DOCUMENT RESUME

ED 325 496 TM 015 719

AUTHOR Johnson, Eugene G.; And Others
TITLE Focusing the New Design: The NAEP 1988 Technical Report.
INSTITUTION National Assessment of Educational Progress, Princeton, NJ.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO ISBN-0-88685-106-8; NAEP-19-TR-20
PUB DATE Sep 90
CONTRACT OERI-G-008720335
NOTE 763p.
AVAILABLE FROM National Assessment of Educational Progress (NAEP), Educational Testing Service, Rosedale Road, Princeton, NJ 08541.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Statistical Data (110)

EDRS PRICE MF04/PC31 Plus Postage.
DESCRIPTORS Annual Reports; Data Analysis; Databases; *Educational Assessment; Elementary Secondary Education; *Federal Programs; Grade 4; Grade 8; Grade 12; *National Surveys; Private Schools; Public Schools; *Research Design; Research Methodology; Sampling; Scaling; Scoring; *Standardized Tests; Test Construction
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

Details of the design and data analysis associated with the 1983 National Assessment of Educational Progress (NAEP) are presented. The 1988 NAEP surveyed American students' knowledge of reading, writing, civics, U.S. history, and geography. Small-scale studies were conducted for mathematics and science. Populations included public school students and private school students (aged 9, 13, and 17 years) as well as students in grades 4, 8, and 12. Geography was surveyed only at grade 12/age 17 years. The objective of these technical notes is to provide information to allow the reader to judge the utility of the design, quality of the NAEP data, reasonableness of the assumptions made, appropriateness of the data analyses, and generalizability of the inferences made from the data. Topics addressed include: development of objectives and items; sample design; assessment instruments; field administration; materials processing and database creation; processing assessment materials; professional scoring; data transcription systems; editing data; quality control of data entry; database products; weighting procedures and estimation of sampling variance; scaling procedures; data analysis for the various subject area assessments; and the statistical summary of the samples and estimates of proficiencies of U.S. students. A total of 204 data tables and 14 figures are provided. Appendices provide: a list of consultants for development of the 1988 NAEP objectives and items; distributions of weight components for the 1988 NAEP samples; contrast codings and estimated effects for 1988 NAEP conditioning variables; 1988 NAEP derived and composite conditioning variables; revision of 1984 NAEP post-stratification weights for grade 4/age 9 years and grade 8/age 13 years; 1988 NAEP item response theory parameters; and NAEP reporting subgroups, composite and derived common background variables, and subject-specific composite and derived reporting variables. A glossary of terms, a list of references cited in the text, and an index are included. (TJH)

ED325496

FOCUSING THE NEW DESIGN:

THE NAEP 1988 TECHNICAL REPORT

FOCUSING THE NEW DESIGN:

THE NAEP 1988 TECHNICAL REPORT

THE NATION'S REPORT CARD naep

Eugene G. Johnson and Rebecca Zwick
in collaboration with

Nancy Allen
John L. Barone
Nancy Caldwell
Anne Campbell
John J. Ferris
David S. Freund
Morris H. Hansen
Lynn B. Jenkins
Janet R. Johnson
Bruce A. Kaplan
Debra L. Kline
Walter B. MacDonald
Nancy A. Mead
Robert J. Mislevy
Ina V.S. Mullis
Norma A. Norris
Alfred M. Rogers
Keith F. Rust
Renee Slobasky
Kentaro Yamamoto

with an Introduction by Albert E. Beaton

September 1990

Prepared by Educational Testing Service under a grant from the National Center for Education Statistics
Office of Educational Research and Improvement • U.S. Department of Education

The Nation's Report Card, the National Assessment of Educational Progress (NAEP), is an education research project mandated by Congress to collect data over time on the performance of young Americans in various subject areas. It makes available information on assessment procedures to state and local education agencies.

The 1988 twelfth-grade geography assessment was funded in part by the National Geographic Society.

This report, No. 19-TR-20, can be ordered from the National Assessment of Educational Progress (NAEP), Educational Testing Service, Rosedale Road, Princeton, New Jersey 08541.

Library of Congress Catalog Card Number: 90-62792

ISBN 0-88685-106-8

The work upon which this publication is based was performed pursuant to Grant No. G-008720335 of the Office of Educational Research and Improvement.

Focusing the New Design: The NAEP 1988 Technical Report

CONTENTS

List of Tables and Figures		vii
Acknowledgments		xiii
Introduction	*Albert E. Beaton*	3

PART I
The Design and Implementation of the 1988 NAEP

Chapter 1	Overview of Part I: The Design and Implementation of the 1988 NAEP *Eugene G. Johnson*	13
Chapter 2	Developing the NAEP Objectives, Items, and Background Questions for the 1988 Assessments of Reading, Writing, Civics, U.S. History, and Geography *Walter B. MacDonald, Ina V. S. Mullis, Anne Campbell, Nancy A. Mead*	33
Chapter 3	Sample Design *Keith F. Rust, Morris H. Hansen*	51
Chapter 4	Assessment Instruments *Janet R. Johnson*	69
Chapter 5	Field Administration *Nancy Caldwell, Renee Slobasky*	85
Chapter 6	Overview of 1988 NAEP Materials Processing and Database Creation *John L. Barone*	99
Chapter 6.1	Processing Assessment Materials *Alfred M. Rogers, Norma A. Norris*	109
Chapter 6.2	Professional Scoring *Lynn B. Jenkins, Anne Campbell*	127
Chapter 6.3	Data Transcription Systems *Alfred M. Rogers*	145
Chapter 6.4	Editing Data *David S. Freund, Alfred M. Rogers*	155
Chapter 6.5	Quality Control of NAEP Data Entry for 1988 *John J. Ferris*	159

Chapter 6.6	Creation of the 1988 NAEP Database *David S. Freund, Alfred M. Rogers*	163
Chapter 6.7	NAEP Database Products *Alfred M. Rogers*	167

PART II
The Analysis of 1988 NAEP Data

Chapter 7	Overview of Part II: The Analysis of 1988 NAEP Data *Rebecca Zwick*	179
Chapter 8	Weighting Procedures and Estimation of Sampling Variance *Eugene G. Johnson, Keith F. Rust, Morris H. Hansen*	187
Chapter 9	Scaling Procedures *Robert J. Mislevy*	229
Chapter 10	Data Analysis for the Reading Assessment *Rebecca Zwick*	251
Chapter 11	Data Analysis for the Writing Assessment *Eugene G. Johnson*	267
Chapter 12	Data Analysis for the Civics Assessment *Nancy Allen*	297
Chapter 13	Data Analysis for the U.S. History Assessment *Norma A. Norris*	317
Chapter 14	Data Analysis for the Geography Assessment *Norma A. Norris*	325
Chapter 15	Data Analysis for Mathematics and Science *Kentaro Yamamoto*	329

PART III
Statistical Summary of 1988 NAEP Data

Chapter 16	Statistical Summary of the 1988 NAEP Samples and Estimates of the Proficiencies of American Students *Bruce A. Kaplan*	343

* * *

Appendix A	Consultants for the Development of 1988 NAEP Objectives and Items	469
Appendix B	Distribution of Weight Components for 1988 NAEP Samples	481

Appendix C	Contrast Codings and Estimated Effects for 1988 NAEP Conditioning Variables	497
Appendix D	1988 NAEP Derived and Composite Conditioning Variables	611
Appendix E	Revision of 1984 NAEP Poststratification Weights for Grade 4/Age 9 and Grade 8/Age 13 *Keith F. Rust*	639
Appendix F	1988 NAEP IRT Parameters	643
Appendix G	NAEP Reporting Subgroups, Composite and Derived Common Background Variables, and Subject-specific Composite and Derived Reporting Variables	679
Glossary of Terms		709
References Cited in Text		715
Index		723

Focusing the New Design: The NAEP 1988 Technical Report

LIST OF TABLES AND FIGURES

Table	1-1	NAEP 1988 student samples	16
	1-2	NAEP subject areas, grades, and ages assessed: 1969-1988	20
	3-1	Reference populations for the components of the NAEP 1988 samples	51
	3-2	Assessment type by age class and assessment period	52
	3-3	Geographic regions used for stratification	53
	3-4	Sampling subuniverses and number of noncertainty strata in each	54
Figure	3-1	Distribution of 94 sampled PSUs by inclusion in fall, winter, and spring	56
Table	3-5	Grade definition of school eligibility for frame inclusion and frame sizes	57
	3-6	School sample sizes, refusals, and substitutes, main samples	60
	3-7	School sample sizes, refusals, and substitutes, bridge samples	62
	3-8	Number of students per school by session type	65
	3-9	Exclusion and participation rates by age class	66
	3-10	Overall participation rates by age class	68
	4-1	Number of main sample subject area cognitive blocks administered	71
	4-2	Booklet contents and number administered, main sample, grade 4/age 9	72
	4-3	Booklet contents and number administered, main sample, grade 8/age 13	73
	4-4	Booklet contents and number administered, main sample, grade 12/age 17	74
	4-5	Main sample block information, grade 4/age 9	75
	4-6	Main sample block information, grade 8/age 13	76
	4-7	Main sample block information, grade 12/age 17	77
	4-8	Booklet contents and number administered, bridge samples, age class 9	78
	4-9	Booklet contents and number administered, bridge samples, age class 13	79
	4-10	Booklet contents and number administered, bridge samples, age class 17	80
	5-1	Summary of NAEP 1988 school participation	92
	5-2	School cooperation rate by age class and type of assessment	92
	5-3	Students sampled, invited, and assessed during the fall assessment	97
	5-4	Students sampled, invited, and assessed during the winter and spring assessments	97
Figure	6-1	Data flow overview	103
Figure	6.1-1	Main menu for the data entry system	111
	6.1-2	First school worksheet data entry screen	113
	6.1-3	Second school worksheet data entry screen	113
	6.1-4	Absentee data entry screen	114
	6.1-5	Main menu for management of the data entry system	118
	6.1-6	Menu for loading of data tapes into data entry system	118
	6.1-7	Student session data entry screen	121
	6.1-8	Student booklet cover data entry screen	121
	6.1-9	Primary menu for entry of questionnaire data	122
Table	6.2-1	NAEP 1988 main sample open-ended items	129
	6.2-2	NAEP 1988 bridge sample open-ended items	130
	6.2-3	Percentages of exact score point agreement and intraclass correlation coefficients for open-ended items in the 1988 main samples	141

Table	6.2-4	Percentages of exact score point agreement and intraclass correlation coefficients for open-ended items in the 1988 bridge samples	143
Figure	6.3-1	NAEP data transcription system (Part 1: Student-related materials)	146
	6.3-2	NAEP data transcription system (Part 2: Questionnaire materials)	147
Table	6.5-1	Summary of quality control error analysis for data entry	161
	8-1	Major subgroups for poststratification in 1988	198
	8-2	Major subgroups for poststratification in 1986 and 1984	200
	8-3	Effect of change in poststratification procedures on relative frequencies and mean reading proficiencies, age 9, 1986	203
	8-4	Effect of change in poststratification procedures on relative frequencies and mean reading proficiencies, age 13, 1986	204
	8-5	Effect of change in poststratification procedures on relative frequencies and mean reading proficiencies, age 17, 1986	205
	8-6	Distributions of design effects across items for the cognitive reading items given in 1988, grade 4	214
	8-7	Distributions of design effects across items for the cognitive reading items given in 1988, grade 8	215
	8-8	Distributions of design effects across items for the cognitive reading items given in 1988, grade 12	216
	8-9	Average design effects by demographic subgroup and grade for mean reading proficiency scores	218
	8-10	Average design effects by grade for simple regression coefficients based on reading proficiency scores	221
	8-11	Effective degrees of freedom for the design effects of the proportion-correct statistics	224
	8-12	Effective degrees of freedom for the design effects for mean reading proficiency scores	226
	8-13	Effective degrees of freedom for the design effects for simple regression coefficients based on reading proficiency scores	227
	9-1	Estimated effects based on full, no, and partial conditioning	247
	9-2	Estimation error variances and related coefficients for the 1988 grade-level reading assessments	248
	10-1	NAEP 1988 reading samples	253
	10-2	Response modes for NAEP 1988 reading items	254
	10-3	NAEP 1988 reading items, BIB samples	257
	10-4	Dichotomization rules for open-ended reading items used in scaling	257
	10-5	Average percent of students failing to reach items for each block position	259
	10-6	Average item percent correct for each block position	260
	10-7	Items for which multiple item response functions were estimated	262
	11-1	Assignment of 1984-1988 writing trend items in 1984 and 1988	269
	11-2	Percentages of exact score agreement and interrater reliability for the primary trait scoring of the 1988 writing trend items	271
	11-3	Mean and standard deviation of (rescore - original) for the 20% rescore of 1984 writing responses	272
	11-4	Distribution of (rescore - original) for the 20% rescore of 1984 writing responses	273
	11-5	Sample sizes for primary trait analyses of trends in writing performance	274
	11-6	Estimable between-item correlations with grade and year	281
	11-7	Sample sizes for mechanics scoring	288
	11-8	Sample sizes for holistic scoring	289
	11-9	Assignment of writing items for cross-sectional writing assessment	290
	11-10	Sample sizes for primary trait analyses of cross-sectional writing performance	290
	11-11	Estimable between-item correlations for the cross-sectional scale with grade	293

Table	11-12	Mean primary trait score by position of item in booklet	294
	12-1	NAEP 1988 civics samples	299
	12-2	Sample sizes and number of items for main assessment grade/age civics samples	300
	12-3	Descriptive statistics for main assessment civics blocks	302
	12-4	Effect of block position on average item percent correct for main focused-BIB civics samples	303
	12-5	Results of the full-information factor analyses of tetrachoric matrices of civics item responses	304
	12-6	Civics means and standard deviations on the calibration scale for grade/age main assessment BIB spiral samples	306
	12-7	Means and standard deviations on the civics proficiency scale	308
	12-8	Civics anchor levels and descriptions	309
	12-9	Dichotomous scoring of the two parts of the open-ended civics item	310
	12-10	Percent agreement for the ratings of the two portions of the dichotomously scored open-ended civics item	310
	12-11	Items deleted from the civics trend analysis because of text changes	311
	12-12	Number of items and sample sizes for the civics trend assessment	312
	12-13	Descriptive statistics for civics 1988 trend blocks	313
	12-14	Items deleted from the civics trend analysis due to lack of model fit	314
	12-15	Means and standard deviations on the civics trend proficiency scale	315
	12-16	Dichotomous scoring of the open-ended civics trend items	316
	12-17	Percent agreement for the ratings of the dichotomously scored open-ended civics trend items	316
	13-1	NAEP 1988 U.S. history samples	318
	13-2	Dichotomized variables for U.S. history	319
	13-3	Descriptive statistics for multiple-choice U.S. history blocks	320
	13-4	Range of percents of items not reached for multiple-choice U.S. history blocks	319
	13-5	U.S. history proficiency scale item information, by sample	321
	13-6	U.S. history item scores excluded from IRT scale	321
	13-7	U.S. history anchoring: Percentages at or above each level, by grade	323
	14-1	NAEP 1988 geography samples	326
	15-1	Mathematics and science samples, 1988 assessment	332
	15-2	Mathematics weighted mean proportion correct	333
	15-3	Coefficients of the linear transformation of the trend scale from original units to the mathematics proficiency scale	334
Figure	15-1	Plot of observed proportion correct of the 1986 BIB spiral and trend assessments with the estimated item response function for an "estimate" item	336
Table	15-4	Science weighted mean proportion correct	338
	15-5	Coefficients of the linear transformation of the trend scale from original units to the science proficiency scale	339
	16-1	Measurement instruments developed for 1988 NAEP	348
	16-2	Numbers of distinct items administered, by age class	349
	16-3	Numbers of distinct cognitive items by age class and sample type	350
	16-4	Characteristics of schools in main NAEP (cross-sectional) samples	351
	16-5	Characteristics of schools in NAEP bridge samples	353
	16-6	Numbers of responses to teacher questionnaire	355
	16-7	Numbers of assessment sessions by sample, type of session, and age class	356
	16-8	Numbers of students assessed and excluded by sample and age class	357
	16-9	Numbers of students in main sample by subgroup, grade 4/age 9	358
	16-10	Numbers of students in main sample by subgroup, grade 8/age 13	359
	16-11	Numbers of students in main sample by subgroup, grade 12/age 17	360

Table	16-12	Numbers of students in bridge to 1984 sample by subgroup, grade 4/age 9	361
	16-13	Numbers of students in bridge to 1984 sample by subgroup, grade 8/age 13	362
	16-14	Numbers of students in bridge to 1984 sample by subgroup, grade 11/age 17	363
	16-15	Numbers of students in bridge to 1986 sample by subgroup, age 9	364
	16-16	Numbers of students in bridge to 1986 sample by subgroup, age 13	365
	16-17	Numbers of students in bridge to 1986 sample by subgroup, grade 11/age 17	366
	16-18	Numbers of students in civics bridge by subgroup, age 13	367
	16-19	Numbers of students in civics bridge by subgroup, age 17	368
	16-20	Numbers of excluded students in main and bridge samples by subgroup, grade 4/age 9	369
	16-21	Numbers of excluded students in main and bridge samples by subgroup, grade 8/age 13	370
	16-22	Numbers of excluded students in main sample by subgroup, grade 12/age 17	371
	16-23	Numbers of excluded students in bridge samples by subgroup, grade 11/age 17	372
	16-24	Numbers of students by sample and age class	373
	16-25	Numbers of students assessed in main assessment, grade 4/age 9	374
	16-26	Numbers of students assessed in main assessment, grade 8/age 13	375
	16-27	Numbers of students assessed in main assessment, grade 12/age 17	376
	16-28	Numbers of students assessed in bridge to 1984, grade 4/age 9	377
	16-29	Numbers of students assessed in bridge to 1984, grade 8/age 13	378
	16-30	Numbers of students assessed in bridge to 1984, grade 11/age 17	379
	16-31	Numbers of students assessed in bridge to 1986, age 9	380
	16-32	Numbers of students assessed in bridge to 1986, age 13	381
	16-33	Numbers of students assessed in bridge to 1986, grade 11/age 17	382
	16-34	Numbers of students assessed in civics bridge, age 13 and age 17	383
	16-35	Weighted percentage of students in main sample by subgroup, grade 4/age 9	384
	16-36	Weighted percentage of students in main sample by subgroup, grade 8/age 13	385
	16-37	Weighted percentage of students in main sample by subgroup, grade 12/age 17	386
	16-38	Weighted percentage of students in bridge to 1984 sample by subgroup, grade 4/age 9	387
	16-39	Weighted percentage of students in bridge to 1984 sample by subgroup, grade 8/age 13	388
	16-40	Weighted percentage of students in bridge to 1984 sample by subgroup, grade 11/age 17	389
	16-41	Weighted percentage of students in bridge to 1986 sample by subgroup, age 9	390
	16-42	Weighted percentage of students in bridge to 1986 sample by subgroup, age 13	391
	16-43	Weighted percentage of students in bridge to 1986 sample by subgroup, grade 11/age 17	392
	16-44	Weighted percentage of students in civics bridge by subgroup, age 13	393
	16-45	Weighted percentage of students in civics bridge by subgroup, age 17	394
	16-46	Weighted percentage of excluded students by subgroup, grade 4/age 9 main and bridge samples	395
	16-47	Weighted percentage of excluded students by subgroup, grade 8/age 13 main and bridge samples	396
	16-48	Weighted percentage of excluded students by subgroup, grade 12/age 17 main sample	397
	16-49	Weighted percentage of excluded students by subgroup, grade 11/age 17 bridge sample	398
	16-50	Numbers of students in main sample with proficiency scores, grade 4/age 9	399
	16-51	Numbers of students in main sample with proficiency scores, grade 8/age 13	400
	16-52	Numbers of students in main sample with proficiency scores, grade 12/age 17	401
	16-53	Numbers of students in bridge to 1984 sample with proficiency scores, grade 4/age 9	402
	16-54	Numbers of students in bridge to 1984 sample with proficiency scores, grade 8/age 13	403
	16-55	Numbers of students in bridge to 1984 sample with proficiency scores, grade 11/age 17	404
	16-56	Numbers of students in bridge to 1986 sample with proficiency scores, age 9	405
	16-57	Numbers of students in bridge to 1986 sample with proficiency scores, age 13	406
	16-58	Numbers of students in bridge to 1986 sample with proficiency scores, grade 11/age 17	407
	16-59	Numbers of students in civics bridge with proficiency scores, age 13	408
	16-60	Numbers of students in civics bridge with proficiency scores, age 17	409
	16-61	Weighted proficiency means, standard deviations, and percentiles with standard errors for main focused-BIB reading samples, by grade and subgroup	410
	16-62	Weighted proficiency means, standard deviations, and percentiles with standard errors for bridge to 1984 reading samples, by age and subgroup	412

Table 16-63 Weighted proficiency means, standard deviations, and percentiles with standard errors for main writing samples, by grade and subgroup 414
16-64 Weighted proficiency means, standard deviations, and percentiles with standard errors for bridge to 1984 writing samples, by grade and subgroup 416
16-65 Weighted proficiency means, standard deviations, and percentiles with standard errors for main focused-BIB civics samples, by grade and subgroup 418
16-66 Weighted proficiency means, standard deviations, and percentiles with standard errors for civics bridge samples, by age and subgroup 420
16-67 Weighted proficiency means, standard deviations, and percentiles with standard errors for main U.S. history samples, by grade and subgroup 422
16-68 Weighted proficiency means, standard deviations, and percentiles with standard errors for main geography sample, by grade and subgroup 424
16-69 Weighted proficiency means, standard deviations, and percentiles with standard errors for bridge to 1986 mathematics samples, by age and subgroup 426
16-70 Weighted proficiency means, standard deviations, and percentiles with standard errors for bridge to 1986 science samples, by age and subgroup 428
16-71 Weighted response percentages and reading proficiency means with standard errors, grade 4, by gender, for main focused-BIB reading samples 430
16-72 Weighted response percentages and reading proficiency means with standard errors, grade 4, by derived race/ethnicity, for main focused-BIB reading samples 431
16-73 Weighted response percentages and reading proficiency means with standard errors, grade 4, by parental education, for main focused-BIB reading samples 432
16-74 Weighted response percentages and reading proficiency means with standard errors, grade 8, by gender, for main focused-BIB reading samples 433
16-75 Weighted response percentages and reading proficiency means with standard errors, grade 8, by derived race/ethnicity, for main focused-BIB reading samples 434
16-76 Weighted response percentages and reading proficiency means with standard errors, grade 8, by parental education, for main focused-BIB reading samples 435
16-77 Weighted response percentages and reading proficiency means with standard errors, grade 12, by gender, for main focused-BIB reading samples 436
16-78 Weighted response percentages and reading proficiency means with standard errors, grade 12, by derived race/ethnicity, for main focused-BIB reading samples 437
16-79 Weighted response percentages and reading proficiency means with standard errors, grade 12, by parental education, for main focused-BIB reading samples 438
16-80 Weighted response percentages and writing proficiency means with standard errors, grade 4, by gender, for main writing samples 439
16-81 Weighted response percentages and writing proficiency means with standard errors, grade 4, by derived race/ethnicity, for main writing samples 440
16-82 Weighted response percentages and writing proficiency means with standard errors, grade 4, by parental education, for main writing samples 441
16-83 Weighted response percentages and writing proficiency means with standard errors, grade 8, by gender, for main writing samples 442
16-84 Weighted response percentages and writing proficiency means with standard errors, grade 8, by derived race/ethnicity, for main writing samples 443
16-85 Weighted response percentages and writing proficiency means with standard errors, grade 8, by parental education for main writing samples 444
16-86 Weighted response percentages and writing proficiency means with standard errors, grade 12, by gender, for main writing samples 445
16-87 Weighted response percentages and writing proficiency means with standard errors, grade 12, by derived race/ethnicity, for main writing samples 446
16-88 Weighted response percentages and writing proficiency means with standard errors, grade 12, by parental education, for main writing samples 447

Table 16-89 Weighted response percentages and civics proficiency means with standard errors, grade 4, by gender, for main focused-BIB civics samples 448
16-90 Weighted response percentages and civics proficiency means with standard errors, grade 4, by derived race/ethnicity, for main focused-BIB civics samples 449
16-91 Weighted response percentages and civics proficiency means with standard errors, grade 4, by parental education, for main focused-BIB civics samples 450
16-92 Weighted response percentages and civics proficiency means with standard errors, grade 8, by gender, for main focused-BIB civics samples 451
16-93 Weighted response percentages and civics proficiency means with standard errors, grade 8, by derived race/ethnicity, for main focused-BIB civics samples 452
16-94 Weighted response percentages and civics proficiency means with standard errors, grade 8, by parental education, for main focused-BIB civics samples 453
16-95 Weighted response percentages and civics proficiency means with standard errors, grade 12, by gender, for main focused-BIB civics samples 454
16-96 Weighted response percentages and civics proficiency means with standard errors, grade 12, by derived race/ethnicity, for main focused-BIB civics samples 455
16-97 Weighted response percentages and civics proficiency means with standard errors, grade 12, by parental education, for main focused-BIB civics samples 456
16-98 Weighted response percentages and U.S. history proficiency means with standard errors, grade 4, by gender, for main U.S. history samples 457
16-99 Weighted response percentages and U.S. history proficiency means with standard errors, grade 4, by derived race/ethnicity, for main U.S. history samples 458
16-100 Weighted response percentages and U.S. history proficiency means with standard errors, grade 4, by parental education, for main U.S. history samples 459
16-101 Weighted response percentages and U.S. history proficiency means with standard errors, grade 8, by gender, for main U.S. history samples 460
16-102 Weighted response percentages and U.S. history proficiency means with standard errors, grade 8, by derived race/ethnicity, for main U.S. history samples 461
16-103 Weighted response percentages and U.S. history proficiency means with standard errors, grade 8, by parental education, for main U.S. history samples 462
16-104 Weighted response percentages and U.S. history proficiency means with standard errors, grade 12, by gender, for main U.S. history samples 463
16-105 Weighted response percentages and U.S. history proficiency means with standard errors, grade 12, by derived ethnicity, for main U.S. history samples 464
16-106 Weighted response percentages and U.S. history proficiency means with standard errors, grade 12, by parental education, for main U.S. history samples 465
16-107 Weighted response percentages and geography proficiency means with standard errors, grade 12, by gender, for main geography samples 466
16-108 Weighted response percentages and geography proficiency means with standard errors, grade 12, by derived race/ethnicity, for main geography samples 467
16-109 Weighted response percentages and geography proficiency means with standard errors, grade 12, by parental education, for main geography samples 468

ACKNOWLEDGMENTS

The comprehensive and important findings reported in the 1988 National Assessment of Educational Progress (NAEP) attest to what can be accomplished when educational policy makers and practitioners, technical experts, and other concerned stakeholders work together to provide the information needed to make informed educational policy decisions.

This report documents the NAEP design and data analysis procedures and the technical thinking that led to their implementation. The report is the result of the contributions of many individuals from many organizations. Individuals from ETS and Westat, Inc., had the responsibility for both carrying out and documenting the procedures reported here. Others have served in facilitating, consulting, and advisory roles. It is appropriate that all of these individuals and organizations be acknowledged here for their contributions to the team effort.

The 1988 NAEP assessment was funded through the Office of Educational Research and Improvement (OERI) of the Department of Education. The staff of OERI have worked in cooperation with ETS and Westat staff in many facilitating roles. In addition, OERI staff have helped to provide the intellectual environment necessary to carry out the sophisticated design and analysis procedures that have become a NAEP trademark. I wish to take this opportunity to thank Emerson Elliott, David Sweet, Gary Phillips, Eugene Owen, Steven Gorman, Maureen Treacy, and their associates at OERI; they have all played a very important role in this endeavor. I also thank the anonymous reviewers for their helpful comments.

The NAEP Design and Analysis Committee (DAC), which oversees the technical aspects of NAEP, has been of exceptional help in both their advisory and supportive roles. The members for the 1988 assessment were Robert Linn (Chair), John Carroll, Robert Glaser, Bert Green, Sylvia Johnson, Ingram Olkin, Tej Pandey, Richard Snow, and John Tukey. New members of the DAC who have also given thoughtful reviews of this report are Jeremy Finn and Huynh Huynh.

I am also grateful to the members of the National Assessment Governing Board (NAGB) for their advice and guidance.

The ETS and NAEP management staffs have been both supporters of and contributors to the technical work of NAEP. Special thanks go to Gregory Anrig, Nancy Cole, Victor Bunderson, and Henry Braun of ETS management and to Archie Lapointe, Ina Mullis, Jules Goodison, and David Hobson of NAEP management.

Ina Mullis deserves special recognition for her leadership and influence on the present and future direction of NAEP. The impact of her experience and expertise can be found in all areas of NAEP. Her wisdom, leadership, and organizational skills have been a source of strength for all those involved in NAEP.

Albert Beaton also deserves special praise for his leadership in the technical aspects of NAEP. He has been the director of NAEP's statistical and psychometric activities from the beginning of ETS' participation, and has helped develop and direct the application of the many new technical innovations that are outlined in this and previous technical reports.

Norma Norris and Jules Goodison deserve special thanks for their demanding role as those responsible for the operational aspects of the 1988 NAEP. I would also like to thank Ina Mullis, Walter MacDonald, and Nancy Mead for their roles in guiding the development of the 1988 assessment objectives and items and Anne Campbell for her part in leading the professional scoring. Ina Mullis and Lynn Jenkins were vital to the process of reporting the results from the assessment; significant contributions to this process were also made by Walter MacDonald and Mary Foertsch. Janet Johnson played an extremely valuable role in facilitating the coordination between NAEP management and NAEP research.

The design and the data analysis of the 1988 NAEP, including the writing of this report, was primarily the responsibility of the NAEP research data analysis staff with significant contributions from both Westat and NAEP management. Many among the NAEP research staff made innovative statistical and psychometric contributions as well as carrying out their more routine supervisory activities. NAEP is particularly indebted for the major contributions of Albert Beaton, Eugene Johnson, Robert Mislevy, Kentaro Yamamoto, and Rebecca Zwick. Nancy Allen, a more recent member of the NAEP research staff, also brought new perspectives and energy to the NAEP tasks. I am grateful also for the contributions of Maxine Kingston and Kathleen Sheehan. Jo-Ling Liang was exceptional in her role as technical assistant.

As in the past, the Data Analysis Department, under the outstanding leadership of John Barone, performed exceptionally in carrying out their responsibilities for developing operating systems and producing data analyses. Alfred Rogers and David Freund deserve special commendation for their excellent leadership in developing and maintaining the large and complex NAEP data management systems. Alfred Rogers also deserves special mention for his role in the development of the new NAEP database products and systemized approach to NAEP data analysis and scaling. In addition, special thanks go to David Freund, Bruce Kaplan, Edward Kulick, Norma Norris, and Jim Ferris for their continuing key roles as supervisors and/or developers of innovative software solutions to NAEP data analysis problems. Other members of the Data Analysis Department who played key roles in the 1988 data analysis were Jennifer Nelson, Michael Narcowich, Inge Novatkoski, Kate Pashley, Minhwei Wang, Ira Sample, and Lee Ann Held.

The staff of Westat, Inc., once again contributed their exceptional organizational skills and talents in all areas of sampling design and data collection. Morris Hansen and Keith Rust developed and supervised the sampling design and Renee Slobasky and Nancy Caldwell carried out the field operations. I would like to thank all of the Westat field staff for their exceptional data collection effort.

Other ETS staff members who contributed to the 1988 design include Margaret Goertz, Ann Jungeblut, and Gita Wilder. I am especially grateful to

William Angoff for his careful reviews and constructive observations during the writing of this report.

Kent Ashworth and Janice Askew were responsible for coordinating the cover design and final printing of the report.

I would like to thank the support staff, including Marge Edmonds, Alice Kass, Donna Lembeck, Kandy Olbrick, June Quattromani, Sharon Stewart, Kay Tyberg, Mary Varone, and Marcy Yates, for their diligence and patience in typing and proofreading the many revisions.

Finally, I would like to give special thanks to Debra Kline, the editor of this and other NAEP technical reports, for her tireless efforts in ensuring that the reports meet the highest standards. As always, she has done an exceptional job of organizing, scheduling, editing, and motivating, and has worked hard to make sure that the final report is coherent and correct. Without her there would be no report.

Donald A. Rock
Coordinating Director
NAEP Research
August 15, 1990

INTRODUCTION

Focusing the New Design: The NAEP 1988 Technical Report

INTRODUCTION

Albert E. Beaton

Educational Testing Service

The 1988 National Assessment of Educational Progress (NAEP) surveyed what students in American schools knew and could do in the subject areas of reading, writing, civics, U.S. history, and geography. Small-scale studies were also carried out for mathematics and science. The populations that were sampled included students enrolled in both public and private schools at ages 9, 13, and 17 as well as students in grades 4, 8, and 12. Geography was surveyed at grade 12/age 17 only.

The purpose of this technical report is to supply the details of the design and data analysis of the 1988 assessment. Our aim is to give the reader sufficient information to judge the utility of the design, the quality of the NAEP data, the reasonableness of the assumptions made, the appropriateness of the data analyses, and the generalizability of the inferences made from the data. For educational psychometricians and statisticians, the report provides a summary of how the technical challenges posed by the 1988 assessment were addressed by NAEP staff. For test development experts, the report provides a detailed account of the evolution and final composition of the 1988 assessment instruments.

The report does not attempt to provide substantive results that might be of interest to educational policy makers; results from the 1988 assessment are provided in a series of NAEP reports on the status of and trends in student performance[1]. This technical documentation is intended to support the

---

[1]*The Reading Report Card, 1971 to 1988: Trends from the Nation's Report Card* (Mullis & Jenkins, 1990); *Learning to Read in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Langer, Applebee, Mullis, & Foertsch, 1990); *The Writing Report Card, 1984-88: Findings from the Nation's Report Card* (Applebee, Langer, Mullis, & Jenkins, 1990); *Learning to Write in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Applebee, Langer, Mullis, Jenkins, & Foertsch, 1990); *The 1988 Civics Report Card: Trends in Achievement from 1976 to 1988 at Ages 13 and 17, and Achievement in 1988 at Grades 4, 8, and 12* (Anderson, Jenkins, Leming, MacDonald, Mullis, Turner, & Wooster, 1990); *The U.S. History Report Card: The Achievement of Fourth-, Eighth-, and Twelfth-grade Students in 1988 and Trends from 1986 to 1988 in the Factual Knowledge of High-school Juniors* (Hammack, Hartoonian, Howe, Jenkins, Levstik, MacDonald, Mullis, & Owen, 1990); and *The Geography Learning of High-school Seniors* (Allen, Bettis, Kurfman, MacDonald, Mullis, & Salter, 1990).

proficiency reports by presenting detailed information on the methods used to derive the results that are presented in those reports.

The technical details of the design and analysis of the 1988 NAEP assessment were the result of a collaborative effort of a large number of persons, including not only the Educational Testing Service (ETS) and Westat, Inc., staff members who collaborated to produce this report, but many others who contributed recommendations, reviews, comments, and other substantial help on technical issues. Of special note are the contributions of the staff of the National Center for Education Statistics, including Emerson Elliott, David Sweet, Gary Phillips, and Eugene Owen, and of the NAEP Design and Analysis Committee (DAC), chaired by Professor Robert Linn (University of Colorado). Other members of the committee included vice-chair Professor Sylvia Johnson (Howard University), Professor Emeritus John B. Carroll (University of North Carolina), Professor Robert Glaser (University of Pittsburgh), Professor Bert Green (Johns Hopkins University), Professor Ingram Olkin (Stanford University), Dr. Tej Pandey (California Department of Education), Professor Richard Snow (Stanford University), and Professor Emeritus John W. Tukey (Princeton University).

FEATURES OF NAEP IN 1988

NAEP has always been innovative. When NAEP first collected assessment data in 1969, it introduced a number of features that were to lead the way in assessment methodology. Then, as now, the National Assessment elected to sample a broad range of knowledge and skills in the subject areas that were surveyed, and so introduced multiple matrix sampling as a way to enlarge the assessment coverage without placing excessive demands on the school time of individual students. NAEP eschewed exclusive reliance on multiple-choice items, and used many open-ended and essay items in its assessment instruments. When funds permitted, NAEP assessments used hands-on methods of measuring student performance. NAEP applied and improved sampling methods and procedures for estimating sampling errors. A brief summary of innovations in NAEP—at its beginning and at this time—is available in *NAEP: On the Cutting Edge of Measurement Since 1969* (ETS, 1990).

The NAEP design underwent a series of major modifications after 1983, when ETS became the grantee. The major goals of NAEP were maintained, but the technology by which they were attained was changed. The new design that was introduced is described in *A New Design for a New Era* (Messick, Beaton, & Lord, 1983). One feature of the new design was the introduction of modern IRT (item response theory) scaling technology in order to summarize efficiently the extensive NAEP data and to communicate the results more effectively to educational policy makers and the public. Another design innovation was the introduction of a more complex form of multiple matrix sampling called BIB (balanced incomplete block) spiraling. BIB spiraling made it possible to maintain the broad coverage of each subject area while adding the ability to estimate the correlations among items, without increasing the amount of testing time for individual students. Another innovation was collecting information from a national probability sample of students who were excluded from the assessment (approximately 5 percent) because they had limited English proficiency, were mildly mentally retarded (educable), or were functionally disabled. Another new feature introduced in the 1984 NAEP was the extension of the sample to cover the modal grades of the students at each age, so that results could be reported by grade as well as by age. A new feature introduced in 1986 was oversampling of Black and Hispanic students, so as to improve the precision of statistics for these subpopulations. The new design also extended the amount of teacher and school information that was made available for analysis.

The new design was introduced in the 1984 assessment and used again with some slight changes in the 1986 assessment. The 1988 NAEP assessment further improved the general design and added a few new features that are discussed below.

It is useful at this point to consider the consequences of introducing a new design into an existing measurement system. There is a clear tension between the need to maintain constant measurement procedures in order to estimate changes in performance and the desire to continue to improve the assessment by using the most modern, best available technology. The new design introduced in 1984 responded to this tension by assessing student achievement in two ways: in one set of samples using the methods of past assessments and in another set using the best available methodology. The samples using the methods of the past were called "bridge" samples, since they provided bridges to the performance of students in past assessments. The result was parallel assessments, using different technologies, that could be compared and for some purposes, perhaps, equated. In this way, innovations could be introduced without losing comparability with the past. Although this flexibility to introduce innovations while maintaining trends has come at the cost of increased complexity, the flexibility does allow NAEP to be responsive to the information needs of policy makers while maintaining the scientific requirements of sophisticated survey research.

The innovative features of NAEP in 1988 were as follows:

Focused-BIB spiraling. BIB spiraling is a special type of multiple matrix sampling in which each item in an assessment is paired with each other item in some booklet so that the interrelationship between any pair of items can be estimated. As originally implemented in the 1984 and 1986 national assessments, the pool of all assessment items in all subject areas was divided into item blocks that typically took a student about 15 minutes to finish. The item blocks were then combined in such a way that each block was paired with each other item block in some assessment booklet. Many assessment booklets were printed and "spiraled" together in a random sequence. Ultimately, each student was assigned an assessment booklet that contained a block of background and attitude questions and three blocks of assessment items. As a result of this process, a student might receive a booklet with items from different subject areas: In 1984, reading and writing were BIB-spiraled together; in 1986, reading, mathematics, science, and computer competence were BIB-spiraled together. The advantage of BIB spiraling across different subject areas was that the correlations among the various subject areas could be explored. The disadvantage of this type of BIB spiraling was that many different assessment booklets had to be printed and assigned to a small number of students. The correlations between subject areas were, therefore, based on small-sized samples.

Focused-BIB spiraling was introduced in 1988 to increase the sample size on which correlations within a subject area were based, at the cost of eliminating the correlations between items in different subject areas. In this case, an assessment booklet consists of a block of demographic background questions, a block of questions about educational experiences in a specific subject area, and three blocks of assessment items all in that same subject area. Proficiency for each student is, therefore, well measured, but only in one subject area. Using this method, far fewer booklets need to be formed, and each different booklet is administered to a larger sample of students. The disadvantage of the focused-BIB spiral design is that no information is accumulated about the relationship between different subject areas. To allow the estimation of correlations between some of the subject areas, the 1988 design also included several special booklets that included blocks of items from different subject areas.

Scaling. The 1988 assessment continued the scaling advances that characterized the 1984 and 1986 assessments. The data collected using the focused-BIB spiral design were carefully checked for unusual properties or differential item functioning in different subpopulations. Although focused-BIB spiraling assures that each student is presented enough items, there is no way to assure that the student will answer enough items in a subject area for precise estimation of his or her proficiency. To avoid the statistical bias introduced by having subsamples that are inadequately measured, the method of plausible values was used in the data analyses. This method also allows for scaling and reporting in narrowly defined content areas. In addition, it provides consistent estimates of changes in proficiency over time, even if there are changes in the number of items per student or the average difficulty of these items. The scales were carefully anchored to enhance public interpretation of various scale points.

Comparable instrument bridges. The 1988 design also contained two equivalent samples of students at each age level that were included to compare the properties of different assessment instruments. The reading assessments in 1984 and 1986 both used BIB-spiraled instruments but the instruments differed in a number of seemingly minor details. In comparing the 1986 reading results to the previously published trend information, the 1986 results seemed anomalous and publication of these results was suspended until further corroborative evidence could be gathered. In 1988, to further investigate the reading anomaly, two equivalent samples of students were selected at each age level, one of which was assigned a booklet from the 1984 assessment and the other of which was assigned a booklet from the 1986 assessment. (Because the booklets from the 1986 assessment contained mathematics and science items, data for these two subject areas were also collected and analyzed.) In both cases, the administrative procedures of the corresponding past assessments were duplicated as carefully as possible. Since the two 1988 samples were randomly equivalent, they are in principle identical in reading proficiency, and any difference between them is due to differences in the assessment instruments and sampling error. These samples were subjected to extensive analysis, and the results were published in a separate report, entitled *The Effect of Changes in the National Assessment: Disentangling the NAEP 1985-86 Reading Anomaly* (Beaton & Zwick, 1990).

Age cohorts. In the original design of NAEP and in the 1984 implementation of the new design, the way in which age was defined at age 17 was different from the way it was defined at ages 9 and 13. As a result of these age definitions, although most 9-year-olds were in the fourth grade and most 13-year-olds were in the eighth grade, most of the 17-year-olds were in the eleventh grade—three instead of four grades ahead. Comparisons across ages were also complicated by the previous NAEP practice of testing 13-year-olds in the fall, 9-year-olds in the winter, and 17-year-olds in the spring of the assessment year.

To facilitate comparisons between different age and grade groups, the 1988 design introduced new age cohorts in the main part of the NAEP design. The definition of 17-year-old students was changed so that the definitions of the ages are similar for the three age groups and the NAEP 9-, 13-, and 17-year-old student cohorts are precisely four years apart. The new definitions now place most of the 17-year-old students in the twelfth grade. In addition, today's population of 9-year-olds will be the population of 13-year-olds sampled in four years and the population of 17-year-olds sampled in eight years. In these samples, students at all ages and grades are assessed in two randomly equivalent half-samples, one assessed between January and mid-March and the other between mid-March and May.

This innovation in the main NAEP sample could not be introduced for the bridge samples used for estimating trends (since the bridge samples had to be comparable to samples from previous assessments), and thus the bridge samples have maintained the former age definitions and times of year for testing.

Experimental samples. Over the years, NAEP has developed a highly successful and innovative way of developing assessment items. However, it is continually exploring new ways to improve its procedures. In 1988, two additional samples were added to NAEP to explore potential measurement improvements. One sample involved the assessment of writing—students were given twice as much time to write essays as in the regular NAEP assessment, and the results of the essays written under different time conditions were compared. This study is reported in *Learning to Write in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Applebee, Langer, Mullis, Jenkins, & Foertsch, 1990). The second special sample was designed to link student performance at ages 13 and 17 to NAEP's 1985 literacy study of young adults.

Poststratification improvements. As discussed in section 8.1.4 of Chapter 8, the process of computing the poststratification adjustments was modified from the procedures followed in 1984 and 1986. The changes were introduced to accelerate the reporting process and to make NAEP results closer to those of the Census Bureau in terms of numbers of students reported in various subpopulations identified by region, race/ethnicity, age, and grade.

Teacher questionnaires. The design for administering teacher questionnaires was modified in 1988 to obtain teacher data for all students at a given grade in a particular subject area, rather than for a few students in all subject areas. In 1988, resources permitted obtaining teacher data for two samples of students. Extensive teacher questionnaires were given: 1) to teachers of fourth-grade students who were assessed in reading and 2) to teachers of eighth-grade students who were assessed in writing. These questionnaires contain not only questions about the teacher's background and teaching practices but also questions for the teacher about the performance of his or her individual students. The results of the assessment of student performance and the teacher questionnaires have been linked for analysis and reported in *Learning to Read in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Langer, Applebee, Mullis, & Foertsch, 1990) and *Learning to Write in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Applebee, Langer, Mullis, Jenkins, & Foertsch, 1990).

Public-use data tapes and other database products. NAEP continues to make available to the public all of its data, except those that would identify its participants. The public-use data tapes are carefully prepared and documented, making it possible for others to duplicate the analyses done by NAEP staff or conduct other analyses of the vast NAEP database.

The NAEP item information database contains all of the descriptive, processing, and usage information for every item developed and administered for NAEP. This database functions as a resource for test development activities, data system control operations, and item linkage to past assessments.

The NAEP restricted-use data files contain all NAEP respondent data, including "secure" files. These files function as NAEP data archives for responses from students, teachers, and school administrators from the booklets and questionnaires used in NAEP from 1970 to 1988.

The NAEP database products are especially useful because they are portable and can be used on a variety of hardware systems; they can be accessed by a variety of software systems (including SAS and SPSS); they are in a "rectangular" file structure that eliminates the need for complex data retrieval processes from dissimilar file formats; and they are well documented.

ORGANIZATION OF THE TECHNICAL REPORT

This technical report is divided into three parts:

Part I presents the details of the 1988 design and contains a summary in Chapter 1 of the steps involved in the process of producing a database ready for analysis. Chapter 1 is followed by chapters in which the details of the several steps in collecting and preparing the data for analysis are described. Chapter 2 describes the specification of the NAEP assessment objectives and the development of the assessment items and background questions for the several assessment instruments and questionnaires. Chapter 3 describes the details of the selection of the NAEP samples. The details of the assessment instruments that were used in the 1988 NAEP are presented in Chapter 4. The administration of NAEP to students in American schools is discussed in Chapter 5. Chapters 6 through 6.7 contain the details of converting the data received from the field into a usable database, including the processing of the raw assessment materials, professional scoring, data transcription systems, editing, quality control, and the creation of the database system, and describes the database products that are available.

Part II begins with Chapter 7, which outlines the analysis of the 1988 NAEP data. Chapter 8 includes general discussions of the weighting and variance estimation procedures used in NAEP analyses. A general discussion of the NAEP scaling methodology is presented in Chapter 9. Chapters 10 through 15 contain the details of the analyses performed for the respective subject areas.

Part III contains some basic data from the 1988 NAEP assessment, including the properties of the measuring instruments, characteristics of the selected sample, and some estimates of the proficiencies of students in American schools. Only a few of the huge number of possible population proficiency estimates are presented, and these include estimates of the means, standard deviations, and selected percentiles of populations of students in various subject areas and grade levels. Estimates are also presented separately for gender, racial/ethnic groupings, and other subpopulations. Estimates of average proficiency values for cross-classifications of selected variables are also reported. Estimated standard errors are reported with all parameter estimates.

PART I

The Design and Implementation of the 1988 NAEP

Chapter 1

OVERVIEW OF PART I: THE DESIGN AND IMPLEMENTATION OF THE 1988 NAEP[1]

Eugene G. Johnson

Educational Testing Service

The 1988 National Assessment collected information on the knowledge, skills, understanding, and attitudes of young Americans in the subject areas of reading, writing, civics, U.S. history, and geography. In addition, information on mathematics and science was collected in a special small-scale study. The basis for this information was a complex sample survey involving more than 126,000 students and consisting of national samples of students aged 9, 13, and 17 as well as national samples of students in grades 4, 8, 11, and 12. This chapter provides a description of the design for the 1988 assessment and gives an overview of the steps involved in the implementation of NAEP from the planning stage through the creation of a database ready for analysis. The major components of the implementation of the assessment are presented here with references to the appropriate chapters in Part I for more details. Not included in Part I are the procedures used for the analysis of the data, these are summarized in the overview to Part II (Chapter 7) and discussed in detail in the remaining chapters in Part II.

The organization of this chapter, and of Part I, is as follows:

- Section 1.1 provides an overview of the NAEP design for 1988 and describes the constituent samples. To provide background, the section also gives the assessment schedule from the inception of NAEP in 1969 through the 1988 assessment.

- Section 1.2 summarizes the four-stage stratified random sampling procedures used for the 1988 assessment with a fuller description provided in Chapter 3.

- Section 1.3 summarizes the development of the objectives for each of the subject areas in the assessment and the development and review of the items written to fit those objectives. Details of the objective and item development processes appear in Chapter 2.

- Section 1.4 discusses the assignment of the cognitive and background questions to assessment booklets and describes the

---

[1]The author is indebted to Albert Beaton and the authors of Chapters 2 through 6 for portions of this chapter, and to Mary Varone for ably typing the manuscript.

focused-BIB spiral design. Chapter 4 provides a complete description of the assessment booklets.

- Section 1.5 summarizes the field administration procedures including the training of field administrators, attaining school cooperation, administering the assessment, and conducting quality control. Further details appear in Chapter 5.

- Section 1.6 describes the flow of the data from their receipt at ETS through data entry, professional scoring, and entry into the database in final form, ready for analysis. Chapters 6 through 6.7 provide a detailed description of the process.

The data collected in the 1988 assessment are available for public use in a set of data tapes containing the data from the sampled students, data about students excluded from the assessment, and data from teachers, principals, and schools linked to the assessed students. The documentation for the public-use data tapes appears in *National Assessment of Educational Progress 1988 Public-use Data Tapes Version 2.0 User Guide* (Rogers, Kline, Johnson, Mislevy, & Rust, 1990).

1.1 THE 1988 NAEP DESIGN

The 1988 NAEP design was intended to address two occasionally competing considerations. The first is NAEP's charge to measure trends in educational achievement. The second is the need for NAEP to evolve as educational issues of interest develop over time and as new technologies of assessment become available. The goal of measuring trends requires a stability in the measurement process; the goal that NAEP evolve implies, however, that the measurement process must be permitted to change.

The 1988 design addresses the competition between stability and change by basing the assessment on two distinct types of samples. The first type of sample, the trend sample, is used for estimating changes in performance from previous assessments in a subject area and uses the same methodology and population definitions as in previous assessments. The second type of sample, a cross-sectional sample, is used for detailed information about the current student population. This sample allows the use of new technology and population definitions and addresses new educational issues.

A number of improvements have been made in the design of NAEP since ETS' first assessment in 1984. Before the 1984 assessment, NAEP used a simple matrix sampling procedure with audiotape pacing—all students in an assessment session received the same booklet of assessment items and an aurally presented stimulus was used to pace the students through the assessment items. In the 1984 assessment, balanced incomplete block (BIB) spiraling (discussed in section 1.4), which does not include aural pacing, was instituted in place of taped matrix sampling. With BIB spiraling, students in an assessment session receive different booklets resulting in a more efficient sample (for reasons given in section 1.4). BIB spiraling also allows the study of the interrelationships between all items included in the balanced incomplete block design. In the 1988 assessment, additional efficiencies were introduced when focused-BIB spiraling was instituted. Focused-BIB spiraling ensures that all correlations between items *within* a subject area can be estimated but, unlike the BIB designs used in 1984 and 1986, does not require that correlations between items in *different* subject areas be estimable. This produces a simpler and more efficient design than that used in 1984 and 1986. Like BIB spiraling, focused-BIB spiraling required the elimination of the audiotape pacing used in earlier assessments.

From its inception, NAEP has sampled students of a given age (9, 13, and 17). Since 1984, the designs have also included samples from the population of the corresponding modal grades (the grade in school of the majority of the students of the age level). As was the case for previous national assessments, the primary populations of inference for the 1988 assessment were in-school students of the specified ages or grades. Thus, youth of the specified age who were not attending public or private school at the time of the assessment (including, in particular, dropouts and early graduates) were excluded from the sample and from the population of inference.

To assure that the grade/age samp`.. measure four years of growth, the definitions of student age were made unif›rm for the 1988 assessment. While NAEP originally defi-.d age by birth within a calendar year for ages 9 and 13 but defined age 17 as being born between October 1 of one year and September 30 of the next, the 1988 assessment placed all ages on the calendar-year basis. This change in age definition changed the modal grade for 17-year-old students from the eleventh grade to the twelfth grade. Since their age definitions were unchanged, the modal grades for ages 9 and 13 remained at grades 4 and 8.

A final change in the 1988 assessme.it, relative to previous assessments, in the direction of greater uniformity and, therefore, greater comparability, was in the time of year that the students were assessed. NAEP traditionally assessed 9-year-olds in the winter, 13-year-olds in the fall, and 17-year-olds in the spring; in the 1988 assessment, all students were assessed in the winter and spring.

1.1.1 The 1988 Samples

The full 1988 assessment consists of four types of samples: main assessment focused-BIB samples, main assessment intercorrelation samples, main assessment special study samples, and bridge (trend) samples. A list of all assessment samples, with 'ey characteristics, appears in Table 1-1. A description of the samples follows.

Main assessment f~ .-BIB samples. These samples form the basis for the cross-sectional a~ of achievement for the 1988 student population and assessed the su' reas of writing (labeled as [Main-Wrt] in Table 1-1), reading [(Mai. Rdg], U.S. history [Main-His], civics [Main-Civ] and geography [Main-Geo]. All but geography were assessed at all three age

Table 1-1
NAEP 1988 Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Defn.	Modal Grade	Sample Size
9[Main-Wrt]	1-7	Print	Grade 4/age 9	Winter, spring	CY	4	6247
13[Main-Wrt]	1-7	Print	Grade 8/age 13	Winter, spring	CY	8	6011
17[Main-Wrt]	1-7	Print	Grade 12/age 17	Winter, spring	CY	12	5740
9[Main-Rdg]	8-14	Print	Grade 4/age 9	Winter, spring	CY	4	6177
13[Main-Rdg]	8-14	Print	Grade 8/age 13	Winter, spring	CY	8	5912
17[Main-Rdg]	8-14	Print	Grade 12/age 17	Winter, spring	CY	12	5768
9[Main-His]	15	Print	Grade 4/age 9	Winter, spring	CY	4	2664
13[Main-His]	15-21	Print	Grade 8/age 13	Winter, spring	CY	8	5988
17[Main-His]	15-21	Print	Grade 12/age 17	Winter, spring	CY	12	5780
9[Main-Civ]	16	Print	Grade 4/age 9	Winter, spring	CY	4	2652
13[Main-Civ]	22-28	Print	Grade 8/age 13	Winter, spring	CY	8	5981
17[Main-Civ]	22-28	Print	Grade 12/age 17	Winter, spring	CY	12	5683
17[Main-Geo]	29	Print	Grade 12/age 17	Winter, spring	CY	12	2446
9[Main-Int]	17-19	Print	Grade 4/age 9	Winter, spring	CY	4	2638
13[Main-Int]	29-31	Print	Grade 8/age 13	Winter, spring	CY	8	2590
17[Main-Int]	30-32	Print	Grade 12/age 17	Winter, spring	CY	12	2438
9[Main-LWr]	20-22	Print	Grade 4/age 9	Winter, spring	CY	4	2634
13[Main-LWr]	32-34	Print	Grade 8/age 13	Winter, spring	CY	8	2586
17[Main-LWr]	33-35	Print	Grade 12/age 17	Winter, spring	CY	12	2438
13[Main-Doc]	35-36	Print	Grade 8/age 13	Winter, spring	CY	8	2533
17[Main-Doc]	36-37	Print	Grade 12/age 17	Winter, spring	CY	12	2425
9[Br84-RW]	51-56	Print	Grade 4/age 9	Winter	CY	4	5188
13[Br84-RW]	51-56	Print	Grade 8/age 13	Fall	CY	8	5500
17[Br84-RW]	51-56	Print	Grade 11/age 17	Spring	not-CY	11	4622
9[Br86-RMS]	91-93	Mixed	Age 9	Winter	CY	4	3711
13[Br86-RMS]	91-93	Mixed	Age 13	Fall	CY	8	3942
17[Br86-RMS]	61-66	Print	Grade 11/age 17	Spring	not-CY	11	4703
17[Br86-His]	67	Print	Grade 11/age 17	Spring	not-CY	11	2349
13[BrCiv]	90	Tape	Age 13	Fall	CY	8	1938
17[BrCiv]	90	Tape	Age 17	Spring	not-CY	11	1786

Legend

Rdg = Reading
Wrt = Writing
His = U.S. History
Civ = Civics
Geo = Geography
Int = Intercorrelation

LWr = Long Writing
Doc = Document Literacy
RW = Reading and Writing
RMS = Reading, Mathematics, and Science

CY = Calender year

classes; geography was assessed only at grade 12/age 17. In these samples, focused-BIB spiraling (and hence printed administration) was used, age was defined on a calendar-year basis and both age populations (9, 13, 17) and modal grade populations (4, 8, 12) were sampled. Each age class sample was divided into two random half-samples, one of which was assessed in the winter and the other in the spring. The purposes of these half-samples were: (1) to allow comparison with other selected samples (assessed only in the winter or spring); and (2) to allow the study of growth in student achievement within a school year.

Main assessment intercorrelation samples. These samples (labeled [Main-Int] in Table 1-1) are intended to permit the measurement of interrelationships in achievement between subject areas. Each booklet in these samples includes blocks of reading, civics, and U.S. history items (as well as some geography items at grade 12/age 17). Print administration was used and age was defined on a calendar-year basis. Both grades and ages were sampled and the assessment was conducted in two random half-samples in the winter and spring.

Main assessment special study samples. These samples are designed to allow the conduct of special studies relating to achievement. The long writing sample [Main-LWr] is intended to measure the relationship between time allocated to the writing task and writing performance; the document literacy sample [Main-Doc] allows the assessment of document literacy and its relation to reading proficiency. Both samples used printed administration, age and grade sampling, and new age definitions and time of testing.

The 1988 assessment also includes a number of additional samples designed to determine the possible effects of changes in age definitions, time of testing, and mode of administration (audiotape pacing versus print administration) and to provide links to the results from previous assessments. Because the purpose of these samples is to provide a linkage between the 1988 data and data from previous assessments, they are referred to as bridge samples (although they are also called trend samples). The various bridge samples are as follows:

Civics Bridge to 1976 and 1982. This bridge (trend) sample, labeled [BrCiv] in Table 1-1, addresses the subject area of civics. The samples for this bridge are comparable to past assessments of citizenship and social studies. Like these past assessments, the civics bridge sample uses tape recorders and pre-1984 definitions of age and time of testing. Since trend data have been traditionally collected only by age, grade sampling was unnecessary. The civics bridge sample consists of one booklet for age 13 and one booklet for age 17. Because there were no reusable civics items from previous assessments of 9-year-olds, an age 9 sample was not needed.

Bridge to 1984. This bridge (trend) sample, labeled [Br84-RW] in Table 1-1, consists of samples comparable to the 1984 main assessment and addresses the subject areas of reading and writing. The samples were collected by grade and age for grade 4/age 9, grade 8/age 13, and grade 11/age 17, using the age definitions and time of testing from 1984. Six assessment booklets were administered at each grade/age. Each booklet consisted of at least one block of reading items and at least one block of writing items. The administration of these booklets was nonpaced (that is, no audiotape was used).

Bridge to 1986, Ages 9 and 13. This bridge consists of trend samples for ages 9 and 13 comparable to those used for the measurement of trends in 1986. The samples are labeled 9[Br86-RMS] and 13[Br86-RMS] in Table 1-1. The samples were collected by age only and used the same age definitions and time of testing as in 1984 and in the 1986 bridge to 1984. The subject areas addressed by this bridge are reading, mathematics, and science. Three assessment booklets were administered to each age group. These booklets were identical to those administered in 1986. Each booklet contains one block of reading, one block of mathematics, and one block of science items. As in 1986, administration of the mathematics and science blocks was paced with an audiotape; the reading blocks were administered without an audiotape.

Bridges to 1986, Grade 11/Age 17. These bridges (labeled 17[Br86-RMS] and 17[Br86-His]) consist of trend samples of grade 11/age 17 students comparable to the 1986 main assessment sample and were selected and administered using the same age definition and time of testing as in that assessment. Since those definitions also apply to samples from the 1984 and earlier assessments, the students in these bridge samples are comparable to the students from these earlier assessments. (However, the performance results are not directly comparable because the earlier assessments had paced audiotape administrations.) The subject areas assessed in the 17[Br86-RMS] bridge were reading, mathematics, and science and consisted of six assessment booklets administered to grade 11/age 17 students. These booklets contained blocks of reading, mathematics, and science items. The 17[Br86-His] bridge involved one booklet consisting entirely of blocks of U.S. history items. The administration of the booklets in both bridges was nonpaced.

Although many of the questions in the assessment booklets for the bridge samples also appeared in the booklets used for the main assessment, the bridge assessment instruments were considered as distinct from the main assessment instruments. Additionally, the procedures used to administer the bridge assessment instruments sometimes differed from those used for the main assessment. A overview of the assessment instrumentation and mode of administration appears in section 1.4, with further details presented in Chapter 4. Details of the analysis of the bridge and main assessment data appears in Part II of this technical report.

1.1.2 NAEP Assessments Since 1969

Table 1-2 shows the subject areas, grades, and ages assessed since the inception of NAEP in 1969. As can be seen, besides the 1988 subject areas of reading, writing, civics, U.S. history, geography, mathematics, and science, many other subject areas have been assessed over the years—social studies, citizenship, literature, music, career development, art, and computer competence. Many subject areas have been reassessed periodically to determine trends over time.

Assessments were conducted annually through 1980, but budget restrictions since then have reduced data collection to a biennial basis. Since its inception, NAEP has assessed 9-year-olds, 13-year-olds, and in-school 17-year-olds, although the age definitions changed in 1986 and again in 1988. Because of budget restrictions, NAEP no longer routinely assesses out-of-school 17-year-olds or young adults. (A separate assessment of young adults or ages 21 to 25 was conducted in 1985 under a separate grant.)

The table also indicates the initiation of data collection by grade as well as by age in 1984, a practice that was continued in the 1986 and 1988 assessments. It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the younger two ages were defined on a calendar-year basis while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. In the 1986 main assessment, all ages were defined on a October 1 to September 30 basis resulting in modal grades of 3, 7, and 11. In 1988, the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12.

1.2 THE 1988 SAMPLE DESIGN

The target population for the 1988 main assessment consists of all students in public and private schools who belong to one of three cohorts: students who were *either* in the fourth grade *or* 9 years old; students who were *either* in the eighth grade *or* 13 years old; and students who were *either* in the twelfth grade *or* 17 years old. The main assessment represents two overlapping samples. The first sample represents students of the grades 4, 8, and 12 (who could be of any age)—these are the modal grades for the students of the specified ages. The second sample represents students of specified ages (who could be of any grade). Students were age-eligible for the main assessment if they were born in the appropriate calendar year (1978, 1974, or 1970). Only students who were attending public or private schools at the time of the assessment were included in the sample (and, therefore, in the target population). Specifically excluded from the sample and the target population are early graduates and dropouts.

For the purposes of analysis, the grade/age samples were treated as two separate samples: 1) a representative sample of students in grades 4, 8, and 12, and 2) a representative sample of students of ages 9, 13, and 17. (A student who was both grade and age eligible, was regarded as a member of both

Table 1-2

National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1988

Assessment Year	Subject Area(s)	Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS[a]	Adult
						Grades/Ages Assessed						
1969-70	Science			x			x			x	x	x
	Writing			x			x			x	x	x
	Citizenship			x			x			x	x	x
1970-71	Reading			x			x			x	x	x
	Literature			x			x			x	x	x
1971-72	Music			x			x			x	x	x
	Social Studies			x			x			x	x	x
1972-73	Science			x			x			x	x	x
	Mathematics			x			x			x	x	x
1973-74	Career and Occupational Development			x			x			x	x	x
	Writing			x			x			x	x	
1974-75	Reading			x			x			x	x	
	Art			x			x			x	x	
	Index of Basic Skills									x	x	
1975-76	Citizenship/Social Studies			x			x			x	x	
	Mathematics[b]						x			x	x	

[a] Age 17 students who had dropped out of school or had graduated prior to assessment.
[b] Small, special-interest assessment conducted on limited samples at specific grades or ages.

Table 1-2 (continued)

National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1988

Assessment Year	Subject Area(s)	Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS[a]	Adult
		Grades/Ages Assessed										
1976-77	Science			x			x			x		
	Basic Life Skills[b]									x		
	Science, Reading, Health[b]											x
1977-78	Mathematics			x			x			x		
	Consumer Skills[b]									x		
1978-79	Writing, Art, and Music			x			x			x		
1979-80[d]	Reading/Literature			x			x			x	x	
	Art						x					
1981-82	Science[b]			x			x			x		
	Math and Citizenship/ Social Studies			x			x			x		
1984[c]	Reading		x	x		x	x	x		x		
	Writing		x	x		x	x	x		x		
1985[c]	Adult Literacy[b]											x

[a] Age 17 students who had dropped out of school or had graduated prior to assessment.
[b] Small, special-interest assessment conducted on limited samples at specific grades or ages.
[c] Assessment conducted by Educational Testing Service.
[d] Because of reduced funding, assessments were conducted biennially after 1979-80.

Table 1-2 (continued)

National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1988

Assessment Year	Subject Area(s)	Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS[a]	Adult
1986[c]	Reading	x		x	x		x	x		x		
	Mathematics	x		x	x		x	x		x		
	Science	x		x	x		x	x		x		
	Computer Competence	x		x	x		x	x		x		
	U.S. History[b]							x		x		
	Literature[b]							x		x		
1988[c]	Reading		x	x		x	x		x	x		
	Writing		x	x		x	x		x	x		
	Civics		x	x		x	x		x	x		
	U.S. History		x	x		x	x		x	x		
	Document Literacy[b]					x	x		x	x		
	Geography[b]								x	x		
	Mathematics[b]			x			x	x		x		
	Science[b]			x			x	x		x		

[a] Age 17 students who had dropped out of school or had graduated prior to assessment.
[b] Small, special-interest assessment conducted on limited samples at specific grades or ages.
[c] Assessment conducted by Educational Testing Service.

37

38

of these samples.) Thus, summary statistics were computed for the age sample or for the grade sample, but generally not for the combined grade/age sample.

The target population for the bridge assessments also consisted of all public and private school students in one of three cohorts, but the age and grade definitions for one cohort differed from that used in the main assessment. All bridge samples contained samples of students eligible by age using the following age definitions (consistent with those used in 1984 and previous assessments):

Students were eligible for the age 9 and the age 13 bridge assessments if they were born in the appropriate calendar year (1978 and 1974, respectively); students were eligible for the age 17 bridge assessments if they were born between October 1, 1970 and September 30, 1971.

For certain bridge samples, students eligible by grade were also selected. These samples consisted of students in grades 4, 8, and 11, the modal grades for the age definitions used by the bridge samples.

The sample for the 1988 NAEP assessment was selected using a complex four-stage sample design involving the sampling of students from selected schools within 94 selected geographic areas, called primary sampling units, across the United States. The sample design is similar to that used in 1986 and is described in detail by Westat, Inc., the firm subcontracted by ETS to select the sample, in *1988 National Assessment of Educational Progress—Sampling and Weighting Procedures, Final Report* (Rust, Bethel, Burke, & Hansen, 1990). The following sections provide an overview of each of the four stages of the sampling design with further details given in Chapter 3.

Stage 1: Primary Sampling Units

In the first stage of sampling, the United States (the 50 states and the District of Columbia) was divided into geographic primary sampling units (PSUs). Each PSU met a minimum size requirement and generally comprised either a metropolitan statistical area (MSA), a single county, or a group of contiguous counties. The PSUs were classified into four regions (Northeast, Southeast, Central, West), each containing about one-fourth of the U.S. population. In each region, PSUs were additionally classified as MSA or nonMSA. In the Southeast and West regions, the PSUs in which 20 percent of the population in the 1980 Census was either Black or Hispanic were further classified as high-minority, while the remaining PSUs in those regions were classified as not high-minority. This resulted in twelve subuniverses of PSUs.

Ninety-four PSUs were then selected from these subuniverses with probability proportional to a measure of their size (the number of school-age children from the 1980 census). Thirty-four large PSUs were designated as certainty units, as it was cost effective to include them in the sample with certainty. Within each major stratum (subuniverse), further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics. Sixty PSUs were selected from the noncertainty strata with probability proportional to size. To enlarge the samples of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups, PSUs from the high-minority subuniverses were sampled at twice the rate of those from the other subuniverses.

The 94 PSUs were used for the main assessments of all three age classes. To facilitate and improve administration of the assessments and to allow for some estimation of growth in achievement during the school year, the main assessment sample was divided into two randomly equivalent subsamples, one to be assessed in the winter and the other to be assessed in the spring. For this purpose, the 94 PSUs were designated as winter PSUs, spring PSUs, or both winter and spring PSUs, according to the scheme detailed in Chapter 3.

The bridge assessments used a subsample of the 94 PSUs used for the main assessment. The grade 4/age 9 bridge assessments, which were conducted in the winter, used the 56 PSUs designated as winter PSUs in the main assessment; the grade 11/age 17 bridge assessments, conducted in the spring, used the 56 PSUs designated as spring PSUs. (The 18 largest PSUs were used in both the winter and the spring assessments.) The grade 8/age 13 bridge assessments, conducted in the fall, used 64 PSUs selected from the complete set of 94 PSUs with probability proportional to the measure of size of the strata from which the PSUs were selected. As for the winter and spring subsamples, the 18 largest certainty PSUs were retained in the fall bridge sample with certainty. Again, the scheme detailed in Chapter 3 shows the relationship between these PSU samples.

Stage 2: Sampling Schools

In the second stage of sampling, the public, private, Catholic, Bureau of Indian Affairs, and Department of Defense schools within each of the 94 PSUs were listed according to the three grade/ages. An independent sample of schools was selected separately for each of the grade/ages so that some schools were selected for assessment of two grade/ages, and a few were selected for all three. Schools within each PSU were selected (without replacement) with probabilities proportional to assigned measures of size. (Details of the probabilities used for school selection appear in Chapter 3.) Overall probabilities of selection for high-minority schools were twice those for other schools in order to enlarge the sample of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups.

The overall school cooperation rate exceeded 85 percent at each grade/age. In certain instances, refusing schools were replaced by substitutes according to the rules indicated in Chapter 3.

For all three grade/ages, a sample of schools was first drawn for the bridge assessments according the procedures detailed in Chapter 3. These schools were then excluded from the frame when the samples of schools were drawn for the main assessments. Appropriate adjustments were made to the sample weights for both bridge and main samples. Schools assigned main assessment sessions were further classified as belonging to the winter main assessment or the spring main assessment.

Stage 3: Assigning Assessment Sessions to Schools

In the third stage of sampling, assessment sessions were assigned to the sampled schools, as described in section 3.3 of Chapter 3. An assessment session typically consisted of 25 to 30 students, all of whom could be assessed following the same procedures. There were two general types of sessions in the 1988 assessment: 1) tape sessions, where every student was administered the same booklet and where audiotape prompts paced the students through at least part of the booklet, and 2) print sessions, where a number of distinct booklets were administered and where no audiotape pacing was used. (Print sessions are also called spiral sessions, since the assessment booklets were spiraled for administration—see section 1.4.1.)

The assignment of sessions to schools was designed to maximize the number of session types conducted within each PSU, where each session type corresponded to a separate sample of the population of students. In most sample schools, four types of sessions were conducted although schools with fewer than 20 eligibles were asked to conduct only a single session.

.age 4: Sampling Students

In the fourth stage of sampling, a consolidated list was prepared for each school of all grade-eligible and age-eligible students for the age class for which the school was selected. To provide the target sample size, a systematic selection of eligible students was made from this list, if necessary. In small and medium-sized schools all eligible students were in the sample. For bridge (or trend) sample schools assigned to more than a single session type, students were assigned by Westat district supervisors to print or paced-tape sessions using specified procedures. A student was not assigned to more than one session. Students assigned to paced-tape sessions who were not age-eligible were dropped from the assessment.

Stage 4a: Excluded Students

Some students selected for the sample were deemed unassessable by school authorities because they had limited English language proficiency, were judged as being mildly mentally retarded (educable), or were functionally disabled. For each of these students, school staff completed an excluded student questionnaire, listing the reason for exclusion and providing some background information.

Stage 4b: Sampling Teachers

The reading teachers of fourth-grade students sampled for the main assessments of reading and the writing teachers of eighth-grade students sampled for the main assessments of writing were identified and asked to complete a questionnaire about the reading or writing capabilities of each selected student and about the kinds of instruction received.

Stage 4c: The School and Principal Questionnaires

A school characteristics and policies questionnaire was mailed to every sampled school by Westat before the assessment. The Westat supervisor then collected the questionnaires and returned them to ETS. The school characteristics and policies questionnaire is described in Chapter 4. The principal questionnaire, distributed to the principal of each sampled school by Westat before the assessment, was used to estimate the number of grade/age-eligible students and to determine the size and type of community used in assigning the STOC codes.

1.3 DEVELOPMENT OF ASSESSMENT OBJECTIVES, ITEMS, AND BACKGROUND QUESTIONS

In 1988, NAEP conducted main assessments of students at all three ages in the subject areas of reading, writing, civics and U.S. history. Also part of the main assessment was a separately funded study of geography achievement at grade 12/age 17 and a special NAEP study of document literacy at grade 8/age 13 and grade 12/age 17. These assessments entailed the generation of a large number of cognitive items. In addition, a large number of background and attitude questions were asked of students and information was collected from principals and teachers. Details of the item development procedures followed for the 1988 main assessment are given in Chapter 2; this section provides an overview. (In addition to the main assessment, bridge studies were performed in reading, writing, civics, U.S. history, mathematics, and science. Since the instruments used for these studies consisted entirely of items used in previous assessments, no developmental tasks were required for their use in the 1988 assessment.)

The development of items for each subject area was supervised by a Learning Area Committee consisting of educators, scholars, and citizens representing many diverse constituencies and points of view. Each Learning Area Committee developed a set of objectives for its subject area, proposing goals that students at each grade/age level should achieve. After careful and extensive review, the objectives were given to item writers to develop assessment items to fit the objectives. Besides specifying the types of cognitive items to be used to measure academic achievement, the Learning Area Committees were also responsible for the development of items to measure student backgrounds, attitudes, experiences, and interests as they relate to the subject area.

Four additional types of instruments were developed for the 1988 assessment: a common student background questionnaire to be given to all assessed students of a given grade/age, a school characteristics and policies questionnaire, teacher questionnaires for teachers of fourth- and eighth-grade students, and an excluded student questionnaire. Each of these questionnaires was developed through a broad-based consensus process.

All items in the assessment underwent extensive reviews by subject area and measurement specialists, as well as careful scrutiny to eliminate any potential bias or lack of sensitivity to any group. Further, the items were field tested on a representative group of students. Based on the results of the field test, items were revised or modified as necessary and then again reviewed for lack of sensitivity to particular groups. With the help of staff and outside reviewers, the Learning Area Committee selected the items to include in the assessment.

Nearly every subject area included both multiple-choice and open-ended items. The exceptions were writing, which consisted entirely of open-ended items, and geography and science, which consisted entirely of multiple-choice items. The open-ended items were professionally scored; the details of the professional scoring process are given in Chapter 6.2.

1.4 ASSESSMENT INSTRUMENTATION

Four types of instruments were used in the 1988 assessment: student assessment booklets, excluded students questionnaires, teacher questionnaires, and a school characteristics and policies questionnaire. This section provides an overview of these instruments; more information about the instruments can be found in Chapter 4.

1.4.1 Student Assessment Booklets—Main Assessment

The student assessment booklets for the main assessment contained both cognitive and noncognitive items. The total testing time was approximately 45 minutes for grade 4/age 9 students and 56 minutes for the older ages. A block of common background questions appeared first in every booklet and required 10 minutes for completion for grade 4/age 9 and 6 minutes for the older students. This was followed by a 5-minute block of subject-specific background questions and (typically) three 15-minute blocks of cognitive items (10-minute blocks for grade 4/age 9).

The assembly of cognitive items into booklets and their subsequent assignment to assessed students was determined by a *balanced incomplete block* (BIB) design with *spiraled* administration. The first step in implementing BIB spiraling is to divide the items within a subject area into units called blocks, where each block was designed to take 15 minutes for the older students to complete. For the grade 4/age 9 students, blocks requiring 10 minutes for completion were created. Some blocks were administered at more than one grade/age; additional items were added to the end of grade 4/age 9 blocks which were also administered to older students.

These blocks were then assembled into booklets containing the background questions and three blocks of subject area items according to a partially balanced incomplete block design. (In a completely balanced incomplete block design, the subject area blocks would be assigned to booklets in such a way that each block appears in the same number of booklets and every pair of blocks appears together in exactly one booklet. This is the *balanced* part of the method; the *incomplete* part refers to the fact that no booklet contains all items and hence incomplete data is yielded for each assessed student. Such a design allows the computation of the correlation between each pair of items but generates a vast number of different booklets, particularly if blocks from different subject areas are to be paired.)

In 1988, the BIB design was focused—that is, each block of items within each of the subject areas was paired with every other block within that subject area but generally, not with blocks of items from other subject areas (special booklets were constructed to measure relationships between subjects). The focused-BIB design used in 1988 called for seven blocks of cognitive items at a given grade/age to be assembled into seven booklets, where each pair of the seven blocks appears in exactly one booklet and where each block appears in three booklets. The focused-BIB design also balances the order of presentation of the blocks of items in the sense that every block of items appears as the first cognitive block in one booklet, as the second block in another booklet, and as the third and last block in a third booklet.

An example of the focused-BIB design with seven blocks (labeled A, B, C, D, E, F, G) and seven booklets (labeled B1 through B7) is as follows:

Booklet	Blocks		
B1	A	B	D
B2	B	C	E
B3	C	D	F
B4	D	E	G
B5	E	F	A
B6	F	G	B
B7	G	A	C

In addition to the focused-BIB booklets, three types of special booklets were created for the main samples. To permit the calculation of correlations among items between subject areas, three special intercorrelation booklets containing one block each of reading, U.S. history, and civics items were created at each grade/age. Three special writing booklets were also created at each grade/age. Each of these booklets consisted of a standard-length writing block and a writing block designed to take twice the regular block time. The purpose of these booklets was to allow an evaluation of the effect of the amount allocated time on writing performance. Finally, two document literacy booklets were created at grade 8/age 13 and grade 12/age 17.

A total of 22 different booklets were assembled for grade 4/age 9, 36 different booklets for grade 8/age 13, and 37 different booklets for grade 12/age 17. These booklets were then spiraled and placed into bundles.

Spiraling involves interleaving the booklets in regular (systematic) sequence so that each booklet appears an appropriate number of times in the sample. Booklets were packaged together in bundles of 25 to 27 booklets, which was large enough to accommodate a typical assessment session. The bundles were designed so that each booklet would appear equally often in each position in a bundle.

The final step in the BIB-spiraling procedure is the assigning of the booklets to the assessed students. The students within an assessment session were assigned booklets in the order in which the booklets were bundled. The result was that, typically, each student in an assessment session received a different booklet and, even in schools with multiple sessions, only a few students received the same booklet or block of items. In the 1988 BIB-spiral design, representative and randomly equivalent samples of about 2,600 grade- or age-eligible students responded to each item (resulting in samples of about 2,000 students eligible by age and 2,000 eligible by grade).

BIB spiraling differs from the simpler matrix sampling scheme used by NAEP prior to 1984 to assign items to students. In the earlier scheme, the pool of items was divided into distinct booklets requiring about 45 minutes to administer, and all students within an assessment session were given the same booklet. Because all students in a session received the same booklet, it was possible to accompany the administration with a paced audiotape of the exercise stimuli, with the aim of minimizing the effect of a student's reading ability on performance in other subject areas. However, since each item appeared in a single booklet, it was impossible to estimate correlations between items appearing in different booklets. Furthermore, the administration of the same items to clusters of students within schools results in an increase in sampling variability over an unclustered sample of the same size because of intracluster correlation.

The BIB-spiral design permits the estimation of correlations between all items within a subject area. Furthermore, since the spiral design presents each block of items to fewer persons in any school, but to more schools, than the simple matrix sampling design, the cluster effect is markedly reduced, leading to a sample with high statistical efficiency. The spiral design does preclude the use of audiotape pacing. Since each student within a session responds to a different set of items, the instructions and the items themselves must be read by the student as audiotape administration would be unmanageable.

1.4.2 Student Assessment Booklets—Bridge Samples

There were four distinct bridge samples in the 1988 assessment, each of which required the creation of special booklets.

Civics Bridge. One booklet was created for each of the ages 13 and 1[illegible] At each age, the booklet consisted of a common background block, a civics background and attitude block and three blocks of cognitive items. The background blocks were from the main assessment; the cognitive blocks contained items used in previous assessments as well as items used in the 1988 assessment. The booklets were administered to all students within a session using audiotape pacing.

Bridge to 1984. Six booklets were used at each of the three grade/age classes. These booklets were identical to booklets used in the 1984 assessments of reading and writing. Each booklet consisted of a common background block and three cognitive blocks, either two reading and one writing or one reading and two writing. All cognitive blocks also contained subject-related background questions. The booklets were administered without audiotape and were spiraled through the assessment session.

Bridge to 1986 for Ages 9 and 13. Three booklets were used at each of the ages 9 and 13 and were identical to booklets administered in 1986. Each booklet contained one block each of reading, mathematics, and science items in addition to a common background block. All cognitive blocks also contained subject-related background questions. In each session, all students were administered the same one of the three booklets. The mathematics and science blocks were paced with an audiotape; the reading blocks were administered without an audiotape.

Bridges to 1986 for Grade 11/Age 17. Seven assessment booklets were created for administration to grade 11/age 17 students. One booklet consisted entirely of blocks of U.S. history items from the 1986 assessment and was administered to the 17[Br86-His] sample. The remaining six booklets consisted of blocks of reading, mathematics, and science items, were identical to booklets administered in 1986, and were administered to the 17[Br86-RMS] sample. The booklets in both bridges were administered without audiotape pacing. All seven booklets from both bridges were administered to students in the same assessment session by spiraling through the session.

1.4.3 Other Instruments

Besides the student assessment booklets, three other instruments provided data relating to the assessment:

Teacher questionnaires were administered to the reading teachers of fourth-grade students assessed for reading and to the writing teachers of eighth-grade students assessed for writing in the main assessment. These questionnaires were designed to gather information about the characteristics of the teachers of the assessed students and about the curricula and teaching methods in the classroom.

School characteristics and policies questionnaires were completed by school principals or a representative and provided information about school administration, staffing patterns, special programs, subject requirements, and school resources.

Excluded student questionnaires were completed by school personnel for each sampled student excluded from the assessment and provided information about the reasons for exclusion as well as basic demographic characteristics of the student.

1.5 FIELD ADMINISTRATION

The field administration of the 1988 assessment was the responsibility of Westat and is documented in Chapter 5. The field operation was organized around a staff at Westat's home office and a larger staff in the field. The home office staff consisted of a field director and a number of assistant field directors who were responsible for coordinating all activities of the Westat home office staff related to field operations and coordinating materials distribution and home-office receipt of assessment reporting forms. The field staff consisted of district supervisors and exercise administrators. The district supervisors, who were trained by Wes at, were each responsible for the assessment activities in one or more PSUs. Each district supervisor was primarily responsible for follow-up contacts to the school districts containing schools selected for assessment (as explained below, ETS made initial contact), recruiting and training exercise administrators to work with them in administering the assessment sessions, making arrangements for the assessments, and selecting the sample of students to be assessed within each school. The district supervisors and the exercise administrators administered the assessments, filled out the necessary forms, performed process control, and shipped the assessment booklets and forms to ETS.

Gaining school cooperation was the responsibility of both Westat and ETS staff. ETS made the preliminary contacts preparatory to obtaining school cooperation by first contacting the Chief State School Officers, informing them that schools within their states had been selected for the assessment and, in a later letter, listing the selected schools and districts. Later mailings were sent to superintendents of public schools and parochial schools and principals of private schools for all schools selected in the assessment. These materials provided an explanation of NAEP, a list of the selected schools in the official's jurisdiction, and a cover letter explaining that a Westat district supervisor would contact them to set up an introductory meeting. Westat district supervisors then scheduled and conducted introductory meetings, worked with the schools to schedule the assessments, and, with the exercise administrators, conducted the assessments. The overall cooperation rate of schools originally selected for all phases of the 1988 assessment was 86.7 percent. Further detail on school participation rates before and after substitution is given in sections 3.2 and 3.3 of Chapter 3.

The main assessment samples were assessed between January 4 and May 18, 1988 at all grade/age levels. The winter portion of the main assessment and the grade 4/age 9 bridge samples were assessed between January 4 and March 11; the spring portion of the main assessment and the grade 11/age

17 bridge samples were assessed between March 14 and May 18, 1988. The grade 8/age 13 bridge sample was assessed between October 12 and December 16, 1987.

Both Westat and ETS participated in the quality control of the field administration. The quality control involved on-site visits by Westat and ETS staff to verify the sampling of the students and to observe the conduct of the assessment by the field supervisors and the exercise administrators. At the end of the assessment, a telephone survey of a 10 percent sample of the assessed schools was conducted to evaluate the field procedures.

1.6 MATERIALS PROCESSING AND DATABASE CREATION

Upon completion of an assessment session, Westat field supervisors and exercise administrators shipped the assessment booklets and forms from the field to ETS for entry into computer files, checking, and forming the database. Careful checking assured that all data from the field were received. More than 130,000 booklets or questionnaires were received and processed. The extensive processing of these data is detailed in Chapters 6 and 6.1.

Items requiring a written response from the student (open-ended items) were included in the assessment instruments for every subject area except geography. Such items had to be professionally scored; the details of the professional scoring appear in Chapter 6.2.

The transcription of the student data into machine-readable form was accomplished by scanning the student instruments with an optical scanning machine. An intelligent data entry system was used for resolution of the scanned data, entry of documents rejected by the scanning machine, and entry of the information on the questionnaires. Additionally, each input datum was checked to verify that it was of an acceptable type, that it was within a specified range or set of ranges of values, and that it was consistent with other data values. The entry and editing of materials is discussed in Chapters 6.3 and 6.4.

Chapter 6.5 discusses the extensive quality control checks undertaken to verify that the information in the database corresponded to the information in the assessment booklets and questionnaires. A random sample of each booklet and questionnaire was selected from the database and compared with the original document. The database was determined to be quite free of errors (with an observed error rate of less than .001).

The final step in the preparation of the assessment data for analysis was the construction of the database and the public-use data tapes. These are described in Chapters 6.6 and 6.7.

Chapter 2

DEVELOPING THE NAEP OBJECTIVES, ITEMS, AND BACKGROUND QUESTIONS FOR THE 1988 ASSESSMENTS OF READING, WRITING, CIVICS, U.S. HISTORY, AND GEOGRAPHY

Walter B. MacDonald, Ina V. S. Mullis, Anne Campbell, and Nancy A. Mead

Educational Testing Service

The subject areas constituting the 1988 assessment were specified by a combination of the legislation in place when development work began and the advice of NAEP's governing board (at that time the Assessment Policy Committee).[1] A brief rationale for the inclusion of each subject area follows.

Reading. Over the past few years, there had been well-warranted concern that an alarming number of American students were underprepared for the literacy demands of an increasingly complex world. Concern had arisen not only for student's intellectual and economic well-being, but also for their ability to participate fully in an information-based society. In addition, because reading is central to proficiency in other subject areas, levels of reading proficiency are likely to serve as an overall barometer of educational progress.

Writing. Good writing is an essential underpinning in students' abilities to express ideas clearly. The importance of assessing writing was exhibited in the NAEP legislation effective through 1988 that required that at least once every five years NAEP collect and report data assessing the writing performance of students at various age or grade levels.

Civics. The civics assessment took place during the bicentennial of the writing of the U.S. Constitution. This was a time of high civic awareness and responsibility which focused attention on students' needs to understand democratic principles in order to appreciate and exercise their rights and to recognize the responsibilities inherent in being a U.S. citizen. Civics was first measured in the context of citizenship assessments in 1970, 1976, and 1982. The Assessment Policy Committee selected civics because of its

---

[1]Although reading, writing, civics, and U.S. history were the four subject areas approved for the 1988 assessment by the Assessment Policy Committee at their October 18-19, 1985 meeting, the committee also authorized NAEP to seek funds for a geography assessment. Thus, the fifth subject, geography, was added later when funding was obtained from the National Geographic Society to help support an assessment of high-school seniors.

importance and because of the time elapsed since it had last been assessed. The 1988 civics assessment represents current trends in civics education and reflects the civics-related portions of past citizenship assessments.

U.S. History. History plays a major role in helping students understand the world in which they live. The Assessment Policy Committee included U.S. history in the 1988 assessment to measure American students' familiarity with the basic timeline and significance of the main issues of U.S. history and the key roles American men and women played in effecting social, political, and economic change.

Geography. The extent to which students understand geography affects their ability to comprehend global relationships and preserve the world's valuable resources. An understanding of geography is also pivotal to the successful study of related disciplines, including history and science. To measure and report on the extent and quality of students' understanding of geography, the National Geographic Society provided support for the inclusion of geography in the 1988 assessment.

From its inception, NAEP has developed assessments through a consensus process and the 1988 assessment was no exception. Educators, scholars, and citizens representative of many diverse constituencies and points of view designed objectives for each of the five subject areas, proposing goals they felt students should achieve in the course of their education. After careful reviews, the objectives were given to item writers, who developed assessment questions appropriate to the objectives. All questions underwent extensive reviews by subject-matter, measurement, and bias/sensitivity specialists. They were assembled and printed into booklets suitable for matrix sampling and then administered by a trained field staff to a stratified, multistage probability sample of students.

The development for the 1988 assessment included questionnaires for students, teachers, and school administrators, as well as a substantial number of cognitive questions for each of the five subject areas.

All 1988 development efforts were governed by four major considerations.

1) As specified in the legislation, the objectives would be developed through a consensus process involving subject matter experts, school administrators, teachers, and parents, and the items would be carefully reviewed for potential bias.

2) As outlined in the ETS proposal for the administration of the NAEP grant, the development of objectives and items for each subject area would be guided by a Learning Area Committee.

3) As described in the *ETS Standards for Quality and Fairness* (ETS, 1987), all materials developed at ETS must be in compliance with specified procedures.

4) All NAEP items must be submitted to a complex Office of Management and Budget (OMB) clearance process and all publications, including objectives booklets, submitted for review by the Office of Educational Research and Improvement (OERI).

In general, developing the objectives and items for the 1988 assessment was a two-year process, beginning in October 1985 when the governing board determined the subject areas to be assessed and ending in October 1987 when data collection began at age 13 for the first trend assessment materials. The schedule called for selecting the Learning Area Committees in the fall of 1987 and beginning objectives development in January 1986. Once the frameworks for the objectives were set, item development proceeded in earnest from July through November 1986 when the clearance materials for the field test were submitted to the U.S. Department of Education. The field testing was conducted in February 1987. Subsequent to the field testing, the Learning Area Committees met to guide selection of the materials for inclusion in the 1988 assessment. The materials for the 1988 assessment were submitted for clearance in May 1987. The objectives booklets were prepared for publication, printed, and disseminated during 1987. It should be noted that the specially funded geography assessment of high school seniors was on an abbreviated schedule, beginning in January 1987, field tested in April 1987, and catching up to the main portion of the assessment in time to be administered beginning in January 1988 together with the reading, writing, civics, and U.S. history materials. The geography objectives booklet was printed in June 1988.

The sections that follow in this chapter include general overviews for setting objectives and developing items and specific details about developing the objectives and the assessments. Included in Appendix A is a list of the more than 400 consultants who participated in the 1988 development process.

2.1 GENERAL OVERVIEW OF PROCEDURES FOR SETTING OBJECTIVES

The general procedures followed for determining the objectives to be measured in each subject area in 1988 were essentially those followed by NAEP in previous assessments (see sections 2.3—2.6 for more detail).

1) The objectives used in the previous assessment were mailed to about 25 specialists for their review, comments, and suggestions. No constraints were placed on this activity and we asked for candid, critical reactions. The individuals involved in this process tended to be educators and specialists in the field and were selected to represent differing points of view, geographical locations, backgrounds, and constituencies. We sought advice from a wide range of sources for recommendations for this activity.

2) Learning Area Committees were established to help guide assessment development procedures within subject areas. The members of each committee were selected on the basis of recommendations from professional organizations, including those related to the specific subject areas, and with great care to represent differing perspectives and backgrounds. In addition to gender and race/ethnicity, care was taken to have representation from classroom teachers, district administrators, and state education agencies as well as researchers and university professors. As much as possible, attention was also given to selecting members familiar with the concerns of the professional organizations appropriate for each of the five subject areas. These committee members worked closely with NAEP staff in developing the 1988 assessment.

3) Comments from the initial objectives review were synthesized and used as input for the first Learning Area Committee meetings.

4) The first assignment of the respective Learning Area Committees was to review and revise the objectives booklets for each subject area based on their professional expertise and the comments of the previous reviewers.

5) The new edition of the objectives was, in turn, mailed to teachers, curriculum specialists (including the state curriculum supervisor f each subject area in each of the 50 states), and school admin. trators practitioners from around the country. These individuals are school administrators and teachers, as well as teacher trainers who live and work in the practical educational environments. Their task was to review these objectives from the point of view of what seems reasonable and practical. As with earlier steps in the objectives development process, care was taken to be certain that appropriate minority group representatives were included to assure proper attention to these sensitivities. Depending upon the results of that review, the objectives were redrafted with the participation of the Learning Area Committee members and others, as necessary.

6) The revised objectives were mailed to the Department of Education for comment and review.

7) Further modifications of the objectives were made as necessary.

8) The Learning Area Committees completed the final review of the product.

9) The objectives were published, printed, and made available for national distribution.

2.2 GENERAL OVERVIEW OF PROCEDURES FOR DEVELOPING THE ITEMS

A carefully developed and tested series of steps, essentially those followed in the past by NAEP, were used to create test items that reflected the objectives and that measured achievements related to them (see also sections 2.3—2.6).

1) Each Learning Area Committee and the staff of NAEP determined what specific aspects of the objectives could be measured given the realistic constraints of resources and the feasibility of implementing the measurement technology. For example, the grant from the U.S. Department of Education set a limit on the financial resources available as well as delineated the general design of the assessment as outlined in Chapter 1. Thus, the measures were constrained to those that could be administered via paper and pencil technology to groups of students. Each respective committee made recommendations about priorities for the assessment and types of items to be developed.

2) The existing pool of items to be used to measure change from previous assessments (trend items) was reviewed in detail.

3) Item specifications were then developed and prototype items were created to reflect the type of questions that had been suggested. Trend items were selected.

4) Item writers with skills and experience in creating items according to specifications were identified from both inside and beyond ETS and scheduled for item development tasks.

5) Newly created items were reviewed and revised by staff and external reviewers.

6) Further language editing and sensitivity reviews were conducted according to ETS quality control procedures.

7) Field test materials were prepared, including the materials necessary to secure OMB clearance.

8) The field test was conducted with a representative group of students from across the country.

9) Field test booklets were scored and the results analyzed.

10) Based on these analyses and the results of the pilot testing, items were revised or modified and re-edited. They once again went through an ETS sensitivity review.

11) With the help of staff and outside reviewers, the Learning Area Committee selected the items to include in the assessment.

12) Items were assembled into "blocks" (15-minute mini-tests) with attention given to balancing content coverage and difficulty levels.

13) After a final review and check to assure that each assessment booklet and each block therein met the overall guidelines for the assessment, the booklets were typeset and printed.

2.3 DEVELOPING THE READING ASSESSMENT

2.3.1 Reading Objectives

The objectives for the 1988 reading assessment[2] carried forward the view expressed in the 1984 reading objectives that the processes of comprehension and the extension of comprehension through interpretation and analysis have a place in reading of all kinds of texts. The 1988 objectives were based on an interactive view of reading. In this view, reading is a dynamic process in which a number of elements interact, including the reader, the material being read, the purposes of the reading, the reader's previous experiences, and the context for reading.

The first objective, "Comprehends What Is Read," is central to the reading process. This objective encompasses the comprehension of various types of written materials as well as the comprehension of materials read for a particular purpose. Objective two, "Extends Comprehension," involves deliberate kinds of analysis, interpretation, and evaluation. Objective three, "Manages the Reading Experience," recognizes that good readers develop a variety of strategies to help them comprehend what they read. These strategies include using the structure and organization of the text; using readers' aids; showing flexibility in one's approach to reading; and selecting reading materials appropriate to the purpose. The fourth objective, "Values Reading," discusses the kinds of appreciation that students can develop. The values include reading as a source of enjoyment; reading to expand understanding and to fulfill personal goals; reading as a means of acquiring knowledge and learning new skills; and recognizing the cultural role of written language. In order to help teachers using the objectives, instructional strategies for each of the objectives were included in the booklet. In addition, because the 1984 reading assessment was so extensive in content coverage and because the proficiency scale developed from it forms the basis for reporting data through the 1980s, a brief description of the reading scale was included.

---

[2] *Reading Objectives, 1986 and 1988 Assessments* (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, June 1987).

2.3.2 Reading Materials

As with previous NAEP reading assessments, a variety of reading stimulus materials were used that reflected the realities of reading passages, charts, and instructions found in texts, newspapers, and source documents. To this end, materials were drawn from many sources, including science and social studies textbooks, peer writing, technical writing, and directions.

2.3.3 Reading Item Development

The main goals of the 1988 reading assessment were to provide a link to the 1984 reading assessment and to broaden the scope of the assessment with some new items. As a result, the items from the 1984 assessment became the foundation for the 1988 assessment and the extensive development of new items was determined to be unnecessary. New items were developed under guidelines that included such factors as type of material (narrative, expository, or peer writing), number of features being matched, number of possible distractors, and correspondence between the question and text. These guidelines were helpful in assembling the items for the field test, particularly in determining an approximate difficulty level for the items. In assembling the items into blocks for the field test, factors that were taken into consideration included the length of the stimulus, the type of reading material, and the estimated difficulty of the items.

Once the field test data had been analyzed, the reading Learning Area Committee met to select trend items from the 1984 reading assessment and to select new items that would complement the pool of trend items. Trend items were selected with careful attention to their psychometric characteristics as evaluated in prior assessments. The new items were selected on the basis of their statistical item analyses from the field test as well as the type of stimulus material. One factor that had to be taken into consideration in compiling the items into blocks was that at least one trend and one new block at each age level had to overlap with the age level above and/or below it. There were to be at least three blocks of trend items at each age level as well. In addition to the criteria for overlap across age levels, trend blocks were selected to represent the best content coverage across the objectives and be representative of the range of performance in previous assessments.

Once the items were selected and the blocks assembled, they underwent final review by ETS subject-area specialists and test editors as well as a review to detect the presence of any bias according to the *ETS Standards for Quality and Fairness* (ETS, 1987). After internal review, the OMB clearance package was prepared and submitted.

2.3.4 Reading Background and Attitude Questions

The 1984 and 1986 assessment provided a wide range of reading background questions for consideration for the 1988 assessment. The Learning Area Committee reviewed this pool of items and selected a number for reuse. Of continuing interest were items that measure student behaviors that indicate interest in reading (reading for fun in your spare time, telling a friend about a good book). Other questions from prior years included spare time spent reading fiction and nonfiction and the student's evaluation of his or her own competence as a reader.

The Learning Area Committee was interested in expanding questions about the home reading environment: whether students subscribe to magazines at home or have books of their own, whether students were read to when they were young, whether students read aloud to someone at home. Also added to the assessment were questions about frequency of various teaching practices—introducing the main idea of a story, pointing out new vocabulary, giving students a list of questions to answer. Finally, the Learning Area Committee wanted to know how often students used various reference books, such as dictionaries and encyclopedias.

2.4 DEVELOPING THE WRITING ASSESSMENT

2.4.1 Writing Objectives

The objectives for the 1988 writing assessment[3] were essentially a revised update of the objectives developed for the 1984 writing assessment. Because a major development effort was expended in the preparation of the 1984 objectives, NAEP anticipated that those objectives still reflected current theory of writing. To make sure, however, that the objectives for the 1988 assessment were up-to-date, NAEP sent the 1984 objectives to teachers and theorists across the nation and asked them to comment on and revise the objectives as they felt necessary. Although many reviewers suggested numerous small revisions, the reviews supported the hypothesis that extensive revisions were unnecessary. The one major concern was integrating the concept "learning through writing" across all writing purposes—informative, persuasive, and personal/imaginative narrative—rather than segregating it as a single objective.

NAEP staff collated the comments and prepared a revised version of the objectives. This revised version was sent to members of the Learning Area Committee prior to their first meeting. When the committee met, they discussed the reviews and the revised objectives and then each member took a section and revised it further to reflect the discussion. Their revisions became the basis for the final version of the objectives.

The writing objectives were based on the premise that individuals write for a purpose to an audience. Reflecting this premise, the first objective was that students use writing to accomplish a variety of purposes: informative, persuasive and, personal/narrative. The second objective—that students manage the writing process—focused on the importance of the process that leads to a piece of writing. The third objective—that students control the forms of written language—concerned itself with such skills as

[3] *Writing Objectives, 1988 Assessment* (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, August 1987).

organizing, elaborating, and using appropriately the conventions of writing (usage and mechanics). The fourth objective was that students value writing and what has been written and so underscored the importance of understanding the value of writing and the roles written works serve in our society. In order to help people put the objectives into practice, ideas and suggestions for teaching to each objective were presented in the final section of the objectives booklet.

2.4.2 Writing Item Development

Because of the similarity between the 1984 and 1988 objectives and because a major effort had also been put forth to develop the tasks for the 1984 writing assessment, NAEP felt that those items could be the foundation for the 1988 assessment. In addition, NAEP had access to the many tasks being developed for the California Assessment Program (CAP). For these reasons, it was possible to keep the new item development effort to a minimum.

First, the Learning Area Committee, made up of seven writing specialists, evaluated and selected the 1984 items by taking into consideration such factors as data from the 1984 writing assessment, the purpose for writing addressed by the task, the appropriateness of the items for more than one age, and the need to use enough items to provide a bridge to the 1984 assessment. The items that the Learning Area Committee selected became the pool for the writing trend bridge assessment and were the foundation of the pool of items for the more comprehensive 1988 main assessment. After selecting the 1984 trend items, the committee then considered what other items were needed to meet the specifications for the 1988 main assessment, reviewed the materials from the CAP, and developed new items. The developers of the CAP writing assessment program, which is built on a theoretical base very similar to that used by NAEP, had worked very long and hard to develop over 80 grade 8 prompts and were pleased to share materials with the NAEP committee. We are particularly grateful for access to this resource and indebted to CAP. In developing the new prompts for the 1988 assessment, particular attention was given to making the writing tasks relevant to students at the grade level being assessed. Consistent with the growing emphasis on process-writing instruction, it was also determined to study the effect on performance of the length of time provided to respond to the writing tasks. A small study on this issue conducted in conjunction with the South Carolina Department of Education had indicated some improvements in performance. Therefore, some tasks were identified to be given to students in two versions. In the regular version, fourth-grade students had 10 minutes to respond and eighth- and twelfth-grade students had 15 minutes to respond. In addition, some of the tasks were also given with longer response times—20 minutes for fourth graders and 30 minutes for students in the upper grades. (Because any assessment context seriously restricts provision for the usual revision strategies, including peer review and outside consulting resources, in either situation the responses were to be viewed as first draft writing and evaluated accordingly.)

Once the pool of prompts had been developed, these items were reviewed by ETS subject-area specialists, sensitivity reviewers, and test editors.

Subsequent to the internal review procedures, NAEP staff submitted the items for OERI and OMB clearance prior to field testing.

After the field test, the Learning Area Committee met once again to review the results and to select the new items that would complete the item pool for the 1988 assessment. Those items underwent final review by ETS subject-area specialists and test editors as well as a review to detect any bias according to the *ETS Standards for Quality and Fairness* (ETS, 1987). The final item pool, including the trend items, was then submitted for OMB clearance for the 1988 assessment.

2.4.3 Writing Background and Attitude Questions

For reasons explained in Chapter 1, the overall NAEP design between 1984 and 1988 changed from one where student background questions were BIB-spiraled across the assessment along with cognitive items to one where all students assessed in writing at a given grade level were given the same five-minute writing background questionnaire. Again, as the result of the extensive development effort for the 1984 writing assessment and also as a result of the design used in 1984, that assessment had contained over 100 background questions and all had been thoroughly analyzed and reported in *The Writing Report Card: Writing Achievement in American Schools* (Applebee, Langer, & Mullis, 1986). This pool provided an ample source of questions for devising a five-minute background questionnaire for each of the three grade levels (10 questions at grade 4, 24 questions at grade 8, and 35 questions at grade 12).

The Learning Area Committee reviewed the questions and the 1984 results and determined which areas to carry forward into the 1988 assessment. The reduction in background coverage for 1988 necessitated difficult decisions, but agreement was reached about the need to focus on information about students' use of and instruction in the writing process as well as the amount of writing actually done by students.

The latter two areas of emphasis were particularly important in view of the teacher questionnaire that was developed for the writing teachers of the eighth-grade students who participated in the assessment (see section 2.7 for further detail about the teacher questionnaires). While both students and teachers were asked some questions relevant to their own experiences, care also was given to obtaining a coordinated set of information from teachers and students. In this way, NAEP was able to obtain the perspectives of both teachers and students about writing instruction and compare the two (see *Learning to Write in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Applebee, Langer, Mullis, Jenkins, & Foertsch, 1990).

2.5 DEVELOPING THE CIVICS AND U.S. HISTORY ASSESSMENT

Because state education agencies had expressed a desire to have greater input into the NAEP assessment development process so that state curriculum concerns could be adequately reflected in the discussions about the objectives and the content of the assessment, the civics and U.S. history assessments were developed together using a new model—one Learning Area Committee with the additional support of a State Advisory Committee. The integrated development of the civics and U.S. history would also be a very efficient protocol for development by decreasing the redundancy of particular aspects of these related subject areas. This model for development was accepted by OERI and ratified by the Assessment Policy Committee at their May 30-31, 1986 meeting. In November 1985, NAEP staff solicited recommendations for reviewers and Learning Area Committee membership from the 50 State Testing Directors.

To enhance state participation in the actual development of the civics and U.S. history assessment, a plan was developed with the State Testing Directors at the Large Scale Assessment Conference held in Boulder, Colorado, June 9-12, 1986. The plan established a State Advisory Committee to work in conjunction with the Learning Area Committee.

2.5.1 Civics and U.S. History Objectives

The eight-member civics and U.S. history Learning Area Committee first met in July 1986 to draft the objectives and suggest prototype items. Unlike reading and writing with ongoing development and extensive efforts in recent assessments, civics had not been assessed since 1976 (as part of citizenship) and although U.S. history had been assessed in 1986, that assessment was a knowledge-based assessment at only grade 11/age 17. Thus, more initial work was needed to update and elaborate on the existing objectives for civics and U.S. history.

For civics, a three-dimensional matrix made up of content, context, and cognition frame three broad objectives.[4] The content dimension included the democratic principles and the purpose of government; the political institutions; the political process; and rights, responsibilities, and the law. Each of the areas were further broken down into numerous assessment topics. The context dimension included home, school, community, state, nation, and world. These addressed the expectations that civics learning begins in childhood, continues through adolescence and schooling, and matures in adulthood when people are participating in society. The cognition dimension included 1) knows and 2) understands and applies. These abilities addressed the notions that students must first be able to recognize factual knowledge and then be able to interpret information and be aware of how concepts and facts are interrelated.

As with the civics objectives, the U.S. history objectives[5] took form as a matrix framed by three broad objectives. The first, chronology of events, persons, and documents, included the following eight historical periods:

---

[4]*Civics Objectives, 1988 Assessment* (Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress, September 1987).

[5]*U.S. History Objectives, 1988 Assessment* (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, August 1987).

1) Exploration and colonization up to 1763; 2) The revolutionary era, the Constitution, and the new Republic, 1763-1815; 3) Economic and social development of the Antebellum Republic, 1790-1861; 4) Crisis of the union: origins of the war, the war, and reconstruction, 1850-1877; 5) The rise of modern America and World War I, 1877-1920; 6) The United States, 1920-1941; 7) World War II and the postwar era, 1931-1968; and 8) Modern post-industrial era: 1968 to the present. Additional topics were specified under each of these chronological periods. The second objective, historical contexts, included the following contextual backgrounds. political life; economic life, cultural, social, and family life; and intellectual life. Reasoning skills formed the third dimension of the framework. The first skill, reference skills and knowledge, addressed the need to know historical aspects and the need to know how to expand an existing knowledge base. The second skill addressed the need for an understanding and a comprehension of the association of ideas and the perception of relationships.

In September 1986, the State Advisory Committee was convened to review these draft objectives and to review items submitted by states for inclusion in the civics and U.S. history assessment. For both the civics and U.S. history objectives, the committee suggested minor improvements to revise or rephrase some of the topic areas. Following the review of the objectives, the committee weighted each of the topic area with consideration to their relevance at a given grade/age level. The revised objectives were subsequently reviewed by the Learning Area Committee who were, by and large, satisfied with the results.

2.5.2 Civics and U.S. History Item Development

Some of the new items used in the 1988 assessment were submitted by some of the states; others were developed by either ETS staff or outside item writers. To facilitate the writing of new items that were appropriate to the curriculum, 1? social studies teachers were convened at an item development conference at ETS to write and revise new items. Following this conference, additional input was obtained from another 15 social studies teachers who reviewed items and suggested other new items during an item development conference at the Social Science Education Consortium in Boulder. The item development process proceeded through the fall of 1986 in accordance with the guidance provided by the Learning Area Committee. The new items were reviewed by subject-matter specialists, edited, and presented to the committee. The committee met in October 1986 to review, revise, and select items for field testing in the spring. To ensure the appropriateness of the items selected for students in the fourth grade, these items were sent to eight fourth-grade teachers for review in November 1986. All materials were submitted to OERI in November 1986, and revised and resubmitted in December to OERI for submission to OMB. After field testing, the Learning Area Committee met for a third time, in April 1987, to review the field test results and select the items for the 1988 assessment of civics and U.S. history. Consistent with routine procedures, the items selected for the 1988 assessment were submitted for internal ETS review and submitted for OERI/OMB clearance.

2.5.3 Civics and U.S. History Background and Attitude Questions

The civics and U.S. history Learning Area Committee was primarily concerned with gathering information about students' instructional experiences. At all grade/age levels, students were asked how much they had studied the various topics covered in the assessment. At the two higher grade/age levels, students were asked about the amount of course work they had taken and their teachers' practices. High-school students also were asked about their homework, their grades, and their attitudes toward the subject areas.

2.6 DEVELOPING THE GEOGRAPHY ASSESSMENT

The 1988 geography assessment was initiated and funded by the National Geographic Society due to well-warranted concerns about the state of geography learning in the United States. The development of this assessment did not start at the same time as the other subjects. The design, planning, and funding for a geography assessment of grade 12/age 17 students commenced man months later in NAEP's 1988 development cycle.

A five-member Learning Area Committee was created with individuals recommended by the National Geographic Society in accordance with the guidelines for committee membership articulated in section 2.1. This committee eagerly took the charge to develop an assessment in a new subject area never before assessed by NAEP. The challenge of a late start and a brand new subject area for assessment was balanced by the modest scope of this assessment relative to the other subject areas—one grade and three blocks of questions versus three grades with approximately seven blocks of questions per grade.

2.6.1 Geography Objectives

The geography Learning Area Committee first convened at the end of January 1987. At this meeting, the objectives of the 1988 geography assessment[6] were formulated to reflect the current trends in geography education. During this process, the Wisconsin geography assessment provided valuable background information upon which to build NAEP's draft framework for a national assessment of geography. This framework was organized around three dimensions of geography: geographic skills and tools, including the use of maps, charts, and globes; geographic knowledge and concepts, including the understanding the area of physical and cultural geography; and geographic inquiry, including the cognitive application of skills, knowledge, and understanding to new situations. The dimension of geographic knowledge and concepts (content) was divided into physical and cultural geography. Physical geography included physical locations, places, and regions; climatology and meteorology; and the evolution of land-form features of the earth's surface,

[6] *Geography Objectives, 1988 Assessment* (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, June 1988).

whereas cultural geography included cultural locations, places and regions, human impacts on the environment; influences of the environment on human activity; and spatial interactions. All of the categories within physical and cultural geography were broken down into numerous subcategories to help characterize topics for assessment

Following the Learning Area Committee meeting, the draft objectives were reviewed by external consultants representing various constituencies and revised as necessary. Because of the hastened development time and also because of their keen interest, the Learning Area Committee remained closely involved throughout the review and revision process.

2.6.2 Geography Item Development

Following the drafting of objectives, item development proceeded at an accelerated pace in order to bring the geography assessment into phase with the assessments in reading, writing, civics, and U.S. history. About 100 cognitive items were written by ETS test development staff and 100 items were written by members of the Learning Area Committee. These items were reviewed by about 10 reviewers and used to assemble six blocks of cognitive items. The rapid preparation did not preclude the standard ETS test development procedures to ensure quality and fairness of the geography item blocks.

The geography field test data were gathered and analyzed in preparation for a May 1987 meeting of the Learning Area Committee. At this meeting, the committee reviewed the items and their statistics to select a pool of items that would be appropriate for the assessment of the geography knowledge and skills possessed by students at grade 12/age 17. Decisions were monitored in terms of the coverage of both the four content areas and the importance of assessing conceptual understanding as well as facts. After the meeting, test developers assembled one test booklet comprising one five-minute block of general background questions, one five-minute block of geography background questions, and three 15-minute blocks of cognitive items, totaling 78 items in all. These materials were reviewed thoroughly by ETS specialists and submitted for OMB clearance.

2.6.3 Geography Background and Attitude Questions

As mentioned above, geography assessment development also included a block of background questions for inclusion into a test booklet with the cognitive blocks. In developing materials for the student questions specific to geography, the Learning Area Committee focused on coursework and content coverage. Students were asked what geography courses they had taken during high school and how much they studied the various topics covered by the objectives and assessment items.

2.7 DEVELOPING THE COMMON CORE STUDENT QUESTIONS AND QUESTIONNAIRES

Additional instruments were developed for the 1988 assessment: a common core of student background questions containing 21 questions at grade 4, 21 questions at grade 8, and 33 questions at grade 12; teacher questionnaires containing 56 questions at grade 4 and 67 questions at grade 8; school characteristics and policies questionnaires containing 91 questions at grade 4, 94 questions at grade 8, and 107 questions at grade 12; and an excluded student questionnaire containing 28 questions.

The student, teacher, and school instruments were designed to collect information about home, classroom, and school factors related to four policy issues that had also been the focus of the 1986 assessment: instructional practice, students at risk, teacher workforce, and effective schools.

The development of policy issues and items was an iterative process that involved staff work, field testing, periodic review by an external advisory group, and review by the Assessment Policy Committee. At an initial meeting, a group of external consultants drafted a set of policy issues and made recommendations regarding the design of the assessment. They were particularly interested in capitalizing on the unique properties of NAEP and not duplicating the purposes underlying other surveys (e.g., The National Survey of Public and Private School Teachers and Administrators and The National Education Longitudinal Study). They recommended a more focused study of the relationship between student achievement and instructional practices. This recommendation provided the impetus for a major redesign of the teacher questionnaire (see below). Items were developed to assess the policy issues and field tested with students, teachers, and principals. The policy issues, items, and field test results were reviewed by the group of external consultants who identified specific items to be included in the final questionnaires. The field test results and the recommendations of the consultants were also reviewed by the Assessment Policy Committee. The items were then assembled into questionnaires and submitted to internal ETS procedures to ensure fairness and quality. The background questionnaires were submitted for OMB clearance together with the cognitive items.

Every student booklet began with a common core of background questions. In many cases the questions used in 1988 were taken from prior assessments. Although many of the questions were common to the three grade/age levels assessed, some were specifically targeted to elementary or high-school students. At grade 4/age 9, the background questions were read aloud to the students and took approximately 10 minutes to complete. At the other two grade/age levels, only the first race/ethnicity question was read aloud, students read and answered the remaining questions on their own during a five minute time period. The questions asked about demographics, home environment, and instruction experiences.

The teacher questionnaire was administered to samples of grade 4 and grade 8 teachers. All grade 4 students assessed in reading were identified and questionnaires were given to their reading teachers (whoever took primary responsibility for the student's reading instruction). All grade 8 students assessed in writing were identified and questionnaires were given to their writing teachers.

The teacher questionnaire included three sections and took approximately 20 minutes to complete. The first section focused on individual student information. All teachers whose students had participated in the assessment were asked to respond to a set of questions about their students. Each teacher answered this set of questions for up to 10 different students. Most of the questions were new.

The second section focused on classroom information. Teachers were asked to respond to a set of questions about the classes in which the students in the assessment were enrolled. Each teacher answered this set of question for up to five different classes. Most of the questions in this section were new in 1988, although some used wording that is parallel to student questions from prior assessments.

The third section focused on teachers' background and experience. Almost all of these items were taken from prior assessments.

The school characteristics and policies questionnaire was given to the principal in each participating school, and took about 15 minutes to complete. Three versions were administered—one for each of the three grade/age levels. The questions asked about the principal's background and experience, school policies, programs, and facilities, and the composition and background of the student body.

The excluded student questionnaire was given to the teachers of students who were identified in the NAEP sample but were determined by the school to be unable to participate because they were mildly mentally retarded (educable), functionally disabled, or had limited English proficiency. This questionnaire took approximately three minutes per student to complete and asked about the nature of the student's exclusion and special programs in which the student participated.

2.8 FIELD TESTS FOR THE 1988 ASSESSMENT

By February 1987, staff had secured school cooperation for participation in field testing for the 1988 assessments of reading, writing, civics, and U.S. history and for the teacher and school policies questionnaires. By this time, staff had assembled 84 15-minute blocks of reading, writing, civics, and U.S. history cognitive items and combined these into 34 field test booklets—13 booklets at grade 4, 11 booklets at grade 8, and 10 booklets at grade 12. Upon receipt of clearance for the field testing, booklets were printed and approximately 30 trained administrators immediately began field testing. The field tests, involving 6,800 students in 52 school districts across the country, were conducted in February 1987. ETS staff members traveled to 12 districts in the Northeast, 17 in the Southeast, 12 in the Central, and 11 in the West. Generally, field tests were conducted at all three levels in each school district including two classes at fourth grade, and one class at each grades 8 and 12. Field tests were carried out in 104 fourth-grade classes, 88 eighth-grade classes, and 80 twelfth-grade classes. Approximately 300 responses were obtained to each question in the field test materials.

Overall the field test was completed on schedule. The data were collected, scored, and analyzed in preparation for meetings with the Learning Area Committees and Background Advisory Committee. Using item analysis, which provides the mean percentage of correct responses for each item in the field test, committee members and NAEP staff reviewed the materials according to five purposes: to determine which items were most related to achievement in the four subject areas; to evaluate the effectiveness of items designed specifically to assess higher-order thinking skills; to determine necessary revisions to items that lacked clarity, or to ineffective item formats; to give priority to items to be included in the full assessment; and to determine appropriate timing for assessment items.

The geography background questionnaire and the blocks of cognitive items were field tested in April 1987. Trained administrators conducted field tests at the twelfth grade at about 12 locations around the country.

Once the committees had selected the 1988 assessment items, all items were rechecked for content, measurement, and sensitivity concerns. The OMB clearance process was initiated April 28, 1987 with the submission of draft materials to the Office of Educational Research and Improvement. The final package was submitted May 27, 1987. Throughout the clearance process revision were made in accordance with changes required by OERI, the IMCD, and OMB.

2.9 FINAL PREPARATION OF THE 1988 ASSESSMENT MATERIALS

2.9.1 Objectives Booklets

The objectives booklets were sent out for external review by teachers, educators, and state representatives. The many good suggestions generated from these reviews were used to revise the booklets. After consensus review and approval by each Learning Area Committee, the objective booklets were sent to OERI for the internal government and peer review process. Subsequent to that 30-day review and incorporation of suggested revisions, the booklets were submitted to ETS internal editing and finally to the composition and printing process The new objectives booklets for reading, writing, civics, and U.S. history were published and released in 1987; the geography objectives booklet was published and released in 1988.

2.9.2 Student Assessment Booklets and Questionnaires

The items earmarked by each Learning Area Committee and submitted to OMB clearance for inclusion in the 1988 assessment were assembled into blocks in accordance with the assessment design. These blocks were assembled to meet content, context, cognition, and psychometric specifications and to conform to the assessment time and administration restrictions. A common core of questions about students' demographics and home backgrounds appeared at the beginning of each booklet. Approximately five minutes of background questions related to a subject area appeared next, followed by blocks of cognitive items in the same subject area. For reading, seven blocks of cognitive items were prepared for each grade/age level; for writing, 10 blocks of cognitive items were prepared for each grade/age level; for civics, three cognitive item blocks were prepared for grade 4/age 9 and 10 for each of the two older grade/age levels; for U.S. history, three blocks of cognitive items were prepared for grade 4/age 9 and seven for each of the two older grade/age levels; and for geography, three blocks of cognitive items were prepared for grade 12/age 17. The questions for school administrators, teachers, and about excluded students were assembled into questionnaires.

In addition to the main assessment special studies were included to measure the relationship between length of time permitted for writing and writing performance and the relationship between document literacy and reading proficiency. Further details about assessment instruments appear in Chapter 4.

Subsequent to assembly, all of the assessment blocks and questionnaires were subjected to the review process established by ETS and NAEP, including scrutiny by subject-matter specialists, measurement specialists, test editors, and persons specially trained to review questions for any potential insensitivity to women or minority groups. As part of the OMB clearance process, all items were also reviewed by OERI staff, by the Office for Quality Assurance, and by the OMB. Subsequent to final OMB clearance on August 11, 1987, the blocks and questionnaires were submitted for composition, printing, bundling, and distribution to the Westat, Inc., field staff responsible for administering the 1988 assessment.

Chapter 3

SAMPLE DESIGN

Keith F. Rust and Morris H. Hansen

Westat, Inc.

The samples for the 1988 NAEP assessment were selected using a complex multistage sample design involving the sampling of students from selected schools within 94 selected geographic areas, called primary sampling units (PSUs), across the United States.

The sample design had four stages of selection:

1) selection of geographic PSUs (counties or groups of counties);

2) selection of schools within PSUs;

3) assignment of session types to schools; and

4) selection of students for session types within schools.

The samples were drawn for the three different age classes, and for each age class the samples were of two distinct types. The first type consisted of the cross-sectional or "main" samples, while the second type consisted of the trend or "bridge" samples. The populations surveyed with each of these sample types are defined in Table 3-1. Separate samples of schools were required for the bridge samples and main samples, because of various differences in the calendar period for test administration, the format of the administration, and, in the case of age class 17, the grade and age definition of the population of interest.

Table 3-1
Reference Populations for the Components of
the NAEP 1988 Samples

Age Class	Main Samples	Bridge to 1986	Bridge to 1984 and Civics Bridge
9	Born 1978 and/or enrolled in grade 4	Born 1978 and/or enrolled in grade 4	Born 1978
13	Born 1974 and/or enrolled in grade 8	Born 1974 and/or enrolled in grade 8	Born 1974
17	Born 1970 and/or enrolled in grade 12	Born 10/70-9/71 and/or enrolled in grade 11	Born 10/70-9/71

In addition to representing the respective populations as a whole, the main samples had as a component a modest oversampling of schools with moderate or high enrollment of Black and/or Hispanic students (see section 3.2). This oversampling was undertaken to increase the sample sizes of such students, thus increasing somewhat the reliability of estimates for these minority groups.

The overall assessment period fell into three time periods - fall, winter, and spring. Not all assessment components were conducted in each time period. Table 3-2 shows the relationship between the various sample components and the assessment periods. The sizes of the PSU and school samples and the procedures for their selection were determined by the assessment period, as well as by the population to be surveyed and the method of administration in each case.

Table 3-2
Assessment Type by Age Class and Assessment Period

Age Class	Fall 10/12/87-12/18/87	Winter 1/4/88-3/11/88	Spring 3/14/88-5/1?/88
9	-	Cross-sectional (part) Bridges to 1984, 1986	Cross-sectional (part)
13	Bridges to 1984 1986, Civics Bridge	Cross-sectional (part)	Cross-sectional (part)
17	-	Cross-sectional (part)	Cross-sectional (part) Bridges to 1984, 1986, Civics Bridge

This chapter gives details of the sample selection procedure, and information on the results of the sampling process. Still fuller details are given in *The 1988 National Assessment of Educational Progress—Sampling and Weighting Procedures, Final Report* (Rust, Bethel, Burke, & Hansen, 1990).

3.1 PRIMARY SAMPLING UNITS

In the first stage of sampling, the United States (the 50 states and the District of Columbia) was divided into geographic primary sampling units (PSUs). Each PSU met a minimum size requirement (a population of at least 60,000 in the 1980 Census) and comprised either a metropolitan statistical area (MSA), a single county, or (more usually in the case of nonMSA PSUs) a group of contiguous counties. In the case of New England MSAs, which are not formed from whole counties, the corresponding New England County Metropolitan Areas, which are defined in terms of whole counties, were designated as the PSUs. The New York City MSA was divided along county/borough lines into three PSUs for reasons of administrative and sampling convenience. Each PSU was contained entirely within one of the four regions defined in Table 3-3. These regions were used to stratify the sample of PSUs, ensuring that each region was adequately represented in the various assessment samples.

Table 3-3
Geographic Regions Used for Stratification

Northeast	Southeast	Central	West
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia*	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia*			Utah
			Washington
			Wyoming

* That part of Virginia that is included in the Washington, DC, metropolitan statistical area is included in the Northeast region, the remainder of the state being included in the Southeast region.

In a few cases an MSA crossed region boundaries. Such MSAs were split into two or more PSUs as necessary (e.g., the Cincinnati OH-KY-IN MSA was split into the Cincinnati OH-IN PSU in Region 3 and the Cincinnati KY PSU in Region 2). Twelve subuniverses of PSUs were then defined as described below.

The 28 largest PSUs were included in the sample with certainty. An additional six very large PSUs (four from the Southeast and two from the West) that had large proportions of Black students and/or Hispanic students were also included with certainty. The 34 certainty PSUs constituted 32 metropolitan areas, since the New York City MSA was divided into three certainty PSUs. The inclusion of these 34 PSUs in the sample with certainty provided an approximately optimum cost-efficient sample of schools and students when samples were drawn within them at the required national sampling rate. The representativeness of the sample for minority groups was enhanced by ensuring that these PSUs were included in the sample, since these minority groups are relatively heavily represented within these certainty PSUs. The remaining smaller PSUs were not guaranteed to be selected for the sample. These were grouped into a number of noncertainty strata (so called because the PSUs in these strata were not included in the sample with certainty), and a sample PSU was selected from each stratum.

The PSUs were classified into four regions, each containing about one-fourth of the U.S. population. These regions were defined primarily by state (Table 3-3). In each region, noncertainty PSUs were classified as MSA or nonMSA. In the Southeast and West regions, the PSUs in which the combined proportion of population which were Black and Hispanic respectively in the 1980 Census exceeded 20 percent, were classified as high minority. The resulting major strata, or subuniverses, are shown in Table 3-4.

Table 3-4
The Sampling Subuniverses
and the Number of Noncertainty Strata in Each

	MSA PSUs		NonMSA PSUs	
Region	Regular Strata	High-minority Strata	Regular Strata	High-minority Strata
Northeast	8	—	2	—
Southeast	4	6	4	6
Central	8	—	6	—
West	4	6	4	2
Total	24	12	16	8

Within each major stratum (subuniverse), further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics, yielding 60 strata. The strata were defined so that the aggregate of the measures of size of the PSUs in a stratum was approximately equal for each stratum, except for strata in the high-minority subuniverses, in which the aggregate was approximately half that of the regular strata. The size measure used was the population from the 1980 Census. The characteristics used to define strata were the percent minority population, the percentage change in total population since 1970, the per capita educational expenditure, and the percent of persons employed in manufacturing (MSA subuniverses only) and the percentages of rural and urban dwellers (nonMSA subuniverses only). One PSU was selected with probability proportional to size from each of the 60 noncertainty strata. That is, within each stratum, a PSU's probability of being the sample selection from that stratum was proportional to its population. Thus the high-minority subuniverses were sampled at approximately twice the rate of the other subuniverses, since they were about half as large. This procedure of oversampling from the high-minority subuniverses was used with the aim of reducing somewhat the level of sampling error for estimates relating to the populations of Black and Hispanic students.

The final sample of 94 PSUs was drawn from a population of about 1,000 PSUs. Primarily because of the use of MSAs as PSUs, PSUs varied considerably as to their probability of selection, since they varied greatly in size. The 34 certainty PSUs consisted of the 26 largest MSAs in the country, based on the 1980 population from the Census, plus six other large MSAs from the Southeast and West regions with in excess of 20 percent of their population being Black or Hispanic. The 36 selected noncertainty MSA PSUs had probabilities of selection ranging from 0.028 to 0.584, while the 24 selected nonMSA PSUs had probabilities ranging from 0.021 to 0.101. The variations in probability depended upon the size of the PSU (1980 population) and whether or not the PSU was in a high minority subuniverse. Parts of 37 states were included in the 94 selected PSUs.

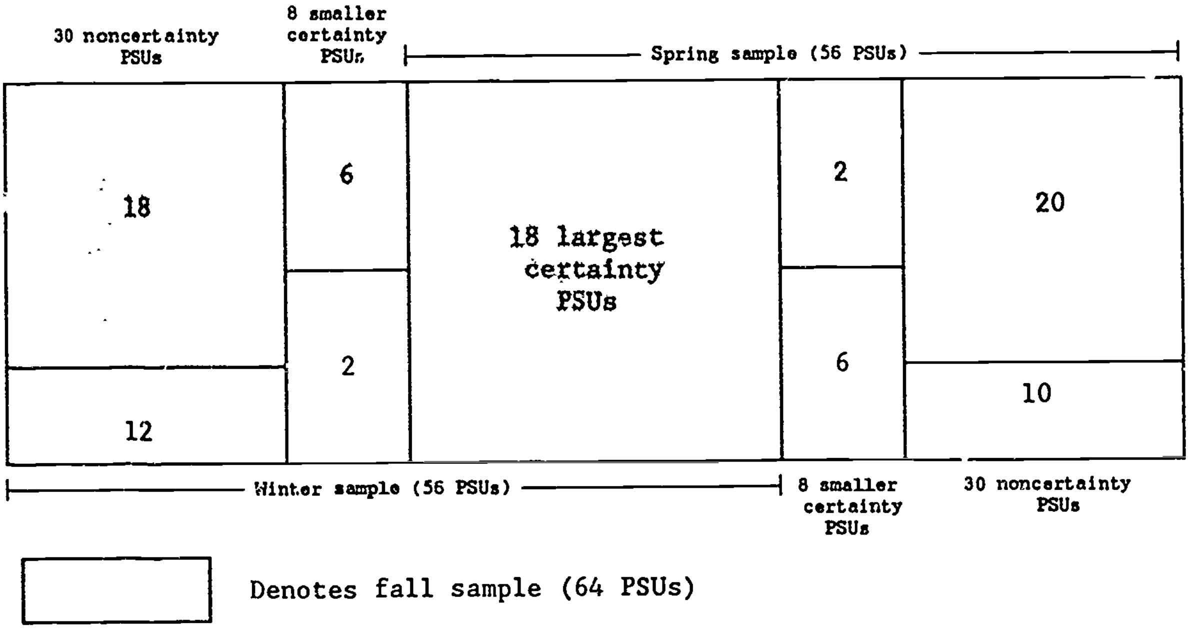
The PSU samples were drawn at one time for the 1986, 1988, 1990, and 1992 assessments. They were drawn so as to provide for the rotation of the PSUs from one assessment to the next, except that certainty PSUs were retained in each assessment year, and some of the larger noncertainty PSUs were retained for two successive assessment years.

The 94 PSUs were used for the main assessments of all three age classes. To facilitate and improve administration of the assessments and to allow for the estimation of growth in achievement during the school year, the assessment sample was divided into two randomly equivalent subsamples, one to be assessed in the winter and the other to be assessed in the spring. For this purpose, the 94 PSUs were designated as winter PSUs, spring PSUs, or both winter and spring PSUs, according to the following scheme. The 18 largest certainty PSUs were designated both winter and spring PSUs, to be included in the sample for both seasons (the sample of schools within each of these PSUs was divided into two stratified random half-samples; one half-sample to be assessed in the winter and one to be assessed in the spring). The 16 smaller certainty PSUs were ordered by region and then alternately designated as winter PSUs or spring PSUs, resulting in eight PSUs for each season. Similarly, alternate members of the set of the 60 noncertainty PSUs, arranged in stratum order within each subuniverse, were designated as winter or spring PSUs. The end result was 56 winter PSUs, including 38 in which assessments were conducted only during the winter, and 18 where assessments were conducted in the winter and the spring, and 56 spring PSUs, consisting of 38 in which assessments were conducted only in the spring and the 18 winter and spring PSUs.

The procedure of designating two half-samples for the main assessment, one for winter and one for spring, differed from the procedure used in 1986. In that year, all of the main (i.e., cross-sectional) sample assessments took place in the spring within the 94 selected PSUs; consequently, it was not necessary to designate half-samples on the basis of season of assessment.

The bridge assessments used a series of subsamples of the 94 PSUs used for the main assessment. The age class 9 bridge assessments, which were conducted in the winter, used the 56 PSUs designated as winter PSUs in the main assessment; the age class 17 bridge assessments, conducted in the spring, used the 56 PSUs designated as spring PSUs. The age class 13 bridge assessments, conducted in the fall, used 64 PSUs selected from the complete set of 94 PSUs with probability proportional to the measure of size of the strata from which the PSUs were selected. The selection of PSUs for the fall assessment was independent of the process of designating PSUs as winter or spring. As for the winter and spring subsamples, the 18 largest certainty PSUs were retained in the fall bridge sample with certainty. The distribution of PSUs among the fall, winter, and spring assessment periods is illustrated in Figure 3-1

Figure 3-1
Distribution of 94 Sampled PSUs by Inclusion
in the Fall, Winter, and Spring Assessment Periods



The age class 9 and age class 13 bridge samples used the same school and student eligibility requirements as the respective main samples. Nevertheless, special bridge samples were required because

1) The conditions for administration of the assessment varied considerably between the main sample and bridge sample sessions.

2) The need in the bridge samples for four distinct session types for age class 9 and five for age class 13 made it infeasible to conduct both a main sample session and the several bridge sessions within a given school.

3) For age class 13, the main samples were conducted at an inappropriate time of the year for bridging purposes, so that a sample of schools was needed to undertake the bridge assessments in the fall of 1987.

A separate sample of schools was required for the bridge sessions and the main sessions for age class 17 primarily because the definitions for student eligibility, based on age and grade, differed substantially between the two samples, even though the same population of schools was surveyed in each case. Conditions of administration varied somewhat also.

3.2 SCHOOLS FOR MAIN SAMPLES

In the second stage of sampling, the public, private, Catholic, Bureau of Indian Affairs, and Department of Defense schools within each of the 94 PSUs were listed according to the grade ranges associated with the three age classes. The population of eligible schools for each age class (Table 3-5) was the same for bridge and main samples in each case. Any school having one or more of the eligible grades, and located within an appropriate PSU, was included on the frame of schools (the list of schools from which the samples of schools were drawn) for a given sample. For each age class, only a fraction of one percent of age-eligible students was enrolled in ineligible schools. Each school within the 94 PSUs with a grade in the range of 2 to 12 was included in at least one age class—a total of 36,290 schools. An independent sample of schools was selected for each of the age classes. Thus some schools were selected for assessment of two age classes, and a few were selected for all three. The lists of schools were obtained from the 1986 list of schools maintained by Quality Education Data, Inc.

Table 3-5
Grade Definition of School Eligibility for Frame Inclusion and Frame Sizes, Main and Bridge Samples

Age Class	Frame included schools with any grade in the range	Number of schools on frame
9	2-5	26,951
13	6-9	28,167
17	9-12	8,485
Total	2-12	36,290

Schools within each PSU were selected (without replacement) with probabilities proportional to assigned ures of size. Roughly equal measures of size were assigned to schools containing estimates of grade/age-eligible students ranging from 20 to 150 (for age class 9), or to 200 (for age class 13 and age class 17). Schools larger than the indicated maximum size were selected within the selected PSUs with probabilities proportional to the number of grade/age-eligible students. As a result, in some instances a large school was selected with certainty within a relatively small selected PSU. Schools with fewer than 20 estimated grade/age eligibles were assigned considerably lower measures of size, and thus lower probabilities of selection, since assessment in these schools involved substantially higher per-student administrative costs.

A school with minority (Black and/or Hispanic) enrollment in excess of 15 percent of total enrollment was given double the probability of selection of a school of similar size in the same PSU with minority enrollment below 15 percent. Overall probabilities of selection for such high-minority schools were twice those for other schools of the same size from a given PSU in order to enlarge the sample of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups. For a given overall size of sample, this procedure reduces somewhat the reliability of estimates for all students as a whole and for those not Black or Hispanic.

The total number of schools selected for each age class was determined to be such that the predesignated student sample sizes would be achieved by selecting all eligible students in a selected school. up to a maximum of 150 (for age class 9) or 200 (for age class 13 and age class 17), allowing for losses due to nonparticipation of selected schools and students and the exclusion of students from the ass ssment.

In each of the 94 PSUs, a minimum of three schools was selected for age class 9, four schools for age class 13, and two schools for age class 17. These minima were established based on the total number of students and hence schools required for the particular age class, and on the proportion of selected schools likely to contain few or no eligible students. This proportion is much higher for age class 13 than for the other age classes because of the inclusion on the frame for this age class of commonly occurring elementary schools with a grade span of kindergarten through grade 6, and high schools with grade 9 through grade 12, which generally enroll few 13-year-olds. The use of these minima for the sample size of schools per PSU was derived as a compromise between two desirable but conflicting objectives. The first of these is to ensure substantial representation from within each selected J (after the impact of nonresponse). The second objective is the need to keep the variability in overall student sampling probabilities (and hence weights) to a low level, so as to control the sampling errors associated with N:EP estimates. The use of a large minimum sample size requirement per PSU would act to reduce unduly the weights of studerts selected from small PSUs.

This design, w:th the important exceptions described above, had the goal of yielding a sample of students in a given age or grade with approximately uniform probabilities of selection. The practical constraints on the sample size within each school, and the need to ensure an adequate ample within each PSU, resulted in some substantial violations of this general goal.

For all three age classes, a sample of schools was fir drawn for the bridge assessments (see below). These schools were then excl ed from the frame when the samples of schools were drawn for the main assessments. Adju ts were made to reflect the appropriate probabilities of selection, and th e sample weights, to yield unbiased estimates for both bridge and main samples. Schools selected for main assessments were further classified as belonging to the winter main assessment or the spring main assessment. For the schools in the PSUs assigned to a single season, this classification corresponded to the classification of the PSU containing the school. Half-samples of the schools in each of the 18 largest certainty PSUs (those PSUs assigned to both seasons) were assigned to the winter and spring assessments respectively.

After selection of the initial sample of schools was completed, information was obtained to update the sample fo new eligible schools. Public school districts and Catholic dioceses of initially selected schools were asked to give information about new schools and schools with changes in grade structure since the 1986 date to which the Quality Education Data list related, for their district/diocese. Schools so identified were given an appropriate chance of inclusion in each of the samples for which they were eligible. The overall probability of inclusion for a given age class for each such school was determined by the estimated number of eligible students enrolled in the school, and the within-PSU sampling rate used to select the initial sample of schools. The conditional probability of selection, used to draw the actual samples of new schools, was obtained by dividing the overall probability for the school by the probability that the school district was represented by a selected school in one of three initial main samples. These district level probabilities were obtained by reconstructing the sample selection procedure to obtain the probability that at least one school from the district was included in a given age class sample, and then combining these independent probabilities across age classes. The process identified a total of 58 new schools, 44 eligible for age class 9, 40 eligible for age class 13, and 11 eligible for age class 17. Three new schools were added to the sample in this way—two at age class 9 and one at age class 13. All three schools were found to have eligible students enrolled, and were invited to participate.

In a few PSUs where school refusals were relatively heavy for a particular sample, substitute school selections were made, replacing the refusals (to the extent feasible) with schools from within the same PSU and similar in size, affiliation (public, Catholic, or other private), grade span, and minority composition. In a very few cases, substitute schools were chosen from another PSU within the same stratum as the original. The goal of this procedure was to maintain the student sample sizes needed, while keeping variance and nonresponse bias at acceptable levels. Table 3-6 shows the number of in-scope schools selected, cooperating, and substituted, in each of the school samples. The participation rates given are based on the initially selected sample of schools. These response rates are comparable with those of previous assessments conducted during the 1980s. Note that since the response rates quoted do not include the substitute selections, the potential for nonresponse bias is likely to be somewhat less than these rates would indicate. This is because the substitute selections were chosen based on their similarity to the initially refusing selections.

The considerable numbers of schools selected with no eligible students enrolled resulted primarily from the fact that, for example, for grade 8/age 13, some schools with grades 6, 7, or 9, but no grade 8, were sampled. Such schools had a reasonable chance of containing some age 13 students. Often they did have a number of eligible students, but sometimes they had none. Because of the grade structure of schools, this occurred most often for grade 8/age 13.

Table 3-6
School Sample Sizes, Refusals, and Substitutes
Main Samples

	Age Class 9	Age Class 13	Age Class 17	Total
Selected, in scope	373	506	383	1,262
Refusals	42	68	66	176
Participation rate of originally selected schools	89%	87%	83%	86%
1986 participation rate	89%*	89%*	81%**	87%
Participating; no eligible enrolled	4	39	13	56
Substitutes participating	9	14	8	31
Final assessed sample	336	413	312	1,061

* The rates quoted for 1986 are for the grade 3/age 9 and grade 7/age 13 spring samples respectively.

** No assessment for grade 12/age 17 was conducted before 1988. The rate quoted here is for the 1986 grade 11/age 17 spring sample.

A school characteristics and policies questionnaire was mailed to every sampled school by Westat before the assessment. The Westat supervisor then collected the questionnaires and returned them to ETS. The school characteristics and policies questionnaire is described in Chapter 4.

A school principal's questionnaire, distributed to each sampled school by Westat before the assessment, was used to refine the estimate of the grade/age-eligible students and to determine in part the size and type of community (STOC) codes (see Appendix G and Rust et al., 1990).

3.3 SELECTION OF SCHOOLS FOR BRIDGE SAMPLES; THE ASSIGNMENT OF SESSIONS TO SCHOOLS

Schools were selected for age class 13 bridge assessments (conducted in the fall of 1987) from the subsample of 64 PSUs that had been designated as the age class 13 bridge PSUs. One school or school cluster was selected in each PSU, except that three schools or school clusters were selected from each of the two largest PSUs (which were substantially larger than the remainder). Initially, four types of bridge sessions were to be administered, and the sample was selected so as to meet this requirement, with a maximum of four sessions to be administered in a school, with each session type to be administered in each PSU. However, a fifth session type was added too late to amend the school sample (although the total required sample size of students was not changed). Thus the requirement that each session type be administered at least once in each PSU was modified so that each session type would be administered at least once in at least 52 PSUs. To avoid the possibility that a particular bridge session might be assigned to a single school with only one or very few eligibles, small schools were grouped with other schools in the same PSU to form clusters of a specified minimum number of eligibles. Bridge sessions were then assigned within each PSU by selecting a school or school cluster with probability proportional to the estimated number of age and grade eligibles within the school(s).

Schools were selected for the age class 9 bridge assessments (conducted in the winter of 1988) from the subsample of 56 PSUs designated for the winter assessment. Four types of bridge sessions were to be administered. The selection was such that each of the session types used in the bridge assessments would be administered in at least one school within each of the 56 PSUs designated as winter or as winter and spring. Schools or school clusters were sampled in the same manner as for age class 13 except that two schools or school clusters were selected per PSU, with six schools or school clusters drawn from each of the two largest PSUs.

In a like manner, schools were selected for the age class 17 bridge assessments (conducted in the spring of 1988) from the subsample of 56 PSUs designated for the spring assessment. Three session types were to be administered. Each of the session types was to be administered within at least one school or school cluster in each of the 56 PSUs. Two schools (or school clusters) were selected per PSU, with six selections being made from each of the two largest PSUs.

Substitute selections were made for nonparticipating bridge sample schools in certain PSUs in similar manner to the approach used for the main sample schools. In the case of the bridge samples, however, no samples of new schools were selected. For the bridge samples, the smaller sample sizes involved meant that the biasing effect of failing to represent such new schools was relatively less, as a component of total error. For the age class 13 sample in particular, there was insufficient time available between the point when districts were contacted and when assessments were scheduled (mid-October through early December, 1987) to ascertain the existence of new schools, sample them, and obtain the participation of the schools.

Table 3-7 shows the school sample sizes and participation rates for the bridge samples for each age class. School participation rates are similar to those seen in comparable samples in 1986—the fall and winter bridges and the grade 11/age 17 spring main sample.

Table 3-7
School Sample Sizes, Refusals, and Substitutes
Bridge Samples

	Grade 4/Age 9 (Winter)	Grade 8/Age 13 (Fall)	Grade 11/Age 17 (Spring)	Total
Selected, in scope	180	219	155	554
Refusals	23	16	34	73
Participation rate of originally selected schools	87%	93%	78%	87%
1986 participation rate	87%	84%	81%	83%
Participating; no eligibles enrolled	3	30	7	40
Substitutes participating	2	8	6	16
Final assessed sample	156	181	120	457

For all three age classes, sessions were assigned to bridge sample schools in the following manner. First, the number of sessions per school was established. This was the maximum number of sessions (up to four) that could be administered without creating unduly small session sizes with few eligibles. Thus, in most bridge sample schools, four sessions were conducted. However, schools with fewer than 20 eligibles, for example, were asked to conduct only a single session.

The number of session types conducted in the assessment varied by age class. Table 3-8 in the following section shows, among other things, the various bridge sample session types conducted for each age class, and the year of the corresponding assessment to which these session types provided a bridge.

The assignment of sessions to schools maximized the number of session types conducted within each PSU. Thus, to the extent feasible, session assignment was delayed until after it was determined that a selected school would participate. On a few occasions, a session could not be conducted in a school that, at the time of session assignment, was expected to participate but subsequently did not. As a result, two types of school nonresponse adjustment factors, denoted school and session, were required for the bridge samples (see Chapter 8).

This procedure was intended to assure that each session type was assigned in each PSU at least once for the age class 9 and age class 17 samples. At age class 13 however, often a PSU was represented in the sample by a single large school. Since it was not feasible to administer each of five different session types in a single school, not all session types were administered in all 64 PSUs, but each session type was administered in at least 52 PSUs for this age class.

3.4 SAMPLING STUDENTS

In the fourth stage of sampling, a consolidated list was prepared for each school of all grade-eligible and age-eligible students for the age class for which the school was selected. A systematic selection of eligible students was made from this list (unless all students were to be assessed) to provide the target sample size. For bridge sample schools assigned to more than a single session type, students were assigned by Westat district supervisors to print or paced-tape sessions using specified procedures. Students assigned to paced-tape sessions who were not age-eligible were dropped from the assessment.

The maxima established for the number of students to be scheduled for assessment were 150 for age class 9, 220 for age class 13, and 200 for age class 17. The limit of 220 for age class 13 was raised from an initial limit of 200 owing to concern that the selected sample of schools would not yield sufficient assessed students. Note that at all three age classes the number of students invited to the assessment in larger schools (those with numbers of enrolled eligibles in excess of these limits) was likely to be significantly below these specified limits. This was because not only were excluded students not invited, but those students in the modal grade who were not age eligible but were selected for assessment in the tape sessions (the majority of the sample at ages 9 and 13) were not invited to participate (unless specifically requested by the school for operational reasons).

The sample for students to be selected in each school was derived in the following manner, both for main and for bridge samples. On the basis of data obtained from the principal questionnaire (or the sample frame when the principal questionnaire data were not obtained in time) an estimate of the number of grade- plus age-eligible students was established for each school. A Session Assignment Form was generated for each school, showing the line numbers (described below) of the students to be selected (and in the case of the bridge samples, indicating the type of session to be taken by each such student). These line numbers were generated using a sampling interval designed to give the appropriate sample size for each school. Thus the overall sampling interval was 1.0 for schools in which all eligible students were to be assessed. The appropriate sampling interval was specified for schools with larger numbers of eligible students, such as to give the appropriate maximum sample size (described above for each age class) in the case that the school had an enrollment of eligible students exactly equal to that predicted.

If the Westat supervisor found that, when applied to the numbered list of eligible students assembled in the field for each school, the line numbers generated gave rise to a sample in excess of 120 percent of the appropriate maximum sample size limit specified above, he or she called Westat's central office. By use of a personal computer, new line numbers based on the actual number of eligible students were generated and relayed to the supervisor. A similar revision to the line numbers was made in the case of a school with a sampling interval in excess of 1.0, and eligible enrollment less than 80 percent of that initially estimated. In this latter case the sample size was increased to the appropriate level. This procedure gave a suitable compromise between control over the sampling rate within each school and operational autonomy and flexibility for Westat field supervisors. Note that in all cases, sampling intervals were generated in Westat's central office, and stored for use in sample weighting. Supervisors were not required to derive or record within-school sampling rates.

Table 3-8 shows the number of students per school who were assessed for each session type. Note that, for the various spiral samples, the number of students assessed per item per school is quite low, even though typically dozens of students were assessed in total in a particular school. Thus the extent of clustering of the sample is in general quite modest, because most bridge sample schools conducted a few different types of sessions with a moderate number of students in each, and more importantly because the use of BIB spiraling in the main samples and print-administered bridge sample sessions greatly alleviated the effects of clustering the samples of students within schools.

3.5 EXCLUDED STUDENTS

Some students selected for the sample were deemed unassessable by school authorities because they had limited English language proficiency, were judged as being mildly mentally retarded (educable), or were functionally disabled. In these cases, school staff completed an excluded student questionnaire, listing the reason for exclusion and providing some background information. For the excluded students, no distinction was made as to the season of the year in which their school was assessed, since the timing was unimportant for these students. For age class 9 and age class 13, no distinction was made between students excluded from bridge assessments and students excluded from the main assessment, since the same grade and age eligibility definitions applied in each case. Conversely, for age class 17, the excluded students from the bridge assessments (with an October-September age definition and modal grade of 11) were treated as separate from the excluded students in the main assessment (with a calendar-year age definition and modal grade of 12).

For all samples, all selected students, whether eligible by age, grade, or both, were considered for possible exclusion, on the grounds that they would not be able to participate meaningfully in the assessment. All students so identified were included in the sample of excluded students. This occurred even if the student, had he or she been assessable, would have subsequently been dropped from a bridge tape session sample because of age ineligibility. This represented a change from the 1986 procedure, where only students who would otherwise have been invited (i.e., met the specific age or grade requirements for the particular type of session for which they were selected) were included in the sample of excluded students. The procedure for 1988 provided consistency in the definition of the excluded student samples (all

Table 3-8

Number of Students per School for Each Session Type

Sample	Session Type	Bridge Type	No. of Schools	No. of Students per Session Type per School: Mean	No. of Students per Session Type per School: Max.	No. of Students per Item per School: Mean	No. of Students per Item per School: Max.
Age Class 9 Bridges	Spiral booklets 51-56	to 1986	152	34.1	81	5.7	14
	Tape booklet 91	to 1984	112	11.4	39	11.4	39
	Tape booklet 92	to 1984	111	11.2	37	11.2	37
	Tape booklet 93	to 1984	112	10.7	22	10.7	22
Age Class 9 Main	Spiral	—	334	68.9	219[a]	8.0[b]	26[b]
Age Class 13 Bridges	Spiral booklets 51-56	to 1986	143	38.5	135	6.4	23
	Tape booklet 90	Civics	105	18.5	55	18.5	55
	Tape booklet 91	to 1984	73	19.2	36	19.2	36
	Tape booklet 92	to 1984	73	17.5	36	17.5	36
	Tape booklet 93	to 1984	71	17.7	56	17.7	56
Age Class 13 Main	Spiral	—	410	89.5	221	6.2[b]	16[b]
Age Class 17 Bridges	Spiral booklets 51-56	to 1986	107	43.2	84	7.2	14
	Spiral booklets 61-67	to 1984	119	59.3	140	6.6[c]	10[c]
						19.8[d]	47[d]
	Tape booklet 90	Civics	97	18.4	27	18.4	27
Age Class 17 Main	Spiral	—	312	104.8	241[e]	7.9[b]	18[b]

[a] The next highest was 159.
[b] The per-item sample sizes for the main samples are for the major cross-sectional samples only. They do not cover such assessment components as geography (at age class 17) or international mathematics (at age class 13).
[c] for booklets 61-66
[d] for booklet 67
[e] The next highest was 200.

age- or grade-eligible students were considered) and also increased the sample size of grade-only-eligible excluded students.

3.6 STUDENT PARTICIPATION AND EXCLUSION RATES

Table 3-9 summarizes the rates of exclusion of selected students and the rates of participation of invited students. The set of invited students consists of the selected students, after removing the excluded students and, in the case of bridge samples, removing those students selected for tape sessions who were not age eligible. For a given session, a makeup session was called for when, for various reasons, more than a tolerable number of invited students failed to attend the originally scheduled session to which they were invited. The participation rates given in the table express the number finally assessed as a percentage of those initially invited in the participating schools. The rates for bridge and main samples are combined for age class 9 and age class 13, since the student eligibility criteria were the same in bridge and main samples for these age classes. For grade 11/age 17 samples, predominantly eleventh graders were selected, whereas for the grade 12/age 17 sample predominantly twelfth graders were selected. For each age class, Table 3-9 presents comparison rates for the most recent assessment (indicated in parenthesis) in which a sizeable sample from the corresponding grade/age cohort was sampled. Note that 1988 was the first year in which a sample of predominantly twelfth graders was assessed.

Table 3-9
Exclusion and Participation Rates by Age Class, Unweighted

	Excluded (%)	Previous Excluded (%)	Number Invited	Participation Rate (%)	Previous Participation Rate (%)
Grade 4/age 9	6.3	3.8 (1984)	34,535	92.6	91.3 (1984)
Grade 8/age 13	5.3	3.6 (1984)	54,466	88.4	87.3 (1984)
Grade 11/age 17	3.0	3.4 (1986)	17,000	79.2	78.9 (1986)
Grade 12/age 17	3.7	—	41,681	78.5	—

The major change from previous assessments is in the proportion excluded at lower ages (in 1986, the exclusion rate for grade 3/age 9 was 3.9 percent, while for grade 7/age 13 it was 3.7 percent). The reason for this increase in the rate of exclusion, at least for grade 4/age 19, is the result of an increase in the proportion excluded for reason of limited English language proficiency. In large part this appears to be attributable to the increased practice of educating native Spanish speakers in Spanish in elementary schools. In a few selected schools for the grade 4/age 9 samples, for example, more than 50 percent of selected students were excluded from the assessment for reason of limited English language proficiency. The reasons for the increased level of exclusion for 13-year-olds are less clear.

3.7 OVERALL STUDENT PARTICIPATION RATES

The combined impact of school nonparticipation and student absenteeism from sessions within participating schools is summarized in table 3-10. The table shows the percentages of students assessed, from among those who would have been assessed if all initially selected schools had participated, and if all invited students had attended either an initial or make-up session. The results show that, consistent with earlier rounds of NAEP, the overall level of nonresponse increases substantially with the increase in age and grade of the students. Levels of nonresponse at the twelfth grade, assessed for the first time in 1988 as part of the main sample for age class 17, were not appreciably higher than for the eleventh grade, assessed as part of the bridge sample for this age class.

The procedures for substituting for nonparticipating schools, or imputing for them (see section 8.1.2.1), and the procedures for imputing for absent students (see section 8.1.2.4) were designed (so far as feasible) to reduce the biases resulting from school and student nonparticipation.

3.8 SAMPLING TEACHERS

The teacher questionnaire was administered to the reading teachers of fourth-grade students sampled for the main assessment of reading and to the writing teachers of eighth-grade students sampled for the main assessment of writing. The purpose of drawing these samples was not to estimate the attributes of the teacher population, but to estimate the number (proportion) of students whose teachers had various attributes and to correlate student characteristics and performance with the characteristics of their teachers

The reading teachers of every fourth-grade student assessed for reading in the main assessment (i.e., respondents to booklets 8-14) were identified in each school. Up to seven of these reading teachers in each school were selected to complete the teacher questionnaire; in schools with more than seven reading teachers with assessed students, a random sample of five of these teachers was selected. Every selected reading teacher was provided a list of all his or her students (up to a maximum of 10) who had been assessed for reading in the main assessment; if more than 10 students fitted the criteria, a random sample of 10 such students was provided. The selected teachers were asked to complete a questionnaire about the reading abilities of each selected student and the kinds of reading instruction the student received. The analysis of the reading teacher questionnaire data is discussed in section 10.3.

Similarly, the writing teachers of every eighth-grade student assessed for writing in the main assessment (i.e., respondents to booklets 1-7) were eligible to complete a teacher questionnaire. Up to seven teachers were selected in each school and each selected teacher was provided with a list of all his or her students (up to a maximum of 10) who had been assessed for writing. The selected teachers were asked to complete a questionnaire about the writing capabilities of each selected student and about the writing instruction the student received. The analysis of the writing teacher questionnaire data is discussed in section 11.2.

Table 3-10

Overall Participation Rates (School and Student Combined) by Age Class

	Age Class 9	Age Class 13	Age Class 17	Overall
Main Samples				
School participation	88.7%	86.6%	82.8%	86.1%
Student participation	92.8%	87.8%	78.5%	85.4%
Overall student participation	82.3%	76.0%	65.0%	73.0%
Number of participating students	23,092	36,699	32,710	92,501
Bridge Samples				
School participation	87.2%	92.7%	78.1%	86.8%
Student participation	92.2%	90.0%	79.2%	85.9%
Overall student participation	80.4%	83.4%	61.9%	72.6%
Number of participating students	8,899	11,423	13,460	33,782
Overall				
School participation	88.3%	87.8%	81.4%	86.3%
Student participation	92.6%	88.4%	78.7%	85.5%
Overall student participation	81.8%	77.6%	64.1%	72.9%
Number of participating students	31,991	48,122	46,170	126,283

Chapter 4

ASSESSMENT INSTRUMENTS[1]

Janet R. Johnson

Educational Testing Service

The 1988 assessment incorporated four types of instruments: student assessment booklets, a questionnaire about excluded students, a teacher questionnaire (grades 4 and 8), and a school characteristics and policies questionnaire. The data collected from these instruments are available on the 1988 NAEP public-use data tapes. This chapter describes the assessment instruments and begins with a discussion of the way in which the items were organized into blocks to create the student assessment instruments.

4.1 MAIN SAMPLE STUDENT ASSESSMENT INSTRUMENTS

Student assessment booklets contained both cognitive and noncognitive items. Cognitive items were used to assess student achievement in the subject areas of reading, writing, civics, U.S. history, and geography. Noncognitive items were used to gather student background and attitude information. Some noncognitive items were presented to every student; these were placed together in a block called the common background block or common core and covered such topics as race/ethnicity, levels of parental education, items in the home, homework, and television watching habits. Other noncognitive items were specific to one of the five subject areas. These items appeared together in a block and were presented in booklets that contained blocks of cognitive items related to the same subject area.

Main assessment focused-BIB booklets achieved a certain degree of uniformity in that each booklet contained five discrete blocks of items: the first block contained the common core block of background items; the second block contained the subject-specific attitude items; the remaining three blocks each contained cognitive items specific to a particular subject area. Each student at all three grade/ages was administered a single booklet.

Main assessment intercorrelation booklets resembled the focused-BIB booklets except that the three blocks containing cognitive items each covered a different subject area: reading, civics, and U.S. history (as well as some geography at grade 12/age 17).

---

[1] The author is heavily indebted to the *National Assessment of Educational Progress 1988 Public-use Data Tapes Version 1.0 User Guide* (Rogers, Kline, Johnson, & Rust, 1989) and in particular to Debra Kline for portions of this chapter.

Main assessment special booklets were designed to allow special studies relating to achievement. To measure the relationship between time allocated to the writing task and writing performance, each of three booklets at each grade/age contained the common background block plus two writing blocks, the second of which was allocated twice as much time as the first. In addition, there were two special study booklets created for grade 8/age 13 and grade 12/age 17 that were composed of the common background block, the reading noncognitive block and three cognitive blocks that were combinations of reading and document literacy blocks. These booklets were designed to allow the assessment of document literacy and its relation to reading proficiency.

Block Timing

The common core background question block appeared first within every booklet. It was read aloud to grade 4/age 9 students and took approximately 10 minutes to complete. The other grade/ages were given five minutes in which to complete this section and only the first question, regarding student race/ethnicity, was read aloud.

At each grade/age, the common core block was followed by a five-minute block of subject-specific background questions about students' experiences and instruction in the subject area. All students read these questions on their own.

The remainder of each student booklet was composed of three more blocks. These were assembled from the pool of cognitive items within a subject area. Students at grade 4/age 9 were given blocks that were each 10 minutes in length, while the other two grade/ages were given three 15-minute blocks.

The overall assessment time was approximately 45 minutes for grade 4/age 9 students and 55 minutes for grade 8/age 13 and for grade 12/age 17 students.

Item Arrangement

For the 1988 assessment, blocks of cognitive items were assigned to booklets in such a way that each subject area block appeared in the same number of booklets and each possible pair of blocks within a subject area appeared in at least one booklet. The total pool of items was so large that no individual student could have received all items.

The assessment booklets themselves were ordered in such a way that typically only several students in any assessment session in any school received the same booklet.

Table 4-1 shows the total number of cognitive blocks created for each subject area in the main sample.

Table 4-1
Number of Main Sample Subject Area Cognitive Blocks Administered

Subject Area	Grade 4/Age 9	Grade 8/Age 13	Grade 12/Age 17
Reading	7	7	7
Writing	7	7	7
Civics	3	7	7
History	3	7	7
Geography	-	-	3
Intercorrelation	9	9	9
Long Writing	6	6	6
Document Literacy	-	3	3
Total	20	31	34

Tables 4-2, 4-3, and 4-4 show which subject area blocks were contained in each booklet for each grade/age, how many of each booklet were administered, and the booklet response method (circling the answer in the booklet, filling in scannable ovals in the booklet, or filling in ovals on a scannable answer sheet).

Tables 4-5, 4-6, and 4-7 display the composition of individual blocks of items administered in the main NAEP assessment. These tables provide the item placement number within each block as well as the item type (e.g. background, subject, multiple-choice cognitive, or open-ended cognitive).

The focused-BIB design of these blocks and booklets generated a total of 22 different booklets for grade 4/age 9, 36 different booklets for grade 8/age 13, and 37 different booklets for grade 12/age 17.

4.2 BRIDGE SAMPLE STUDENT ASSESSMENT INSTRUMENTS

The 1988 assessment included four distinct bridge (or trend) studies. These are each described below. Tables 4-8, 4-9, and 4-10 show which subject area blocks were contained in each bridge booklet, how many of each booklet were administered, and the booklet response method.

Civics Bridge (Booklet 90)

The civics bridge sample links the 1988 main assessment data to data from citizenship/social studies assessments conducted in 1976 and 1982. The sample consists of one booklet each for ages 13 and 17. (Because there were no cognitive civics items per se from 1976 and 1982 for age 9, a trend sample for that age was not necessary.) Booklet 90 for both ages contains the common background block (B1), the civics background and attitude block from the main assessment (C1), and three blocks of cognitive items (C9, C10, and C11). Blocks B1 and C1 each took five minutes of assessment time; the three cognitive blocks were administered in 45 minutes. The cognitive blocks

Table 4-2

Booklet Contents and Number of Booklets Administered
Main Sample, Grade 4/Age 9

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Writing	1	B	B1	W1	W2	W3	W5	896
	2	B	B1	W1	W3	W4	W6	892
	3	B	B1	W1	W4	W5	W7	894
	4	B	B1	W1	W5	W6	W8	900
	5	B	B1	W1	W6	W7	W2	901
	6	B	B1	W1	W7	W8	W3	882
	7	B	B1	W1	W8	W2	W4	882
Reading	8	B	B1	R1	R2	R3	R5	889
	9	B	B1	R1	R3	R4	R6	879
	10	B	B1	R1	R4	R5	R7	876
	11	B	B1	R1	R5	R6	R8	885
	12	B	B1	R1	R6	R7	R2	883
	13	B	B1	R1	R7	R8	R3	884
	14	B	B1	R1	R8	R2	R4	881
U.S. History	15	B	B1	H1	H2	H3	H4	2664
Civics	16	B	B1	C1	C2	C3	C4	2652
Intercorrelation	17	B	B1	X1	H2	C2	R2	869
(Reading,	18	B	B1	X1	R3	H3	C3	884
U.S. History, Civics)	19	B	B1	X1	C4	R7	H4	885
Long Writing	20	B	B1	W1	W6	W9		880
	21	B	B1	W1	W8	W10		877
	22	B	B1	W1	W5	W11		877
				TOTAL MAIN SAMPLE BOCKLETS. . .				23012

† B = scannable booklet

Table 4.3

Booklet Contents and Number of Booklets Administered
Main Sample, Grade 8/Age 13

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Writing	1	B	B1	W1	W2	W3	W5	866
	2	B	B1	W1	W3	W4	W6	854
	3	B	B1	W1	W4	W5	W7	862
	4	B	B1	W1	W5	W6	W8	860
	5	B	B1	W1	W6	W7	W2	848
	6	B	B1	W1	W7	W8	W3	862
	7	B	B1	W1	W8	W2	W4	859
Reading	8	A	B1	R1	R2	R3	R5	855
	9	A	B1	R1	R3	R4	R6	856
	10	A	B1	R1	R4	R5	R7	849
	11	A	B1	R1	R5	R6	R8	845
	12	A	B1	R1	R6	R7	R2	855
	13	A	B1	R1	R7	R[illegible]	R3	795
	14	A	B1	R1	R8	R2	R4	857
U.S. History	15	A	B1	H1	H2	H3	H5	854
	16	A	B1	H1	H3	H4	H6	872
	17	A	B1	H1	H4	H5	H7	851
	18	A	[illegible]1	H1	H5	H6	H8	857
	19	A	B1	H1	H6	H7	H2	865
	20	A	B1	H1	H7	H8	H3	854
	2[illegible]	A	B1	H1	H8	H2	H4	835
Civics	22	A	B1	C1	C2	C3	C5	859
	23	A	B1	C1	C3	C4	C6	855
	24	A	B1	C1	C4	C5	C7	845
	25	A	B1	C1	C5	C6	C8	823
	26	A	B1	C1	C6	C7	C2	[illegible]5
	27	A	B1	C1	C7	C8	C3	[illegible]1
	28	A	B1	C1	C8	C2	C4	863
Intercorrelation (Reading, U.S. History, Civics)	29	A	B1	X1	H6	C7	R3	859
	30	A	B1	X1	R5	H5	C2	858
	31	A	B1	X1	C6	R6	H2	873
Long Writing	32	B	B1	W1	W6	W9		858
	33	B	B1	W1	W8	W10		869
	34	B	B1	W1	W4	W11		859
Document Literacy	35	B	B1	R1	R2	D2	R5	1267
	36	B	B1	R1	D3	R7	D4	1266
					TOTAL MAIN SAMPLE BOOKLETS. . .			31601

† A = scannable answer sheet, B = scannable booklet

Table 4-4

Booklet Contents and Number of Booklets Administered
Main Sample, Grade 12/Age 17

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Writing	1	B	B1	W1	W2	W3	W5	827
	2	B	B1	W1	W3	W4	W6	835
	3	B	B1	W1	W4	W5	W7	813
	4	B	B1	W1	W5	W6	W8	827
	5	B	B1	W1	W6	W7	W2	821
	6	B	B1	W1	W7	W8	W3	815
	7	B	B1	W1	W8	W2	W4	802
Reading	8	A	B1	R1	R2	R3	R5	822
	9	A	B1	R1	R3	R4	R6	814
	10	A	B1	R1	R4	R5	R7	824
	11	[illegible]	B1	R1	R5	R6	R8	826
	12	A	B1	R1	R6	R7	R2	823
	13	A	B1	R1	R7	R8	R3	838
	14	A	B1	R1	R8	R2	R4	821
U.S. History	15	A	B1	H1	H2	H3	H5	831
	16	A	B1	H1	H3	H4	H6	827
	17	A	B1	H1	H4	H5	H7	830
	18	A	B1	H1	H5	H6	H8	820
	19	A	B1	H1	H6	H7	H2	829
	20	A	B1	H1	H7	H8	H3	824
	21	A	B1	H1	H8	H2	H4	819
Civics	22	A	B1	C1	C2	C3	C5	818
	23	A	B1	C1	C3	C4	C6	809
	24	A	B1	C1	C4	C5	C7	817
	25	A	B1	C1	C5	C6	C8	807
	26	A	B1	C1	C6	C7	C2	810
	27	A	B1	C1	C7	C8	C3	814
	28	A	B1	C1	C8	C2	C4	808
Geography	29	A	B1	G1	G2	G3	G4	2446
Intercorrelation (Reading, U.S. History Civics, Geography)	30	A	B1	X1	R2	H6	G4	817
	31	A	B1	X1	H2	R5	C4	815
	32	A	B1	X1	G2	C6	R8	806
Long Writing	33	B	B1	W1	W6	W9		799
	34	B	B1	W1	W8	W10		811
	35	B	B1	W1	W4	W11		820
Document Literacy	36	B	B1	R1	R2	D2	R5	1217
	37	B	B1	R1	D3	R7	D4	1208
				TOTAL MAIN SAMPLE BOOKLETS. . .				32710

† A = scannable answer sheet, B = scannable booklet

Table 4-5

1988 Main Sample Block Information, Grade 4/Age 9

Block	Type	Order of Background Items	Order of Cognitive Items	Total Items	Total Cognitive Items	Total Open-ended Cognitive Items
9B1	Common Bkgd.	1-21		21		
9R1	Rdg. Bkgd.	1-10		10		
9R2	Reading		1-14	14	14	
9R3	Reading		1-7	7	7	
9R4	Reading		1-7	7	7	
9R5	Reading		1-15	15	15	1
9R6	Reading	14	1-13	14	13	
9R7	Reading		1-15	15	15	
9R8	Reading		1-11	11	11	1
9W1	Wrt. Bkgd.	1-10		10		
9W2	Writing		1	1	1	1
9W3	Writing		1	1	1	1
9W4	Writing		1	1	1	1
9W5	Writing		1	1	1	1
9W6	Writing		1	1	1	1
9W7	Writing		1	1	1	1
9W8	Writing		1	1	1	1
9W9	Long Writing		1	1	1	1
9W10	Long Writing		1	1	1	1
9W11	Long Writing		1	1	1	1
9C1	Civics Bkgd.	1-9		9		
9C2	Civics		1-15	15	15	
9C3	Civics		1-16	16	16	
9C4	Civics		1-20	20	20	
9H1	Hist. Bkgd.	1-9		9		
9H2	History		1-15	15	15	
9H3	History		1-15	15	15	
9H4	History		1-15	15	15	
9X1	Intercorr. Bkgd.	1-10		10		

Table 4-6

1988 Main Sample Block Information, Grade 8/Age 13

Block	Type	Order of Background Items	Order of Cognitive Items	Total Items	Total Cognitive Items	Total Open-ended Cognitive Items
13B1	Common Bkgd.	1-21		21		
13R1	Rdg. Bkgd.	1-20		20		
13R2	Reading		1-14	14	14	
13R3	Reading		1-9	9	9	
13R4	Reading		1-10	10	10	
13R5	Reading		1-14	14	14	
13R6	Reading		1-18	18	18	
13R7	Reading		1-19	19	19	
13R8	Reading		1-15	15	15	1
13W1	Wrt. Bkgd.	1-24		24		
13W2	Writing		1	1	1	1
13W3	Writing		1	1	1	1
13W4	Writing		1-2	2	2	2
13W5	Writing		1	1	1	1
13W6	Writing		1	1	1	1
13W7	Writing		1	1	1	1
13W8	Writing		1	1	1	1
13W9	Long Writing		1	1	1	1
13W10	Long Writing		1	1	1	1
13W11	Long Writing		1	1	1	1
13C1	Civics Bkgd.	1-25		25		
13C2	Civics		1-26	26	26	
13C3	Civics		1-24	24	24	
13C4	Civics		1-30	30	30	
13C5	Civics		1-27	27	27	
13C6	Civics		1-24	24	24	
13C7	Civics		1-22	22	22	
13C8	Civics		1	1	1	1
13H1	Hist. Bkgd.	1-27		27		
13H2	History		1-26	26	26	
13H3	History		1-26	26	26	
13H4	History		1-26	26	26	
13H5	History		1-26	26	26	
13H6	History		1-28	28	28	
13H7	History		1-28	28	28	
13H8	History		1	1	1	1
13D2	Doc. Lit.		1-10	10	10	7
13D3	Doc. Lit.		1-14	14	14	10
13D4	Doc. Lit.		1-13	13	13	10
13X1	Intercorr. Bkgd.	1-25		25		

Table 4-7

1988 Main Sample Block Information, Grade 12/Age 17

Block	Type	Order of Background Items	Order of Cognitive Items	Total Items	Total Cognitive Items	Total Open-ended Cognitive Items
17B1	Common Bkgd.	1-33		33		
17R1	Rdg. Bkgd.	1-32		32		
17R2	Reading		1-14	14	14	
17R3	Reading		1-11	11	11	1
17R4	Reading		1-19	19	19	
17R5	Reading		1-14	14	14	
17R6	Reading		1-15	15	15	1
17R7	Reading		1-19	19	19	
17R8	Reading		1-18	18	18	
17W1	Wrt. Bkgd.	1-35		35		
17W2	Writing		1	1	1	1
17W3	Writing		1	1	1	1
17W4	Writing		1-2	2	2	2
17W5	Writing		1	1	1	1
17W6	Writing		1	1	1	1
17W7	Writing		1	1	1	1
17W8	Writing		1	1	1	1
17W9	Long Writing		1	1	1	1
17W10	Long Writing		1	1	1	1
17W11	Long Writing		1	1	1	
17C1	Civics Bkgd.	1-34		34		
17C2	Civics		1-26	26	26	
17C3	Civics		1-26	26	26	
17C4	Civics		1-25	25	25	
17C5	Civics		1-27	27	27	
17C6	Civics		1-24	24	24	
17C7	Civics		1-22	22	22	
17C8	Civics		1	1	1	1
17H1	Hist. Bkgd.	1-36		36		
17H2	History		1-28	28	28	
17H3	History		1-25	25	25	
17H4	History		1-25	25	25	
17H5	History		1-26	26	26	
17H6	History		1-28	28	28	
17H7	History		1-28	28	28	
17H8	History		1	1	1	1
17G1	Geog. Background	1-28		28		
17G2	Geography		1-26	26	26	
17G3	Geography		1-26	26	26	
17G4	Geography		1-26	26	26	
17D2	Doc. Lit.		1-10	10	10	7
17D3	Doc. Lit.		1-14	14	14	10
17D4	Doc. Lit.		1-13	13	13	10
17X1	Intercorr. Bkgd.	1-36		36		

Table 4-8

Booklet Contents and Number of Booklets Administered
Bridge Samples, Age Class 9

BRIDGE TO 1984, GRADE 4/AGE 9

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading and Writing	51	C	CC	‡	C	L	Q	884
	52	C	CC	‡	H	E	R	879
	53	C	CC	‡	C	K	J	860
	54	C	CC	‡	G	O	E	853
	55	C	CC	‡	M	G	N	861
	56	C	CC	‡	V	R		851
					TOTAL 1984 BRIDGE BOOKLETS. . .			5188

BRIDGE TO 1986, AGE 9

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading, Mathematics, and Science	91	B	B1	‡	R1	M1	S1	1274
	92	B	B1	‡	S2	R2	M3	1240
	93	B	B1	‡	M2	S3	R3	1197
					TOTAL 1986 BRIDGE BOOKLETS. .			3711

---

† B = scannable booklet, C = circled-answer booklet
‡ Subject area background questions are included in the cognitive blocks for this booklet.

Table 4-9

Booklet Contents and Number of Booklets Administered
Bridge Samples, Age Class 13

BRIDGE TO 1984, GRADE 8/AGE 13

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading and Writing	51	C	CC	‡	M	K	D	907
	52	C	CC	‡	C	L	Q	915
	53	C	CC	‡	[illegible]	E	R	924
	54	C	CC	‡	N	C	D	927
	55	C	CC	‡	G	O	E	906
	56	C	CC	‡	G	J	P	921

TOTAL 1984 BRIDGE BOOKLETS. . 5500

CIVICS BRIDGE, AGE 13

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Civics	90	B	B1	C1	C9	C10	C11	1938

TOTAL CIVICS BRIDGE BOOKLETS. . . 1933

BRIDGE TO 1986, AGE 13

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading, Mathematics, and Science	91	B	B1	‡	R1	M1	S.	1405
	92	B	B1	‡	S2	R2	M3	1281
	93	B	B1	‡	M2	S3	R3	1256

TOTAL 1986 BRIDGE BOOKLETS. . . 3942

† B = scannable booklet, C = circled-answer booklet
‡ Subject area background questions are included in the cognitive blocks for this booklet

Table 4-10

Booklet Contents and Number of Booklets Administered
Bridge Samples, Age Class 17

BRIDGE TO 1984, GRADE 11/AGE 17

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading and Writing	51	C	CC	‡	M	K	D	774
	52	C	CC	‡	C	L	Q	768
	53	C	CC	‡	H	E	R	770
	54	C	CC	‡	N	C	D	777
	55	C	CC	‡	G	O	E	[illegible]
	56	C	CC	‡	G	J	P	762
					TOTAL 1984 BRIDGE BOOKLETS. . .			4622

BRIDGE TO 1986, GRADE 11/AGE 17

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading, Mathematics, and Science	61	B	B1	‡	R2	M4	M2	799
	62	B	B1	‡	M1	R2	M9	786
	63	B	B1	‡	S1	S11	R1	798
	64	B	B1	‡	S2	S4	R5	779
	65	B	B1	‡	S3	R6	M3	780
	66	B	B1	‡	R3	R4	R2	761
U.S. History	67	B	B1		H2	H3	H4	2349
					TOTAL 1986 BRIDGE BOOKLETS. . .			7052

CIVICS BRIDGE, AGE 17

Subject Area	Booklet	Response Method†	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Civics	90	B	B1	C1	C9	C10	C11	1786
					TOTAL CIVICS BRIDGE BOOKLETS. . .			1786

† B = scannable booklet, C = circled-answer booklet
‡ Subject area background questions are included in the cognitive blocks for this booklet.

contained items that were used in the 1982 assessment of citizenship and social studies as well as several items that were newly developed for the 1988 assessment for the purpose of maintaining consistent timing across blocks of items.

In order to match the 1976 and 1982 assessment characteristics, age-only samples of students were defined using the 1976 and 1982 age definitions, 13-year-olds were tested in the fall, 17-year-olds were tested in the spring, and administration of the civics booklets was paced with an audiotape.

Bridge to 1984 (Reading and Writing, Booklets 51-56)

The 1984 bridge samples are comparable to the 1984 main assessment of reading and writing. Samples of students at grade 4/age 9, grade 8/age 13, and grade 11/age 17 (the same age and grade target populations assessed in 1984) were determined using the 1984 age definitions and were assessed during the same time of the year as in 1984. The assessment booklets (six at each grade/age) were administered without an audiotape.

At each grade/age, booklets 51 through 56 all contain a common background block (CC) and three cognitive blocks, either two reading and one writing, or one reading and two writing. All blocks are identical to those used in the 1984 assessment. Blocks C, D, E, and G are writing blocks; blocks H, J, K, L, M, N, O, P, Q, and R are reading blocks. At grade 4/age 9, block V, a double-length block, contains both writing and reading items. All writing and reading blocks also contain subject-related background and attitude questions.

The common background block was administered in approximately 15 minutes to grade 4/age 9 and six minutes to the other grade/ages; 14 minutes were allowed to complete each cognitive block (except for block V, which was allowed 28 minutes).

Bridge to 1986 for Ages 9 and 13 (Reading, Mathematics, and Science, Booklets 91-93)

These bridge samples are comparable to those used for the measurement of trends in reading, mathematics, and science in 1986. Three assessment booklets each were administered to samples of 9-year-old and 13-year-old students. The age-only samples were obtained using the same age definitions and times of testing as were used for both the 1984 main assessment and the 1986 bridge to 1984. The mathematics and science blocks were administered using an audiotape; the reading blocks were administered without an audiotape.

The contents of booklets 91 through 93 are identical to booklets used in the 1986 assessment. Each booklet contains a common background block (B1) and three blocks of cognitive items—one reading block (R1-R3), one science block (S1-S3), and one mathematics block (M1-M3). The cognitive blocks also contain subject-related background and attitude questions.

The common background block took approximately 15 minutes at age 9 and six minutes at age 13. The reading blocks took 13 minutes each at age 9 and 16 minutes at ages 13 and 17.

Bridge to 1986 for Grade 11/Age 17 (Reading, Mathematics, Science, and U.S. History—Booklets 61-67)

These bridge samples are comparable to the 1986 main assessment. Seven assessment booklets were administered to samples of grade 11/age 17 students. The grade/age samples were obtained using the same age definitions (*not* calendar year) and time of testing (spring) used for both the 1984 and 1986 main assessments. The booklets were administered without an audiotape.

Booklets 61 through 66 all contain a common background block (B1) and three blocks of cognitive items in various combinations of reading (R1-R6), science (S1-S4, S11), and/or mathematics (M1-M3, M9). Booklet 67 contains a common background block (B1) and three history blocks (H2-H4). The cognitive blocks also contain subject-related background and attitude questions. The booklets were constructed and cycled for administration using the BIB spiral design.

The common background block took five minutes of assessment time; each cognitive block took 16 minutes of assessment time.

4.3 QUESTIONNAIRES

In addition to the student assessment booklets, three questionnaires were administered to collect data about school characteristics, teachers associated with sampled students, and students excluded from the assessment.

The Teacher Questionnaires

NAEP gathered information on curricula and teaching methods from two distinct samples of teachers. The teacher questionnaires were administered to a sample of the reading teachers of fourth-grade students who were assessed in reading for the main assessment and a sample of the writing teachers of eighth-grade students who were assessed in writing for the main assessment. (The method by which teachers were sampled is described in Chapter 3.) These teachers completed a questionnaire that surveyed years of teaching experience, course curricula, use of classroom time, instructional practices, homework assignments, and teaching materials used.

Note: The purpose of this questionnaire was to collect additional information about <u>students</u> by gathering information about their teachers, not to describe the attributes of the teacher population.

The School Characteristics and Policies Questionnaire

The school characteristics and policies questionnaire was completed by the school principal or his or her representative for every school included in any of the 1988 samples. The questionnaire was used to gather information about school administration, staffing patterns, special programs, subject requirements, and school resources.

The Excluded Student Questionnaire

This questionnaire was completed by school personnel for every student selected for inclusion in the NAEP sample who was unable to respond to items because he or she was classified by the school as being limited in English language proficiency, mildly mentally retarded (educable), or functionally disabled. The questionnaire was used to gather information about special education, language, and other student programs.

Chapter 5

FIELD ADMINISTRATION

Nancy Caldwell and Renee Slobasky

Westat, Inc.

As a subcontractor to ETS, Westat, Inc., was responsible for field activities that included contacting selected districts and schools, administering the assessment sessions, and delivering completed booklets and questionnaires to ETS. This chapter summaries the Westat field organization and operations for the 1988 assessment. Details of field administration activities are available in the Westat *Report on Field Operations and Data Collection Activities—NAEP 1988* (Caldwell, Moore, & Slobasky, 1989).

This chapter begins with an overview of the field organization, followed by discussions of the training of supervisors; procedures for gaining cooperation of districts and schools; the supervisors' responsibilities, including making arrangements for the assessment, sampling, conducting the assessments, and collecting questionnaires; and the results of the assessments.

5.1 ORGANIZATION OF THE FIELD OPERATION

The field operation was organized around a core group of home office and field staff supplemented by additional staff in preparation for and during the winter and spring assessment. The core home office staff consisted of the field director and assistant field director. Throughout the study, the field director coordinated all activities in the home office related to field operations. The assistant field director coordinated the distribution of materials and the receipt of reporting forms. During the fall, the supervisors were assigned for routine reporting to the assistant field director. When the size of the field staff tripled for the winter/spring assessment, two more assistant field directors were added to the home office staff to share telephone-reporting and trouble-shooting responsibilities.

The sample of schools for the 1988 assessment was selected from eligible schools in selected geographic areas. These areas were a county or group of counties, called primary sampling units (PSUs). The sample of areas consisted of 94 PSUs, including a core group of 64 PSUs used for the fall bridge samples plus 30 additional PSUs for the winter and spring samples. The field staff was similarly organized around a core group of 11 supervisors who were responsible for fall pre-assessment and assessment activities in the 64 core PSUs. Pre-assessment activities in the 94 PSUs for the winter and spring assessments were conducted by 37 supervisors, including the original 11, in the late fall after the awarding of the new contract.

For the winter and spring, when the number of PSUs increased and the workload in terms of both the number of schools and students to be assessed within each PSU also increased significantly, supervisory regions were rearranged and expanded in number from 11 to 37. During the fall, each supervisor was responsible for assessments in an average of 5-6 PSUs. During the winter and spring, supervisors worked in an average of 1-2 PSUs.

Of the 11 fall supervisors, six had worked on NAEP during at least one of the past assessments. With the awarding of the new contract and expansion of work for the winter and spring, 26 additional supervisors were required. To fill these positions, we rehired five supervisors who had worked on NAEP in the past. To find the remaining supervisors for the winter/spring, we first searched Westat supervisory files, and then recruited from outside the company.

During the field period, supervisors were replaced in four of the 37 supervisory regions. Two of the supervisors resigned because of personal problems, the other two because the position was more demanding than they had anticipated.

Each district supervisor was responsible for a variety of different tasks. During the fall pre-assessment phase, the 11 supervisors contacted school districts with schools scheduled to be assessed in the fall to follow up on introductory materials that had been mailed earlier explaining the assessment program. During these calls, supervisors also scheduled introductory meetings with representatives of the sampled schools. The purposes of the introductory meetings were to explain the program in greater detail and to set a schedule for the assessment in each school. While in the area conducting meetings, the supervisors recruited exercise administrators to work with them in administering the assessment sessions.

Except for the introductory materials mailed by ETS, no contact was made with districts containing schools scheduled only for the winter and/or spring until the new contract was awarded in September, 1987. In October, 26 supervisors and two alternates were hired and trained. They followed the same process described above in contacting the winter/spring school districts and conducting introductory meetings during the months of November and December, 1987. Also during this time, the 11 fall supervisors who were in the process of conducting the fall assessments expanded their schedules to include the contacts with the districts in their regions in which there were winter/spring schools.

5.2 SUPERVISOR TRAINING

The 11 fall supervisors came to Bethesda, Maryland, for a four-day training session from September 2-5, 1987. Also in attendance were representatives from the ETS Princeton and regional offices who were there to help the supervisors gain cooperation from schools and districts. The training was conducted by the Westat project director and field director. ETS Princeton office staff also made presentations and provided explanatory notes throughout the session.

With the awarding of the new NAEP contract, 26 additional supervisors and two alternates were hired and brought to Rockville, Maryland. for training on October 26-27, 1987. This training session, conducted by the Westat project director and field director, focused only on scheduling and conducting introductory meetings for the winter/spring assessments.

To assist supervisors in conducting their introductory meetings, a slide presentation with a script for the supervisor to read was developed and implemented for the first time during the fall introductory meetings for the 1988 assessment. The slides and script formalized the supervisors' presentations on background information of National Assessment, the general stages of sampling schools and students, the role of the school in the assessment, criteria for student eligibility, and questionnaires to be completed by school personnel. Supervisors reported that the slide presentations were well received at introductory meetings attended by several school representatives. However, they often chose not to use the presentation during small, one-on-one meetings.

Procedures for conducting the winter/spring assessments were the focus of a three-day training session attended by all supervisors on December 16-18, 1987.

5.3 OBTAINING COOPERATION OF SCHOOL DISTRICTS AND SAMPLE SCHOOLS

5.3.1 Preliminary Contacts

During June, July, and August 1987, while Westat was recruiting supervisors and developing materials, ETS was making preliminary contacts preparatory to obtaining school cooperation. The schedule of these preliminary activities is outlined below and discussed in more detail following the outline.

Date	Contact Made By	Activity
June 29	ETS	An initial letter was mailed to Chief State School Officers informing them that schools within their states had been selected for NAEP.
July 7	ETS	A second letter was sent to Chief State School Officers containing a list of the school districts and private schools selected within the state.
July 22	ETS	An initial letter and a NAEP report were mailed to superintendents of public and parochial schools and principals of private schools selected for the entire 1988 assessment within the state.

August 5	ETS	A followup letter and NAEP materials were mailed to superintendents and private school principals. The materials included • a cover letter explaining that the supervisor would contact them to discuss NAEP and schedule an introductory meeting; • a list of the selected schools in the official's jurisdiction; • a fact sheet on NAEP; and • examples of assessment items.

As can be seen from the outline, recruiting of schools for NAEP actually began in June, once the sample of schools had been selected and their corresponding school districts identified. ETS contacted the Chief State School Officers in each state and asked them to notify the school district superintendents. In July, ETS sent a letter to the superintendents and heads of private schools inviting their participation. Informational materials on NAEP and a list of the sampled schools in the district were also sent. These initial contacts, which were completed prior to supervisor training, paved the way for the telephone contacts to follow.

Once the supervisors and ETS regional office staff had been trained, they began working to obtain cooperation. The schedule of these contacts follows.

Date	Contact Made by	Activity
Sept. 8-Oct. 9	District supervisors for fall assessment, with ETS assistance	Calls were made to districts with schools selected for fall assessments. The calls were made to introduce the 1988 assessment, establish participation, and make arrangements for introductory meetings. The supervisor filled out the Introductory Meeting Form(s) and the School Update Form and summarized the conversation in the Results of Contact form. Copies of the forms were sent to the home office. Districts that had schools selected for winter/spring as well as fall assessments were told that further contact concerning winter/spring assessments would occur later.

Sept. 8- Oct. 9	ETS	Calls were made to districts with four or more schools selected only for winter/spring assessments, and were told that the supervisor would contact the district in October-November.
Sept. 8- Oct. 9	Westat home office	A confirmation letter was sent to superintendents after the Introductory Meeting Form was received from the supervisor. Principals of selected schools within the superintendent's district were sent a package containing • a cover memo introducing the study and confirming the meeting date; • a memo to the principal giving the assessment schedule and outlining the school's role; • the principal's questionnaire; and • other informational materials (the same as those sent to superintendents).
Sept. 14- Oct. 9	Fall district supervisors and ETS	Introductory meetings were conducted for schools in the fall assessment.
Oct. 28- Dec. 4	District supervisors for winter/spring assessment, with ETS assistance	Calls were made to districts with schools selected for winter/spring assessments. The supervisor filled out the Introductory Meeting form, the School Update Form, and the Results of Contact form. If the district had schools in the fall assessment, the primary purpose of the call was to make arrangements for introductory meeting with winter/spring schools; the School Update Form would have already been filled out by the fall supervisor.
Oct. 28- Dec. 15	Westat home office	A confirmation letter was sent to superintendents when the Introductory Meeting Form was received from the supervisor. Principals of selected schools in the superintendent's district were sent a package containing information similar to that sent to principals with schools in the fall assessment.

Contacting districts to solicit cooperation and schedule the meetings was primarily the responsibility of the district supervisor, unless we had some reason to believe that getting the cooperation of a particular district or school was going to be a problem. In those cases, ETS regional or national staff made the initial contacts. An ETS representative also made the initial call to all districts containing four or more schools in the assessment.

During these calls to establish cooperation and to set up the introductory meeting, supervisors also updated our information on schools in the district. Topics raised with school district staff included school creation and closing, and changes in enrollment and grade span. Three forms were used to record this information: the School Update Form, the Introductory Meeting Form, and the Results of Contact form. The originals of these forms were mailed to the home office and then used as the basis for mailing packages of materials to the persons scheduled to attend the meeting. Information from the School Update Form was used to revise and update information in the home office files on the schools in the sample. New schools identified were also given a chance to be selected for the study.

There were basically two waves of telephoning and introductory meetings—one for districts with schools in the fall assessment and another for districts with schools only in the winter/spring. For the districts that had schools in the winter/spring as well as the fall assessment, a second telephone call was made to arrange for an introductory meeting for winter/spring schools, and to obtain any updated information about schools in the district not obtained by the fall supervisor.

5.3.2 Schools Added to the Original Sample

Due to a variety of factors, described in *The 1988 National Assessment of Educational Progress—Sampling and Weighting Procedures, Final Report* (Rust, Bethel, Burke, & Hansen, 1990), it was sometimes necessary to add schools to the original sample. Since the process of adding schools did not begin until October, while introductory meetings were taking place, the procedures for contacting and gaining cooperation from these schools necessarily differed from those described for the original sample. For the added schools, ETS first mailed a letter to the district superintendents and heads of private schools. Then, the district supervisor telephoned the contact person in the superintendent's office and asked him or her to notify the sample schools. If convenient, the supervisor then met with school representatives in person. If it was not possible to meet in person because of scheduling problems, the supervisor conducted the introductory meeting by telephone. ETS regional and national staff provided assistance as needed in contacting districts and schools.

5.3.3 Results of School Cooperation Effort[1]

There were 1,776 schools originally selected for one or more grade/age samples in 1988. This is comparable to the 1,682 schools selected for 1984, but less than the original sample for 1986 of 2,309. The 1986 sample was made larger to accommodate inclusion of the language minority component and because of the number of different subject areas assessed that year.

Table 5-1 presents the results of the school cooperation effort. Of the 1,776 schools originally selected for the 1988 assessment, 1,412 cooperated. An additional 88 would have cooperated, but they did not have any eligible students enrolled and 230 refused. The other 46 schools were closed or out-of-scope.

The cooperation rate (86.7 percent) for the 1988 assessment was about the same as in 1986 but less than in 1984 (88.1 percent).

As Table 5-2 indicates, the school cooperation rate varied among the three age classes and, more noticeably between the main NAEP and bridge assessments. For the main NAEP assessments, the rate varied from 88.7 percent for grade 4/age 9 schools to 82.8 percent for grade 12/age 17 schools.

The bridge assessments for age class 13 were held in the fall of 1987 and had the highest school cooperation rate (92.7 percent). The bridge assessments for age class 9 were held in the winter and had a school cooperation rate of 87.2 percent. The bridge assessments for age class 17, held in the spring of 1988, had a school cooperation rate of only 78.1 percent.

5.4 OVERVIEW OF THE SUPERVISORS' MAJOR TASKS

During the assessment phase of the project, the supervisors were responsible for carrying out the tasks listed below.

- Recruit and train exercise administrators.
- Check each shipment from ETS for quantities of assessment booklets and excluded student questionnaires and, in the winter/spring, teacher questionnaires. Check shipment from Westat for the other supplies needed.
- Review the Session Assignment Forms for the PSU and formulate a plan for scheduling assessment activities in each school.

[1]The summary numbers presented in Table 5-1 refer to individual <u>schools</u>, whereas Tables 5-2, 16-4, and 16-5 refer to grade/age samples. Since a school may be selected for more than one grade/age sample, the school appears in the counts for <u>each</u> of the appropriate assessment samples in Tables 5-2, 16-4, and 16-5.

Table 5-1

Summary of NAEP 1988 School Participation

	Number of Schools
Total original sample	1,776
Out-of-range or closed	46
No eligibles enrolled	88
District refused	149
School refused	81
Cooperating	1,412
Cooperation rate*	86.7%
[1986]	86.4%
[1984]	88.1%
Replacement sample for refusals	86
Out-of-range or closed	4
No eligibles enrolled	0
District refused	24
School refused	11
Cooperating	47

Table 5-2

NAEP 1988 School Cooperation Rate* by Age Class and Type of Assessment

	Number Cooperating	Number of Refusals	Cooperation Rate
Age class 9 main NAEP	327	42	88.7%
Winter bridge	154	23	87.2%
Age class 13 main NAEP	399	68	86.6%
Fall bridge	173	16	92.7%
Age class 17 main NAEP	304	66	82.8%
Spring bridge	114	34	78.1%

$$* \text{ Cooperation rate } \quad \frac{\text{Cooperating + no eligibles enrolled}}{\text{Cooperating + no eligibles enrolled + refused}}$$

- Contact each school one week before the assessment week and confirm the exact days for the assessment; remind the school coordinator that the Student Listing Forms must be completed before the supervisor arrives.

- Contact each school two days before the sample selection was to begin to be sure the school was prepared.

- Contact exercise administrators to inform them when and where to meet on the first day in a PSU.

- Collect school characteristics and policies questionnaire from each school.

- Complete the sample selection of students in each school a few days to a week before the assessment began. Arrange with the school coordinator the exact time and place for each session and complete an Administration Schedule for each session.

- Prepare an excluded student questionnaire for each student who had been sampled from the Student Listing Form but had been excluded from the assessment by the school. Give the school coordinator the excluded student questionnaires.

- Supervise and coordinate the assessment process.

- In the winter/spring, complete sample selection of teachers for the teacher questionnaires and distribute the teacher questionnaires.

- Perform quality control checks on each exercise administrator.

- Collect, or have the exercise administrator collect, completed excluded student questionnaires (and teacher questionnaires in the winter/spring).

- Complete Roster of Questionnaires for school characteristics and policies questionnaire and excluded student questionnaire and, as appropriate, the Teacher Survey Roster for the teacher questionnaire.

- On a daily basis, take back completed work from exercise administrator and distribute new materials for the next day.

- Make arrangements for makeup sessions as needed.

- Review each booklet to make sure that all coding was accurate. Compare the front cover to the Administration Schedule and make any corrections needed.

- Complete the School Worksheet.

- Ship materials to ETS and mail appropriate reports to Westat.

- Call the field director <u>immediately</u> if there was any problem completing the assessments in a PSU.

5.5 MAKING ARRANGEMENTS FOR THE ASSESSMENTS

To make arrangements for the assessments, the supervisors worked with the schools to accomplish the following:

- Schedule the date and time for the assessment sessions.

- Make sure that appropriate space would be available.

- Identify a school staff member to be the school coordinator who would work with the supervisor on assessment arrangements.

- Arrange for the necessary number of exercise administrators and, if necessary, additional school staff to monitor the sessions.

- If the school required any parental notification or permission, make sure that the appropriate letters or forms were distributed to the students.

- Select the sample and prepare all assessment materials.

5.6 SELECTING THE SAMPLE OF STUDENTS

At the introductory meetings, schools received instructions and forms to be used to prepare lists of eligible students. The instructions specified who should be listed and the information (birth date, gender, race/ethnicity, and grade) to be provided. The forms were to be used if the school chose to do so. Since experience in previous years showed that an increasing number of schools could, and preferred to, produce computer-generated lists of eligible students, a set of instructions for computer listing of students was also made available. The instructions made it clear that computer lists were perfectly acceptable as long as the necessary information on each student was included.

Two weeks prior to a school's assessment date, the supervisor contacted the school coordinator to make sure that the lists of eligible students were prepared and that all arrangements were set as agreed. The supervisor then visited the school (or district office) to select the sample of students to be assessed. The time interval between the selection of the sample and the assessment varied, depending on several factors; however, the average elapsed time was about a week. Schools with large assessments and those requiring parental notification generally preferred that the sampling be done as early as possible to give them time to make final arrangements.

In scheduling sample selection, the supervisors had to balance several, often competing, concerns. Their own time constraints and travel schedule limited their flexibility in scheduling sampling. Also, the schools' preference for early sampling had to be balanced against the importance of having an up-to-date sample. The more time that elapsed between sampling and assessment, the more likely that the sample would include students who had dropped out and that it would exclude newly enrolled students. These problems were most severe in the large urban high schools.

Once the supervisor had reviewed the lists for completeness and accuracy, he or she selected the sample following the instructions on the Session Assignment Form. In most cases the supervisors worked from lists prepared by the school and did the sampling in the schools. In some instances, districts prepared computer lists for their schools and some of these districts required that the sampling be done in the district office rather than at the school.

Because of the complexity of the sampling, supervisors were required to do all sampling themselves, although they could use exercise administrators to help check the sampling lists and to fill out forms. Having exercise administrators participate in the sampling was very helpful to the supervisors and expedited the sampling process.

Once sampling had been completed, the supervisors and exercise administrators could make out the lists of students to attend each assessment session. The supervisor reviewed the plans for the assessment with the school coordinator before making out the lists or Administration Schedules. At this time, and for the first time, the exact number of students sampled was known. Working with the coordinator, the supervisor updated the plans for the assessment and determined which students would go to which sessions. Sometimes the coordinator had very specific ideas about the organization of the assessment. In elementary schools, for example, the usual preference was for all students in a particular teacher's class to be assessed together and, if possible, in their own classroom.

The supervisor had a great deal of flexibility in arranging spiral assessments in main NAEP schools to suit the needs of the school. There was not as much flexibility with the bridge spiral or tape sessions because each student was sampled for particular type of session had to attend the session when that booklet type was administered.

Once the arrangements had been set and the Administration Schedules filled out, the supervisor distributed the excluded student questionnaires.

5.7 DISTRIBUTING AND COLLECTING NAEP QUESTIONNAIRES

The school characteristics and policies questionnaire, excluded student questionnaire, and teacher survey questionnaire were distributed in the schools to be completed by school personnel.

All schools were mailed the school characteristics and policies questionnaire by Westat prior to the assessment. This form was to be filled out by the principal or another staff member knowledgeable about the school's administrative policies and staff characteristics. The supervisors picked up the questionnaire when at the school for sampling or for the assessment.

An excluded student questionnaire was to be filled out for every student who was sampled for the assessment but excluded by the school. Schools could exclude students with limited English speaking ability and those who were mildly mentally retarded (educable) or functionally disabled, if in the judgment of school staff or if a review of the school records indicated that they were unable to take the assessment.

Main NAEP schools selected for the grade 4/age 9 and grade 8/age 13 assessments were asked to participate in a post-assessment teacher survey. The supervisor used the completed Administration Schedules to identify those fourth-grade students who completed reading booklets and eighth-grade students who completed writing booklets. To identify the teachers for the survey, the school coordinator was asked to identify the reading/English teachers of those students. These were the teachers selected for the survey. The supervisor gave the school coordinator a teacher questionnaire to distribute to each selected teacher. The supervisor (or exercise administrators) returned to these schools to pick up the teacher questionnaires a few days after the assessment.

The supervisor attempted to collect all the completed questionnaires on the assessment day except for the teacher questionnaires. If the questionnaires were not ready on the assessment day, and it was convenient for the supervisor or an exercise administrator to return to the school later to pick up the questionnaires, they would do so. Otherwise, the supervisor gave the coordinator a postage-paid envelope to be used to mail the forms to ETS.

5.8 PREPARING REPORTS AND SHIPPING MATERIALS

Once the assessments were finished in a school, the supervisor and exercise administrators edited the booklets, filled out the necessary forms and shipped the booklets and forms to ETS. A copy of all forms was sent to Westat so that progress in the field could be monitored.

5.9 RESULTS OF THE FALL ASSESSMENT

Table 5-3 shows data on the number of students who were sampled, invited and assessed during the fall assessment of 13-year-olds.

The original sample included 13,494 students. Of these, 806 students were sampled but excluded from participation by the school because of they had limited English speaking ability, were mildly mentally retarded (educable) or were functionally disabled. The rate of exclusion (6.0 percent) is somewhat higher than in the previous assessment, when it was 5.4 percent. The number

Table 5-3

Students Sampled, Invited, and Assessed
During the Fall Assessment

Number of Students	Bridge to 1984	Civics Bridge	Bridge to 1986 Booklets 91	92	93	Overall
	Sample Type					
Invited	6,189	2,165	1,534	1,377	1,423	12,688
Assessed	5,500	1,938	1,405	1,281	1,256	11.380
Percent Assessed	88.8%	89.5%	91.6%	93.0%	88.3%	89.7%

Table 5-4

Students Sampled, Invited, and Assessed
During the Winter and Spring Assessments

1988	Winter and Spring Age Class 9	Age Class 13	Age Class 17	Overall
Number sampled	36,961	44,118	60,833	141,912
Number excluded	2,253	2,213	2,065	6,531
Number invited	34,708	41,905	58,768	135,381
Number assessed	31,911	36,699*	46,170	114,780*
Percent assessed	91.9%	87.6%	78.6%	84.9%

* Number includes 5,098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences: An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

of students invited to the assessment was 12,688. Of these, 11,380 actually were assessed.

The overall attendance rate (89.7 percent) is the same as in 1986 (89.7 percent) and slightly higher than in 1984 (87.3 percent) and 1982 (85.5 percent).

5.10 RESULTS OF THE WINTER/SPRING ASSESSMENT

Table 5-4 provides information on the number of students sampled, invited to assessment, and assessed during the winter and spring.

Of the almost 142,000 students sampled for assessment, 6,531, or 4.6 percent were excluded by their schools. Of those students invited to assessment, just under 85 percent were assessed. During the 20 weeks of the winter and spring assessment, 114,780 students were assessed. Overall, the assessment rates for the three age classes have remained relatively stable over the past few assessments.

Chapter 6

OVERVIEW OF 1988 NAEP MATERIALS PROCESSING AND DATABASE CREATION

John L. Barone

Educational Testing Service

Chapters 6.1 through 6.7 detail the receipt, processing, and final disposition of the 1988 NAEP assessment materials at ETS. These processes resulted in the integration of all respondent data collected during the 1988 assessment into a NAEP database—a database that ensures data quality and provides for efficient analysis and reporting.

Chapters 6.1 through 6.4 describe the methods used to transcribe the materials to computer-readable form. Chapter 6.5 discusses the quality control processes that were followed during data transcription and provides a summary of quality control error analyses. Chapter 6.6 discusses the integration of the transcribed NAEP data into the NAEP database/information system that was used for data analysis and reporting. Chapter 6.7 describes the database products that result from the NAEP information system.

This chapter describes the flow and evolution of the operational procedures used to process the 1988 NAEP data, and provides some detail on the amounts of materials that were processed.

Also described in this chapter are the three distinct data units that make up the integrated NAEP database:

1) the item information database, which contains information about every assessment item used in a 1988 assessment booklet;

2) the restricted-use data files, which contain all data collected from the 1988 NAEP respondents; and

3) the public-use data files, which contain a nonconfidential subset of the restricted-use data files, and are available to external users via the 1988 NAEP public-use data tapes package.

The flow of materials, creation of data files, and creation of the NAEP database are depicted as an ordered set of processes that are applied either to the assessment materials or to the transcribed data. Chapters 6.1 through 6 7 describe these processes in detail.

6.0.1 THE 1988 ASSESSMENT: SCOPE OF WORK

The scope of the effort required to process the 1988 assessment materials is evidenced by the following numbers.

For the 1988 assessment,

- more than *130,000* assessment booklets or questionnaires were received and processed.

This processing included

- optically scanning more than *1 million* double-sided pages from test booklets and questionnaires;

- professionally scoring more than *300,000* student responses from 130 open-ended items;

- manually key-entering and verifying more than *15,000* assessment booklets;

- using the NAEP minicomputer-based transcription system to track, audit, edit, and resolve more than *22 million* characters of information;

- selecting and comparing a quality control sample of more than *160,000* characters of transcribed data to the actual responses in assessment booklets;

- cataloging more than *1.5 million* characters of information on a total of 3,800 assessment items and derived variables, as part of a comprehensive item information database;

- developing a public-use data tape package containing more than *160 million* characters of useful information.

These numbers alone indicate the staggering size of 1988 NAEP materials processing and database operations. However, the full extent of this effort becomes clearer when one considers that over 90 percent of the data transcription activities described in this chapter were completed within six months, with a conservatively estimated accuracy rate of fewer than 2.5 errors for every 10,000 characters of information transcribed.

6.0.2 NAEP DATA PROCESSING SYSTEMS

Materials processing and database creation for the 1988 assessment closely paralleled the processes used in the 1984 and 1986 assessments. This allowed the use of in-place, proven operational procedures and computer systems. This fact will be emphasized and highlighted throughout the following chapters.

A major improvement in the 1988 assessment was the introduction of scannable answer sheets. This innovation greatly reduced the amount of paper handling and scanning required to process assessment materials. The NAEP data systems were adapted to accept this type of answer document.

The large volume of collected data and the complexity of the NAEP design, with its spiraled distribution of many booklets, required the development and use of NAEP-specific data entry and management systems, including carefully planned and well-defined editing, quality control, and auditing procedures. This chapter discusses the original 1984 design and implementation of these systems, and the adaptation and use of these systems and processes as applied to the 1986 and 1988 assessments. The result was effective, responsive data management procedures that ensured the quality and integrity of NAEP data. And, a NAEP database that met the original objectives of integrity and usefulness, while exceeding stringent standards for accuracy and quality.

Types of Assessment Booklets and Answer Documents Used in NAEP

NAEP data processing flow and systems are determined to a large extent by the type of assessment booklets and answer documents that need to be processed. As a consequence of the 1988 NAEP design, three types of answer documents were used and processed for the assessment. key-entered booklets, scannable booklets, and scannable answer sheets. The evolution of these response documents is described below.

In the 1984 assessment, students circled their responses in their test booklets. These responses were then manually transcribed, item by item, to a computer file. In 1986, because of a higher volume of data and a shorter time period for processing, a new type of booklet was introduced in which students filled in ovals to indicate their responses. Each page of these booklets was entered into a computer scanning device to create the data file. The NAEP data transcription systems were modified to accept the output of the scanning devices.

As part of the continual effort to improve the efficiency of NAEP data processing, the use of separate, scannable answer sheets was introduced in the 1988 assessment. Main sample students in grade 12/age 17 and grade 8/age 13 filled in ovals corresponding to their responses to multiple-choice items on a separate answer sheet, instead of in the test booklet. Because the answer sheets were only one or two pages in length—versus a test booklet, the length of which could be up to 30 double-sided pages—the amount of paper handling and scanning required to process these samples was substantially reduced. The NAEP data transcription systems were adapted to accommodate the new answer sheets in addition to the scannable booklet and direct entry methods that were already in place.

For those booklets that contained open-ended response items, students recorded their written responses on the answer document. Later, professional scorers at ETS scored the items and, depending on the type of answer document, wrote, circled, or filled in an oval for the students' scores on the answer document. The document was then transcribed to a computer file.

Students in some of the 1988 samples did not receive separate answer sheets. Main sample grade 4/age 9 students filled in ovals next to responses in the test booklet, because there was some concern that separate answer sheets would be confusing to them. Bridge sample students (students in samples selected to link 1988 results with past results) were given instruments that matched those used in the corresponding previous assessments.

In addition to the student assessment booklets, three questionnaires were administered to collect data about school characteristics, teachers associated with sampled students, and students excluded from the assessment. The excluded student questionnaire was a scannable document. The school and teacher questionnaires were manually transcribed, item by item, to a computer file.

6.0.3 PROCESS FLOW OF NAEP MATERIALS AND DATABASE CREATION

Figure 6-1 is a flow diagram that shows the conceptual framework of ordered processes that were applied to the NAEP materials and data files. The vertical line through the center of the figure divides the processes into two sets—processing assessment materials and database creation—both of which a e described below.

The processes represented by rectangular boxes in the flow diagram were performed at ETS on the paper materials or computer files. The three processes enclosed in rounded boxes (assembling the sample of schools, planning and conducting the field administration, and deriving the sampling weights) were performed by Westat and are discussed respectively in Chapters 3, 5, and 8. Two Westat reports, the *Report on Field Operations and Data Collection Activities—NAEP 1988* (Caldwell, Moore, & Slobasky, 1989) and *The 1988 National Assessment of Educational Progress—Sampling and Weighting Procedures, Final Report* (Rust, Bethel, Burke, & Hansen, 1990) discuss the field operations and sampling procedures in detail.

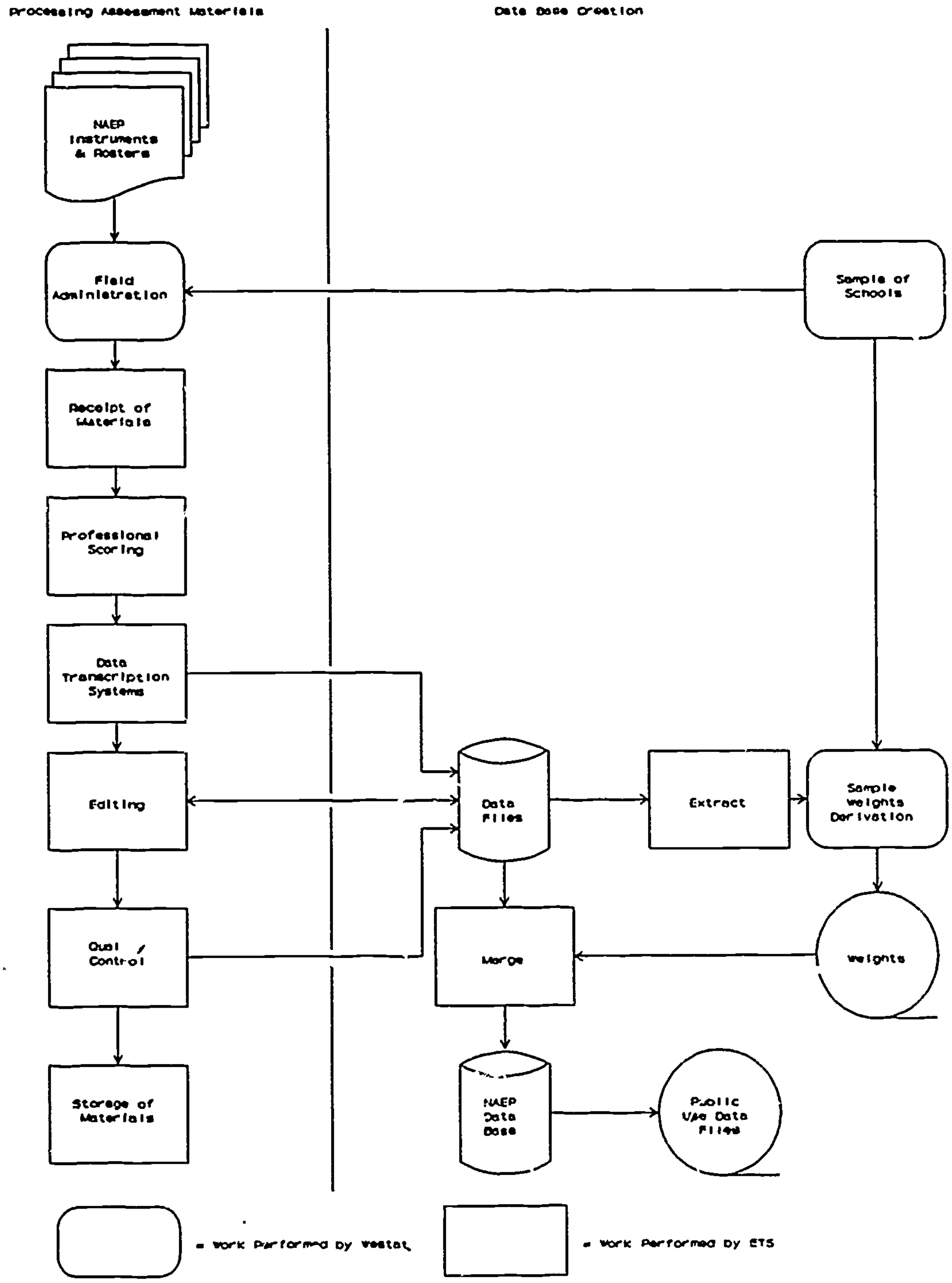
Processing Assessment Materials

The left side of Figure 6-1 depicts the flow of NAEP printed materials. Chapter 6.1 describes this flow in detail and discusses how information from the field rosters, schedules, and worksheets was used to control the processing of materials. The figure follows the path of each assessment instrument (student test booklets, school characteristics and policy questionnaires, teacher questionnaires, and excluded student questionnaires), absentee rosters, school worksheets, and administration schedules as they are tracked through the appropriate processes that result in the final integrated NAEP database.

The following is a brief description of the materials processing activities as shown on the left side of Figure 6-1. Each description refers

Figure 6-1

Data Flow Overview



the reader to the section(s) or chapter(s) in which the activity is discussed in detail.

Field Administration is the conduct and monitoring of the NAEP assessment in the schools. Chapter 5 summarizes this process.

Receipt of Materials refers to receipt and processing of assessment materials at ETS. Section 6.1.1 describes the procedures and forms that were used to check and verify the receipt of documents from the field. It also discusses the follow-up procedures that were initiated when discrepancies were identified and the subsequent assembling of NAEP materials for further processing and data transcription.

The *Professional Scoring* process was carried out for responses to open-ended items for reading, writing, civics, U.S. history, mathematics, and science. Chapter 6.2 describes the items, types of scoring used, scoring operation, reliability checks, and resolution of scoring discrepancies. Entry and editing of these data are discussed in sections 6.1.4 and 6.4.4.

Data Transcription Systems refers to the methods used to transcribe NAEP materials into computer-readable form. The transcription method used for each type of NAEP instrument is discussed in Chapter 6.1. Chapter 6.3 describes the design, structure, and development of the data entry system used to transcribe most of the NAEP materials to computer files; it also discusses the tracking and audit mechanisms that were built into the system to ensure that all data were properly processed and accounted for.

Originally implemented for the 1984 assessment, NAEP's data transcription system has proven to be accurate, efficient, and flexible. In 1984, manual key entry and verification was the primary method of entering data. In the 1986 assessment, the system was modified to accept scannable booklets as the main source of input. For the 1988 assessment, the system was modified to accept scannable answer sheets as a third method of data entry.

Editing refers to the ETS procedures that ensured the correctness and integrity of the NAEP data files by (1) validating every field of NAEP data that was entered into computer-readable form, (2) identifying any invalid or inconsistent values, and (3) correcting or flagging as unresolvable those values identified as invalid or inconsistent. Chapter 6.4 describes these procedures.

ETS *Quality Control* procedures were used to assess the accuracy of the data transcription and editing operations. Chapter 6.5 discusses the quality control procedures used in NAEP data processing and provides a summary of the likely error rates.

Storage of Materials refers to the final disposition of NAEP printed materials after processing had been completed. Chapter 6.1 discusses materials storage.

Database Creation

The right-hand side of Figure 6-1 depicts the evolution of the integrated NAEP database from the transcribed data to the final database, available to external users via the public-use data tapes. Chapter 6.6 describes the processes through which the database evolved.

The remainder of this section contains a brief description of each process involved in database creation as shown on the figure. Each description also refers the reader to the section(s) or chapter(s) in which the process is discussed in detail.

Sample of Schools refers to the process performed by Westat to select the schools to be included in the assessment. This process is discussed in Chapter 3.

Data Files refers to (1) the data files created by the ETS/NAEP data transcription, editing, and resolution systems and (2) the labeling files (discussed in Chapter 6.6) that contain descriptive information on every item used in NAEP.

Extract is the process (discussed in section 6.6.1) that created data files containing specific demographic data fields extracted from the ETS/NAEP data files. These files were required by Westat to derive sampling weights.

Sample Weights Derivation was performed by Westat and is discussed in Chapter 8. This process produced computer tape files containing sampling weights for every student and school assessed by NAEP.

Merge refers to the final integration of NAEP data files into the NAEP database. This process, discussed in section 6.6.2, merged the NAEP data files, labeling files, and sampling weights into one database.

NAEP Database is the final, integrated NAEP database that contains all 1988 NAEP data and is made available to external users via the public-use data tapes. The structure of the internal NAEP database is discussed in Chapter 6.6; the public-use data tapes, which contain all of the nonconfidential data fields from the internal database, are discussed in Chapter 6.7.

6.0.4 NAEP DATABASE: DESCRIPTION AND HISTORY

The 1988 NAEP data collection resulted in seven classes of data file (student, school, teacher, excluded student, sampling weight, item information, and derived variable files). These files are operationally merged into an integrated database that efficiently links student, school, and teacher data in ways that permit simultaneous analysis. The database system can operationally aggregate data while avoiding the necessity of creating intermediate files.

The structure and internal data format of the 1988 NAEP database is a continuation of the integrated design originally developed by ETS in 1983.

Structure of the NAEP Database

ETS has placed all NAEP information from the 1970-1988 assessments into three data systems that make up the complete NAEP database:

The *item information database* contains all of the descriptive, processing, and usage information for every item developed and administered for NAEP. This database functions as a resource for test development activities, data system control operations, and item linkage to past assessments. A complete description of the contents, documentation, and use of this database is found in *A Guide to the NAEP Item Information Database* (Rogers, Barone, & Kline, 1989) and *A Primer for the NAEP Item Information Database* (Rogers, Kline, Barone, Mychajlowycz, & Forer, 1989).

The *restricted-use data files* contain all NAEP respondent data, including "secure" files. These files function as NAEP data archives for responses from students, teachers, and school administrators from the booklets and questionnaires used in NAEP from 1970 to 1988. The contents, documentation, and use of these files for trend analyses are described in *A Guide to the NAEP Restricted-use Data Files* (Rogers, Barone, & Kline, 1989).

The *public-use data tapes* contain all nonconfidential respondent and item information for public dissemination. Their contents and use are documented in *The 1988 NAEP Public-use Data Tapes Version 2.0 User Guide* (Rogers, Kline, Johnson, Mislevy, & Rust, 1990).

These three database products, developed by ETS, are especially useful because

- They are *portable* and can be used on a variety of hardware systems. They can be accessed by a variety of software systems, including SAS and SPSS.

- They are in a *"rectangular"* file structure that is easy to understand and use. This data structure eliminates the need for complex data retrieval processes from dissimilar file formats.

- They have *standardized documentation*, including complete data layouts and codebooks.

- They are supported by *user guides* that have been published by ETS/NAEP.

Size of the NAEP Database

The complete NAEP database is massive. Spanning the years 1970 through 1988, NAEP has:

- collected data on the performance of over 1.6 million students and young adults, and background information about their teachers and schools;

- collected data on over 10,000 cognitive, background, and attitude items in 14 subject areas and on several special topics; and

- derived over 3,000 variables (composites, transformations, scale scores, sampling weights) and added them to the data files.

These data are maintained as part of the NAEP item information or restricted-use data files. Except for a small percentage of personal or secure item information, all data are made available to the public on the public-use data tapes.

While the selection of data from a given subject area or assessment year for analysis and reporting is straightforward (documentation, user guides, and data-variable codebooks are available for every NAEP data file), the following table puts the total size of the NAEP database into perspective.

NAEP Database	Approximate Size 1970-1986	1988
Item Information Database	81,000,000 bytes	4,000,000 bytes
Restricted-use Data Files	2,000,000,000 bytes	100,000,000 bytes
Public-Use Data Files	2,000,000,000 bytes	100,000,000 bytes

Chapter 6.1

PROCESSING ASSESSMENT MATERIALS

Alfred M. Rogers and Norma A. Norris

Educational Testing Service

This chapter describes the procedures through which NAEP instruments, schedules, and worksheets were received at ETS, and the methods used in the subsequent scoring, scanning, loading, editing, and resolution of NAEP data.

6.1.1 RECEIPT OF MATERIALS

At the completion of the assessment administration in each school, it was the responsibility of the Westat district supervisor to complete and mail a postcard to ETS containing the assessed school identification, the number of boxes shipped, and the mode of shipment. The receipt of this card at ETS alerted staff to expect arrival of the shipment within seven working days. If after seven days the shipment had not arrived, ETS staff were directed to notify Westat, who in turn would initiate a trace of the shipment.

The shipment from each school contained the school worksheet; administration schedule; questionnaire roster; school, teacher, and excluded student questionnaires; and assessment booklets, bundled by session. The format and content of these instruments are documented in the Westat Field Administration Report. The following discussion of check-in procedures presumes an understanding of the information contained in and the interrelationships among these instruments.

The school worksheet contained summary counts of the booklets used in all assessment sessions in each school. The booklets used within each session were counted and checked against the count written on the school worksheet. All discrepancies in the counts were referred to the administration schedules for resolution. The booklet numbers from the bundle in question were compared against the listing of booklet numbers on the schedule. If the discrepancy could not be resolved by this process, Westat was notified, and they in turn contacted the appropriate district supervisor for resolution of the discrepancy.

Two identification codes, the session code and the batch code, were then assigned to each column on the worksheet and to the corresponding bundle of booklets. The two-digit session code distinguished main sessions from bridge sessions and regular sessions from makeup sessions.

The use of a batch identification code was necessitated by the use of machine-scannable documents in this assessment. A preprinted, scannable header sheet was attached to each bundle of student booklets to be used to identify the bundle through all subsequent scoring, scanning, entry, and resolution processing. This batch header sheet was pregridded with a unique four-digit sequence code. As each header sheet was drawn from the pile, it was gridded with the age group code, the school and session codes, the current batching date, and the number of booklets to be processed. The age group code was either "N," "T," or "S" corresponding to the 9-, 13-, or 17-year-old cohorts. The batch identification code, which consisted of this age group code and the sequence number from the header sheet, was then recorded at the bottom of the session information on the school worksheet.

From the aspect of materials processing, a significant difference of the 1988 assessment from the 1986 assessment was the introduction of student instruments with removable, machine-scannable answer sheets. For the grade 8/age 13 and grade 12/age 17 cohorts these instruments were spiraled with the machine-scannable instruments whose format was used in the 1986 assessment (the grade 4/age 9 cohort were administered only these machine-scannable instruments). The bundles of student materials for each spiral session for these older cohorts thus contained a mixture of scannable booklets and scannable answer sheets.

The scanning machine program, which had been written for the 1986 assessment, was used again to scan the booklets; a new program was written for scanning the answer sheets. Since these programs could not be run concurrently, the bundles of instruments had to be split into separate batches of booklets and answer sheets, and a batch header sheet assigned to each. In order to maintain the integrity of each session throughout the processing phase, the same batch identification number had to be assigned to each pair of booklets and answer sheet bundles from the same session.

The teacher questionnaires and the excluded student questionnaires were then counted and compared against the questionnaire roster. All discrepancies in the teacher questionnaires and the excluded student questionnaire counts were referred to Westat and again, in turn, to the district supervisor for resolution. As the field administration procedures permitted a separate shipment of these questionnaires, the questionnaire roster listed questionnaires not included in the shipment, alerting the receiving staff to expect a later shipment.

If the supervisor was unable to collect the questionnaires on the day of the assessment, a pre-addressed envelope was left at the school so that the school coordinator could mail the questionnaires directly to ETS. There was no other follow-up activity to obtain uncollected questionnaires from school personnel; efforts to encourage school cooperation were focused primarily on student assessment activities.

When all of the student-related materials for a school had been received and checked in, the administration schedules, school worksheet, assessment booklets, and questionnaires were forwarded to the data operations coordinator for transcription processing. The operations coordinator separated these materials according to the appropriate data entry procedures: the administration schedules, the school worksheet and the teacher and school questionnaires were sent directly to data entry systems; the excluded student questionnaires were accumulated and shipped in batches to the optical scanning area; and the assessment session bundles were forwarded to the professional scoring area.

The absentee data from the administration schedules and the school worksheet data were entered into the data entry system on a daily basis. The teacher and school questionnaires were batched and held for data entry until scheduling permitted.

6.1.2 SCHOOL WORKSHEETS

Each column of the school worksheet contained information pertaining to the administration activity of each session within a school. This information included the date, time, and location of the administration, the exercise administrator code, and the counts of the students who were sampled, those who were absent, and those who were assessed. Additionally, each column contained a session code and batch identification code that were recorded by receipt processing staff. This information was entered into the system by selecting the first option on the main data entry menu (Figure 6.1-1).

Figure 6.1-1

Main Menu for the NAEP Data Entry System

```
                N A E P   E N T R Y   S Y S T E M   M E N U

                          OPTION: _

                 1  School Worksheet Entry

                 2  Student Data Entry/Verification/Resolution

                 3  Questionnaire Data Entry/Verification/Resolution

                 4  Absentee Data Entry

                 5  Questionnaire Roster Entry

                 X  Quit

   Enter Option Code:
```

The worksheet entry program received its input through two entry screens. The first entry screen (Figure 6.1-2) requested school-level information, namely, the PSU and school codes and the total number of assessment sessions that were conducted in that school. This count was further broken down into the four types of session administration: regular spiral, makeup spiral, regular bridge, and makeup bridge. The program would then display the second entry screen (Figure 6.1-3) once for each session, requesting the session-level information. When all sessions for a school had been entered, the program would redisplay the first entry screen, ready to process the next worksheet. The operator could either enter new information or press ENTER to return to the main menu.

The assessment session was the primary unit at which the entry system controlled the processing of student data and maintained statistics on data entry activity. A separate tracking file was established for this purpose, each record of which contained all control and reporting information for one session. The entry of the school worksheet information thus generated a new record on the tracking file for each session, setting initial values for those parameters that would control entry processing and record entry events.

The operations coordinator was provided with procedures for periodically monitoring and reporting activity on the data entry system. These procedures compared the counts of booklets processed at each stage with the initial counts from the worksheet, and flagged discrepancies. This, in turn, alerted the coordinator to possible missing or extra booklets. If the school worksheet information was determined to be in error, the operations coordinator had the facility to correct the tracking file data to prevent reappearance of the discrepancies in the activity report.

The school worksheets were retained by the operations coordinator in anticipation of later queries, since they could be stored conveniently and referenced easily.

6.1.3 ADMINISTRATION SCHEDULES

The administration schedules contain the demographic characteristics of the students selected for the assessment. This information, which included the gender, race/ethnicity, grade, and birth date of the sampled students, was used by Westat in the derivation of sampling weights. The booklet numbers of the students who participated were transferred to the schedule at the time of the assessment, and the demographic information was in turn transferred to the front covers of the booklets after the assessment.

The demographics of the students who were sampled but did not participate in the assessment (exclusions and absentees) were used to adjust the sampling weights of the students who did participate. The excluded student information could be obtained from the excluded student questionnaire data, but the information on absentees could be found only on the

Figure 6.1-2

First School Worksheet Data Entry Screen

```
                         SCHOOL WORKSHEET

                             PSU #: ___

                          SCHOOL #: ___

                TOTAL NUMBER OF SESSIONS: __

          NUMBER OF SPIRAL SESSIONS (0-10): __
   NUMBER OF MAKEUP SPIRAL SESSIONS (0-10): __
 NUMBER OF ORIGINAL TAPE SESSIONS (0,1,2): _
   NUMBER OF MAKEUP TAPE SESSIONS (0,1,2): _
```

Figure 6.1-3

Second School Worksheet Data Entry Screen

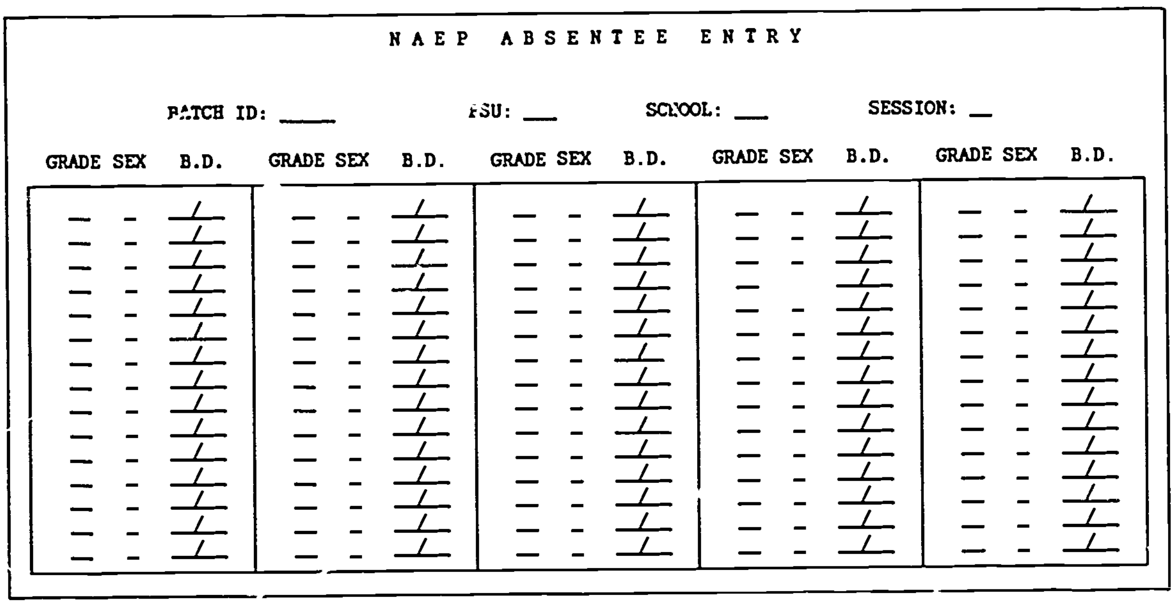
```
                         SCHOOL WORKSHEET

       PSU #: ___
    SCHOOL #: ___
                     TAPE/SESSION #: __

                               DATE: __/__/__
                               TIME: __:__
                      EA'S INITIALS: ___
                            EA'S ID: ____

                  # TO BE ASSESSED: ___
                        # ASSESSED: ___
                          # ABSENT: ___

                      BATCH NUMBER: _____
```

administration schedules. It was therefore necessary to transcribe the information for these absentees to computer-readable media and combine it with the assessed and excluded student data.

The absentee data were entered into the system by selecting the fourth option on the main data entry menu. The system then presented a screen for the entry of all absentee information for a single session (Figure 6.1-4). The program first requested entry of the batch identification number and the PSU, school and session codes. The batch code was used to locate the corresponding session record in the tracking file and the remaining codes were checked for correctness against the corresponding fields in the tracking record.

Figure 6.1-4

Absentee Data Entry Screen



The remainder of the screen could accommodate data entry for 75 absent students. Only three data fields were required for each absentee: grade, gender and birth date. These data were ultimately used by Westat to adjust the sample weights. As each field was entered, the program checked for appropriateness of range according to the age cohort and session type. At the completion of data entry for an assessment session, the operator pressed ENTER and the program would present a fresh screen for the entry of another session's complement of absentee data. The operator could then either enter another batch identification code or press ENTER again to return to the main menu.

If the operator entered a session code for which absentee data had already been entered, the program would display the data for all absentees in the record for that session at that time. The operator then had the facility to correct, delete, or add new data. When the ENTER key was pressed, the data displayed on the screen were transcribed to the current session record and the record was rewritten to the absentee data file.

The assessment schedules were retained by the operations coordinator in anticipation of future questions about and references to the sample. As the schedules for all three grade/age assessments could be contained in three storage boxes, their retention proved to be the most efficient and compact means of referring to the relevant raw data.

6.1.4 STUDENT ASSESSMENT INSTRUMENTS

The student assessment booklets were forwarded directly to the scoring area as the complete set of materials was received from each school. The booklets and answer sheets were batched separately for each session, with a batch header sheet attached to the top of each bundle. This preprinted, scannable sheet contained the PSU, school, and session codes, and a unique batch identification code to identify each session. The header sheets were retained with the batches throughout entry processing.

6.1.4.1 Professional Scoring

The batches of student booklets and answer sheets were sent from the receipt processing area to the scoring area where the open-ended reading, writing, mathematics. and civics items were read and scored by trained readers. The procedures and guidelines followed in scoring these items are more fully described in Chapter 6.2.

Each open-ended item was provided with a set of scannable ovals to be filled in by the reader. The ovals were usually at the bottom of the page on which the item was printed to avoid distracting or confusing the student. All open-ended items were provided with an extra set of ovals to permit secondary scoring of the primary trait scores for interrater reliability analysis. Several of the reading and writing items that were to be evaluated for secondary traits had an additional set of ovals for each secondary trait score.

The primary reader would examine each booklet in a batch and determine if it contained any open-ended items. If so, the reader wrote in his or her identification code and gridded in the first column of ovals in the reader identification area on the inside front page. The reader would then locate and read each of the open-ended items for that booklet and grid the first primary trait score and all secondary trait scores into the appropriate ovals. On every fifth booklet read, the reader would place a piece of tape over each set of primary trait score ovals and designate this book for secondary scoring. The sampling rate of one in five assured a 20 percent rate of secondary scoring, and the tape was a device to avoid influencing the secondary reader. The completed booklets were stacked in the same order in which they were received and the completed batch was placed in a designated area.

The secondary reader selected the designated booklets from each batch and entered his or her identification code into the second column of the reader identification area. The reader then located the items with concealed primary trait scores, read and scored them, and removed the pieces of tape. The completed batch was placed in a second holding area, whence it would be forwarded to the scanning processing area.

6.1.4.2 Scanning

The heart of the scanning process was a programmable computing machine that could "read" the pencil marks from both sides of a sheet of paper at a very high speed, convert those readings to "response" codes, and transcribe those codes to a magnetic computer tape in specified format. This section addresses the functions that constitute the body of the scanning process: the preparation of the scannable materials, the operation of the scanning machine, and the activity of the scanning machine operator.

Before the batches of scored booklets and scored answer sheets were sent to the scanning area, they were grouped by age cohort and placed into "capsules" that were then arranged sequentially on "carts." The capsules were cardboard boxes with one side open to facilitate access by the scanning and resolution staff, and the other side equipped with hangers to permit removal from the carts. The carts were transportable, two-sided hanging shelves with sloping sides to permit the capsules to hang with the open sides out while keeping the documents in. The carts were shipped to the scanning area.

The first step in the scanning process was to separate each booklet into its component pages for single-sheet processing by the scanner. Each booklet was secured by three staples along the left edge. Each answer sheet was a large, single sheet of paper folded along the left edge. The timing marks for the scanner were also printed along this edge. Two special machines were used to cut off the stapled edge without damaging the timing marks. The cutting machine could cut three or four booklets at a time but required a slower, manual setup process. The slitting machine was more automatic, processing one booklet at a time, but was less precise than the cutting machine. Careful handling of these booklets was imperative once they were cut, as the scanning program depended on the correct sequencing of pages within each booklet. The cut booklets were placed back in their capsules and the completed cart sent to the scanning machine.

The scanning machine operator first determined which scanning program to be used according to the age cohort and instrument type (booklet or answer sheet), mounted a fresh magnetic tape on the machine's tape drive, and started running the program. Scanning was initiated by placing the sheets from the first capsule into the input hopper of the scanning device. The scanner then read both sides of each sheet and placed it into one of two hoppers. If no errors in readability or sequencing were detected, the sheet went into the output hopper and the next sheet was read from the input hopper. If an error was indicated, the sheet was diverted into the shunt hopper, the program wrote a message to the operator's console, and the scanner stopped processing while the operator took appropriate action.

Each page of every booklet and answer sheet had been printed with a set of identification marks next to the timing marks. The front cover of each booklet within a numbered type was assigned a unique set of these marks, and the pages within each item block type were similarly identified by block code and sequence number. As the scanner read a cover sheet, the program identified the booklet number and referred to an internal table to determine which blocks should follow and which page formats should appear within each block.

If the program indicated a page sequence error, the operator instructed the program to treat the page as missing and placed the shunted page into the input hopper to be read again. If a page within a block was unreadable, the operator instructed the program to treat it as missing and placed the sheet perpendicularly on top of the output stack. If a block sequence error or unreadable booklet cover was indicated, the operator instructed the program to insert a dummy record and removed the remaining pages of that booklet and placed them perpendicularly on top of the output stack.

As the scanning program completed scanning each batch, the batch was removed from the output hopper and placed back in its capsule. The next batch was taken from its capsule and placed into the input hopper and the machine resumed processing. When the machine had completed processing the last batch, the operator terminated the program, dismounted the tape, and removed the listings from the printer.

The output data tapes were forwarded to the VAX computer area for loading processing. The scanned documents were returned in their original cartons to the resolution processing area.

6.1.4.3 Loading

The scanning tapes were recei . and checked in by an operator at the VAX computer area. The operations coordinator, having received notification of the tape's transmittal, initiated the data entry management procedure on the VAX computer from which the "Load Scanning Tapes" option was selected (Figure 6.1-5). A second menu provided the supervisor with the choice of loading the scannable booklets, answer sheets, or excluded student questionnaires (Figure 6.1-6).

The program's first input request was the tape number, a six-digit code printed on an external label on the tape and coded internally by the scanning program. The VAX operator then mounted that tape on the tape drive and put the drive online, which logically connected the tape to the program. The program checked that the right tape number and type had been mounted and proceeded with the loading process. As it processed the tape, the program

Figure 6.1-5

Main Menu for Management of the NAEP Data Entry System

```
                N A E P   D A T A   E N T R Y   M A N A G E M E N T

                              OPTION: _

                     1  Tracking & Data File Management

                     2  Status Reports

                     3  Questionnaire Audit

                     4  Load Scanning Tapes

                     5  Quality Control

                     6  Spool Data for Final Edits

                     X  Quit

   Enter Option Code:
```

Figure 6.1-6

Menu for Loading of Data Tapes into NAEP Data Entry System

```
      N A E P   D A T A   T A P E   L O A D I N G   F U N C T I O N S

                              OPTION: _

                     1  Load Tape of Scannable Booklets

                     2  Load Tape of Scannable Answer Sheets

                     3  Load Tape of Excluded Student Questionnaires

                     X  Quit

   Enter Option Code:
```

printed the batch code and record count for each batch to the operator's terminal to assure the operator that the program was running properly When the program reached the end of the tape file, it printed out three listings, rewound and dismounted the tape, and returned to the main menu. The three listings consisted of an error log, a batch listing, and an audit listing.

The error log was a running commentary and summary of the processing of the tape. Each log was identified with the tape number, file name, and date of the loading run. The start of each batch was recorded with the batch number and its corresponding school and session codes. Any disagreement between these codes and those entered from the school worksheet was recorded at this point. Any booklets that did not belong to the session type (e.g., bridge booklets in a spiral session) were also listed here as well as all unscannable booklets. At the end of processing each batch, the program printed the number of scannable and unscannable booklets it had counted in that batch.

The batch listing reported the information from the front cover fields of each booklet within each batch. This listing could be checked against the administration schedules for discrepant or missing information.

The audit listing identified the data problems found within each batch. Each data anomaly was identified by the batch sequence number, booklet number, section, and item number to facilitate location of the data in the actual instruments by resolution staff.

The printed output was forwarded to the resolution area to be joined with the scanned materials. The tape was retained in the VAX computer area.

6.1.4.4 Resolution Processing

The error log and batch listing were retained by the operations coordinator. The audit listings were separated by batch number and matched with the appropriate scanned materials. If the error log indicated any unscannable booklets within a batch, they were identified and extracted from the bundle and manually entered and verified through the data entry system. Upon completion of verification processing, the system produced an updated audit listing to replace the one output from the load process.

Resolution processing was not permitted to proceed until *all* materials from a session had been scanned, loaded and received, after which the separate booklet and answer sheet batches from a session were matched, combined, and forwarded to resolutions staff.

Staff assigned to resolution processing reviewed the audit listing, checked the actual responses in the documents wherever asterisks or question marks were indicated, determined the appropriate value(s) to be coded in the data file, and wrote these new codes on the audit listing. The asterisks indicated multiple gridding of a single-response item, question marks flagged critical fields from the front cover, such as gender or birth date, that were incorrectly gridded, and fields from unscannable pages.

Access to the student data for entry, verification, or resolution processing was gained through the second option on the main data entry menu The first screen (Figure 6.1-7) requested the identification number of the batch to be processed; the PSU, school, and session codes as a secondary check on the batch code; and a code for the processing mode: entry, verification, or resolution. The second entry screen (Figure 6.1-8) prompted for input of the batch serial number and for the student ID number as a secondary check.

The processing mode was a program state that determined how much data the terminal operator would see, how to process data entered by the operator, and the management of data within the program. The entry mode permitted the creation of a new data record by the operator. The verification mode presumed the existence of a data record, permitted the "creation" of a second data record, and performed a field-by-field comparison of the two records, alerting the operator of data disagreements. The resolution mode displayed the entire contents of a data record and interpreted any data entered by the operator as a correction of a data field.

The resolution mode of the entry system permitted the operator to read verified or loaded data records, display the field values, and make corrections to individual fields. A change in any data field under resolution mode also generated a record for the audit file, and the program produced an updated audit listing at the completion of resolution processing for each batch. There was no limit to the number of times a session or data record could be processed under resolution. On completion of resolution processing, each session bundle was stored in a labeled box and held for final editing and quality control processing.

If the program was in the entry mode and no data record for the booklet could be found, the program would prepare to create a new record and request entry of the booklet cover data. If in verification mode and the data record had not been already verified, the program would request re-entry of the cover data and compare them against the data record. If in resolution mode and the data record had been through verification or loading processing, all data fields were displayed and the operator could either modify these fields or advance to the rest of the entry screens for that booklet.

A final validation was performed when the data entry work files were merged and copied, or spooled, onto a master student data file. This spooling program checked every data field of every student record for out-of-range values and question marks. A listing similar to the audit listings for each session was produced, which the resolution staff then used to identify and correct the remaining data anomalies.

The quality control process first selected a random sample of each booklet type from the master student file, identifying them by batch and sequence number. The designated booklets were located, extracted from their storage boxes, and forwarded to the quality control staff. The responses in each booklet were then compared with their coded data values in the data file On completion of quality control processing, the booklets were returned to

Figure 6.1-7

Student Session Data Entry Screen

```
                    NAEP YEAR 19 STUDENT DATA

                         BATCH: ____

                           PSU: ___
                        SCHOOL: ___
                       SESSION:  __

                          MODE:   _
```

Figure 6.1-8

Student Booklet Cover Data Entry Screen

```
                     STUDENT ASSESSMENT BOOK

      BATCH SERIAL #: ____

                                              BOOK

                                                            PART
  ID:______             P/S:__/___                         1
                                              87-88        2
  R:_          A:_      ETS:__/___                         3
                                             AGE           4
  G:__        SEX:_       B:__/_                           5

                              SCORERS
                      S1:__    S2:__    S3:__
```

their boxes. The full details and results of the quality control process are presented in Chapter 6.5.

At the completion of resolution processing of the student data, the session bundles were separated by booklet or answer sheet and sorted into individual stacks by instrument number. These stacks were then placed into storage boxes, identified by age cohort and instrument number, and shipped to the ETS data retention area for long-term storage.

6.1.5 QUESTIONNAIRES

The questionnaire instruments were separated by type and accumulated by the operations coordinator as they were received from mail processing. The teacher and school questionnaires were eventually transcribed through the NAEP data entry system, but on a lower priority basis than the student booklets. The excluded student questionnaires were batched and sent to scanning at regular intervals, since the demographics of the excluded students were used in deriving the sampling weights of the assessed students. In order to allow the two files to be completed at the same time, every effort was made to keep the processing rate of these instruments in pace with the student data entry.

Processing of the questionnaire data was initiated by selecting the third option on the data entry menu. The first entry screen (Figure 6.1-9) prompted for input of the questionnaire type, age group, and processing mode. The questionnaire entry programs followed the same model as the student entry program with the absence of a tracking file and session batching. Entry, verification, and resolution modes were available; audit reports were initiated by the operations coordinator.

Figure 6.1-9

Primary Menu for the Entry of Questionnaire Data

NAEP YEAR 19 QUESTIONNAIRE MENU

TYPE: _	AGE: _	MODE: _
1 SCHOOL	1 AGE 9	1 ENTRY
2 TEACHER	2 AGE 13	2 VERIFICATION
3 EXCLUDED STUDENT	3 AGE 17	3 RESOLUTION

The data for the excluded student questionnaires for the age 13 bridge sample were entered through the data entry system. The program for entry of the excluded student questionnaire data first displayed a screen for entry of the front cover data. The operator was prompted for the serial number of the booklet to be processed. An error condition occurred when either a record with that serial number was found under entry mode or when no record was found under the verification or resolution mode. In either case, the operator was asked to verify that the correct number had been entered. If the problem persisted, it was referred to the operations coordinator for resolution. The remaining cover information, including PSU and school code, student gender, race/ethnicity, grade, and birth date, was processed according to the same criteria as were the data from the student booklet covers. The program then displayed a single screen for processing the responses within the questionnaire. When the operator pressed ENTER to terminate processing for that booklet, the program redisplayed the cover entry screen, ready to process another booklet. A blank field entered in the serial number field returned the program to the primary menu.

The excluded student questionnaires for the age 9 bridge, age 17 bridge, and all main samples contained detachable machine-scannable answer sheets for the recording of identification information and responses to the questions. These sheets were checked for correctness and completeness of identification information, separated from their questionnaires, and placed on one of three piles, according to age cohort. On a weekly basis, the operations coordinator would grid a batch header sheet for each pile with the appropriate age cohort information, place it on the appropriate pile, and send these batches to the scanning area.

The scanning program for the excluded student answer sheets was executed once for each batch, creating a separate output tape for each age cohort. At the completion of scanning processing, the tapes were sent to the VAX computer area and the batches sent back to the data entry area. Upon notification of the tape serial numbers, the operations coordinator started the data entry management procedure on the VAX computer, selected the "Load Scanning Tapes" option, and chose the third option to initiate loading of the excluded student data.

The loading program for the excluded student data performed many of the same functions as the program for loading the assessed student data: checking the demographic information for appropriateness to age cohort and sample, validating the questionnaire responses for range, and reformatting the output records for compatibility with the data entry system. The entry system maintained a single excluded student data file and a single audit file for each age cohort. The load program wrote the edited data records to the appropriate data file, using the unique book serial number as an indexing key for insertion into the file, and for later retrieval by the resolution program. The program also wrote the audit records to the appropriate audit file for each age cohort, "appending" them to the end of the file. When the program completed loading the data, it produced an audit listing of data anomalies found in that batch.

Resolution processing started by comparing the anomalous data in the audit listing with the marked responses on the answer sheets. All corrections were recorded on the listing, which was given to a data entry operator. The operator selected the third option from the main menu to display the questionnaire data entry menu, then indicated resolution of excluded student questionnaire data for the appropriate age cohort. After entering the identification number for the record to be corrected, the operator confirmed that the right record had been retrieved and made the corrections to the erroneous data. An account of each data correction was written to the audit file.

The operations coordinator had the ability to produce, either as needed or at the conclusion of all resolution processing for an age cohort, a complete audit listing for any questionnaire. This listing was organized by transaction within data field within identification number, to facilitate visual analysis of resolution activity. At the conclusion of the excluded student data processing for an age cohort, a special summary program was executed that listed and counted the processed questionnaires within each school. This listing was compared against the individual questionnaire rosters for each school to determine if any shipments were still incomplete. The rosters contained enough information to generate "dummy" records for each missing questionnaire, which were designated with a special code for use by Westat in deriving the sample weights. These records were added to the file that was sent to Westat along with the absentee and assessed student data files.

The program for entry of the teacher questionnaire data first displayed a screen for entry of the cover information. It processed the serial number in the same fashion as did the entry program for the excluded student questionnaire. The cover information included only the PSU, school, and teacher codes. As the longest questionnaire instrument, the teacher questionnaire required three screens for entry processing. Completion of processing for each booklet returned the program to the cover entry screen, where the entry of a blank serial number returned the program to the primary menu.

The program for entry of the school questionnaire data also started with a display of the cover entry screen. The only information requested for this instrument, however, was the PSU and school code, which also served as the booklet identification number. Because of the large number of questions in this questionnaire, entry processing required two screens. Completion of processing for each booklet returned the program to the cover entry screen, where the entry of a blank PSU and school code returned the program to the primary menu.

After all questionnaires had been received and processed through the entry system, a validation program was run against all data values in all records. All remaining data errors or discrepancies were then corrected using the resolution mode of the entry system. A final audit listing was generated, recording all entry activities for each questionnaire.

The questionnaires were subjected to the same quality control procedures received by the student data. The details of the sampling rates and results are discussed in sections 6.5.2 through 6.5.4.

At the completion of quality control processing, the questionnaires were packed into boxes and shipped to the ETS data retention area for long-term storage.

Chapter 6.2

PROFESSIONAL SCORING[1]

Lynn B. Jenkins and Anne Campbell

Educational Testing Service

Like previous NAEP assessments, the 1988 assessment included a variety of open-ended items—or items that ask students to provide written responses. Open-ended items were administered as part of the main assessments in reading, writing, document literacy, civics, and U.S. history, and the bridge assessments in reading, writing, civics, mathematics, and science. Some of the items requested extended writing, and these appeared alone in a block so that students had 10 to 15 minutes to respond. Others requested shorter written responses; these were interspersed with other items in a block.

The 1988 main and bridge (or trend) assessments included the following numbers of open-ended items.

1988 Main Assessment

	Grade 4/Age 9	Grade 8/Age 13	Grade 12/Age 17
Reading	2	1	2
Writing	7	8	8
Document Literacy	N/A	27	27
Civics	0	1	1
U.S. History	0	1	1

1988 Bridge (Trend) Assessments

	Age Class 9	Age Class 13	Age Class 17
Reading	5	8	12
Writing	6	6	6
Civics	0	1	2
Science	0	0	2
Mathematics	28	27	54

[1]The authors would like to acknowledge Debra Kline, Walter MacDonald, and Ina Mullis for their contributions to the text of this chapter and Bruce Kaplan, David Freund, Rebecca Zwick, and Jim Ferris for providing statistical data.

In both the bridge and main assessments, some of the same items were administered at more than one age class. Tables 6.2-1 and 6.2-2 give an overview of the main and bridge assessment items, including their NAEP identification numbers, response time provided, age classes, and score ranges.

In the 1988 assessment, three types of answer documents were used: key-entered booklets, scannable booklets, and scannable answer sheets. The placement of the scores and the manner in which the scores were recorded varied according to the type of answer document that was used. Scores for open-ended items in the bridge to 1984 were recorded on the back covers of the booklets and subsequently key-entered. Machine-scannable booklets were used for the bridge to 1986, for the main assessment of grade 4/age 9 students in all subject areas, and for the main assessment of grade 8/age 13 and grade 12/age 17 students in writing and document literacy. Scores for open-ended items in these samples were gridded in ovals at the bottom of the pages on which the items appeared. Scannable answer sheets were used for the main reading, civics, and U.S. history assessments of students in grade 8/age 13 and grade 12/age 17. Scores for open-ended items in these samples were gridded on the page of the answer sheet where the response was written.

Three teams of readers worked simultaneously to score the open-ended items. One team scored responses to the mathematics bridge items, while a second team scored responses to the bridge items in the other subject areas. A third team scored all open-ended item responses from the main assessment.

The rest of this chapter includes a description of the scoring operation, including scoring guides, training, work flow, and the measures used to monitor the reliability of the scoring procedures.

6.2.1 DESCRIPTION OF SCORING

Each open-ended item included in the 1988 assessment had a unique scoring guide that identified the range of possible scores for the item and defined the criteria to be used in evaluating students' responses. To enable NAEP to accurately measure changes in performance across time, the scoring guides for open-ended trend items readministered in 1988 were identical to those used in the previous assessments. The following sections summarize the scoring guidelines used to evaluate responses to the open-ended items for each subject area assessed in 1988.

Reading and Document Literacy

The scoring guides for the open-ended reading items included in the bridge and the main assessments focused on students' ability to perform various reading tasks—for example, identifying the author's message or mood and substantiating their interpretation, making predictions based on given details, supporting an interpretation, and comparing and contrasting information.

Table 6.2-1
NAEP 1988 Main Sample Open-ended Items*

NAEP ID	Description	Response Time (Mins.)	Age Class 9	13	17	Score Range	Secondary Trait
Writing							
N000311	Recreation Opp.	15		X	X	0-4, 7-9	Yes
N000331	Recreation Opp.	30		X	X	0-4, 7-9	Yes
N000411	Food on Frontier	15		X	X	0-4, 7-9	
N000551	Dissecting Frogs	7½		X		0-4, 7-9	
N000941	Radio Station	10	X			0-4, 7-9	
N000951	Radio Station	7½		X		0-4, 7-9	
N007711	Ghost Story	15		X	X	0-4, 7-9	Yes
N007721	Ghost Story	20	X			0-4, 7-9	Yes
N007731	Ghost Story	30		X	X	0-4, 7-9	Yes
N007741	Ghost Story	10	X			0-4, 7-9	Yes
N014741	Plants	10	X			0-3, 7-9	
N014821	Spaceship	20	X			0-4, 7-9	Yes
N014841	Spaceship	10	X			0-4, 7-9	Yes
N018051	Space Program	7½			X	0-4, 7-9	
N021051	Bike Lane	7½			X	0-4, 7-9	
W000141	Summary of Story	10	X			0-4, 9	
W000221	Favorite Animal	20	X			0-4, 9	Yes
W000241	Favorite Animal	10	X			0-4, 9	Yes
W000341	Three Wishes	10	X			0-4, 9	
W000411	Favorite Story	15		X	X	0-4, 9	
W000511	TV Habits	15		X	X	0-4, 8, 9	Yes
W000531	TV Habits	30		X	X	0-4, 8, 9	Yes
W000611	Memorable Event	15		X	X	0-4, 9	
Reading							
R000206	Dove and Ant	(I)**	X			0-5, 9	
R000807	Grandpa and Wind	(I)	X	X		0-5, 9	
R002406	Small Fruits	(I)			X	0-2, 9	
N015905	High Tech Pizza	(I)			X	0-4, 7-9	
U.S. History							
H024901	Settlers	15		X		0-5, 9	
H025002	Presid. Power Part 1	15			X	0-2	
H025003	Presid. Power Part 2	15			X	0-4, 9	
Civics							
P018201	Presid. Resp. Part 1	15		X	X	0-1	
P018202	Presid. Resp. Part 2	15		X	X	0-4, 9	

* Not including open-ended items scored as right/wrong.

** (I) denotes that the item appeared in a 10-minute block at grade 4 or a 15-minute block at grades 8 and 12 that contained several multiple-choice content items.

Table 6.2-2
NAEP 1988 Bridge Sample Open-ended Items*

NAEP ID	Description	Response Time (Mins.)	Age Class 9	13	17	Score Range	Secondary Trait	Holistic
Writing								
N000302	Recreation Opp.	15		X	X	0-4, 7-9	Yes	Yes
N000402	Food on Frontier	15		X	X	0-4, 7-9		Yes
N000502	Dissecting Frogs	7½		X		0-4, 7-9		
N000602	XYZ Company	7½	X	X		0-3, 7-9		
N000902	Radio Station	15	X	X		0-4, 7-9		
N001002	Appleby House	15	X	X	X	0-4, 7-9		
N007602	Flashlight	15	X			0-4, 7-9		Yes
N014702	Plants	15	X			0-3, 7-9		
N014802	Spaceship	7½	X			0-4, 7-9		Yes
N018002	Space Program	7½			X	0-4, 7-9		
N019002	Job Application	7½			X	0-4, 7-9		Yes
N021002	Bike Lane	15			X	0-4, 7-9		
Reading								
N001507	Nuts	(I)**	X	X	X	0-6, 7-9	Yes	
N001904	Charley	(I)		X	X	0-5, 7-9	Yes	
N002302	The Door	(I)		X	X	0-9		
N002804	Bethune	(I)	X	X	X	0-5, 7-9		
N003104	Goods to Market	(I)	X	X	X	0-4, 7-9		
N003704	Web Life	(I)	X	X	X	0-4, 7-9		
N004303	Javelin	(I)		X	X	0-4, 7-9		
N004605	Jobs	(I)		X	X	0-5, 7-9		
N008905	Mother and Dog	(I)	X			0-6	Yes	
N015905	High Tech Pizza	(I)			X	0-4, 7-9		
N021301	Jacob	(I)			X	0-4, 9	Yes	
N021801	Eggplant I	(I)			X	0-5, 9	Yes	
N021805	Eggplant II	(I)			X	0-4, 9	Yes	
Civics								
P021001	Democracy	(I)			X	10-15, 20-24, 77, 88		
P021101	Newspaper Publishers	(I)		X	X	10-16, 20-21, 7, 8		
Science								
N430801	Pendulum	(I)			X	0-3, 9		
N437001	Battery/Bulb	(I)			X	0-4, 9		

* Not including open-ended items scored as right/wrong.

** (I) denotes that the item appeared in a 10-minute block at grade 4 or a 15-minute block at grades 8 and 12 that contained several multiple-choice content items.

The guides for the reading items varied somewhat, but typically included the distribution of score points shown below.

Outline for Scoring of Open-Ended Reading Items

Score	Definition
5	Elaborated reference or interpretation. These responses exceeded the requirements of the task by including illustrative examples or details and demonstrating a high level of cohesiveness.
4	Satisfactory reference or interpretation. These responses identified at least two relevant examples or reasons to support a given interpretation.
3	Minimal reference or interpretation. These responses identified at least one relevant example or reason to support a given interpretation.
2	Unsatisfactory reference or interpretation. These responses did not give evidence to support a stated interpretation.
1	No reference or interpretation. These responses did not provide an interpretation, but instead digressed or avoided the task.
0, 7, 8, 9	These responses were, respectively, blank, indecipherable, completely off-task, or included a statement to the effect that the student did not know how to do the task. (In the scoring guides for the main assessment, scores of 7, 8, and 9 were collapsed into the score point of 9).

Some of the guides for the main assessment items included secondary scores, which typically involved categorizing the kind of evidence or details the student used as support for an interpretation. The document literacy items, most of which required short answers, were scored on a right-wrong basis.

Writing

There are widely divergent views as to what constitutes good writing. In response to these different conceptions, writing researchers have developed a variety of methods for evaluating students' writing abilities.

To provide multiple perspectives on students' writing performance, NAEP uses three scoring approaches—primary trait, holistic (or general impression), and mechanics scoring—to evaluate responses to the writing assessment tasks. Selected writing items in the bridge assessment were scored using all three approaches, while the remaining items were scored using the primary trait method only. The primary trait method was used to score items in the main assessment.

As described in the sections that follow, the purposes of the three scoring approaches used by NAEP are quite different. Primary trait scoring focuses on students' ability to accomplish the core purpose of a particular writing task, holistic scoring focuses on overall fluency, and mechanics scoring focuses on students' grammar, punctuation, and spelling. Previous research has revealed moderate correlations between the results from holistic and primary trait scoring (ranging from .29 to .60); however, the two approaches evidently capture different aspects of writing performance (Applebee, Langer, & Mullis, 1989). The range of the correlation coefficients between the two sc' ' g approaches was from .39 to .66 for the trend results for the 1984 and 1988 writing assessments (Applebee, Langer, Mullis, & Jenkins, 1990).

Primary Trait (or Task Accomplishment) Scoring. As noted above, the primary trait scoring method focuses on the writer's effectiveness in accomplishing specific tasks. It is sensitive to the writer's understanding of the audience as well as to the inclusion of specific features needed to accomplish the specific purpose of a task. The primary trait scoring criteria defined five levels of task accomplishment: not rated, unsatisfactory, minimal, adequate, and elaborated. The scoring guide for each item described these levels in detail. A general explanation of the score points is given below.

Levels of Writing Task Accomplishment

Score	Definition
4	Elaborated. Students providing elaborated responses went beyond the essential, reflecting a higher level of coherence and providing more detail to support the points made.
3	Adequate. Students providing adequate responses included the information and ideas necessary to accomplish the underlying task and were considered likely to be effective in achieving the desired purpose. (For two of the items, this was the highest possible score.)
2	Minimal. Students writing at the minimal level recognized some or all of the elements needed to complete the task but did not manage these elements well enough to assure that the purpose of the task would be achieved.
1	Unsatisfactory. Students who wrote papers judged as unsatisfactory provided very abbreviated, circular, or disjointed responses that did not even begin to address the writing task.
0, 7, 8, 9	Not Rated. A small percentage of the responses were blank, indecipherable, or completely off task, or contained a statement to the effect that the student did not know how to do the task; these responses were not rated.

Some items also were scored for secondary traits, which involved indicating the presence or absence of elements that were of special significance to a particular item (e.g., whether notes were made before writing or whether critical information was filled out on a form).

Mechanics Scoring. Mechanics scoring focuses on the extent to which the writer can control the conventions of written English—specifically, grammar, punctuation, and spelling. In addition, the procedures include identifying sentence structures and computing sentence length to gauge the sophistication of students' syntax. In the mechanics scoring, conducted after the main scoring effort had been completed, two teams of readers joined together to score a subset of responses to selected open-ended items from the writing bridge assessment. One item was chosen at grade 4 ("Spaceship") and a second item was chosen at grade 8 and grade 11 ("Recreation Opportunities"). A random probability sample of approximately 500 essays was selected from each grade level for the 1984 and the 1988 assessments, for a total of 1,000 essays at each grade level. The set of essays selected from each grade level for each year included responses from approximately 200 students who were Black and approximately 300 students who were not. Black students were oversampled to ensure that the comparisons of performance between Black and White students were reasonably precise.

Prior to the scoring, the responses were duplicated with the student's identification number shown on the copy. The essays were then bundled by grade by assessment year. As the readers selected bundles to score, they alternated among the different grade levels and years.

Rather than assigning a single score to each paper, as was done in the primary trait scoring, the mechanics scorers marked each paper with a series of symbols, addressing the elements of sentence construction, word choice, spelling, punctuation, and capitalization. These symbols, written in red ink, designated each word or punctuation mark in error and indicated sentence type or faulty sentence construction.

To analyze the mechanics data, criteria were devised to derive information from scoring codes (see Campbell, 1987 for a description of these criteria). The analyses included calculations of:

1) the average number of words in an essay;
2) the average number of sentences in an essay;
3) the average number of letters in a word;
4) the average number of errors in an essay;
5) the percentage of different types of sentence construction: and
6) the rate of punctuation errors and omissions.

Holistic Scoring. In holistic scoring, readers evaluate the fluency of each student's writing compared to the writing of other students at the same grade or age level who responded to the same task. Unlike primary trait or mechanics scoring, where the reader focuses on the presence or absence of particular elements, holistic scoring takes a global view of the ideas, language facility, organization, mechanics, and syntax of each paper taken as a whole—as its name implies. "The chief assumption that underlies holistic scoring of essays is that the whole text or composition is more than the sum of its parts...To look at a composition as a whole in order to judge its quality as an entity in itself is to score it holistically" (Breland, Camp, Jones, Morris, & Rock, 1987, p. 18).

The holistic scoring was conducted by a large group of readers in a session that was conducted separately from the primary trait and mechanics scoring sessions. The tasks scored holistically were "Spaceship" and "Flashlight" at grade 4, and "Recreation Opportunities" and "Food on the Frontier" at the upper grade levels (8 and 11). Trained readers evaluated the relative fluency of students' writing on a 6-point scale. A small percentage of papers—such as those that were blank or indecipherable—were not rated.

The holistic scale was anchored by chief readers and table leaders chosen for their expertise in holistic scoring. This group studied the pool of student responses to select papers that represented each point on the holistic scale, then used these sample papers to train a group of approximately 50 raters. Using the sample papers as a guide, the raters wer trained to determine whether papers corresponded to the top half or the bottom half of the holistic scale, then to make finer distinctions between adjacent points on the scale.

To conduct the scoring, the readers were divided into two group One large group was responsible for evaluating eighth and eleventh graders' responses to the two tasks common to those grade levels, while a smaller second group was responsible for evaluating fourth graders' responses to he two tasks given only at that grade level. Because the emphasis of the holistic scoring was on detecting trends across time at each of the three grade levels assessed, the tasks given at grades 8 and 11 were rated separately, although by the same readers. A training session preceded the scoring of responses to each task at each grade level.

Student papers are evaluated relative to one another in holistic scoring, rather than against specific criteria, as with primary trait scoring. Therefore, for each task at each grade level, the distribution of scores for the total sample of papers should be approximately normal, with scores evenly distributed around the center of the scale. To detect changes in writing fluency across time at each grade level, papers from the 1984 and 1988 assessments were randomly mixed prior to scoring. Thus, if more papers from either assessment were judged to be in the top half of the scale, the results would indicate changes across time in overall writing fluency.

Mathematics

Because the open-ended mathematics items in the 1988 assessment tended to focus on computational skills, all were scored on a right-wrong basis, where 1—correct and 2—incorrect. Omitted responses were scored as 0. Answers written on the answer lines were the primary basis for the scores; however, if the student left the answer line blank, consideration was given to answers written under the item or answers written where the student had worked out the item.

Science

The scoring guides for the two open-ended science items ("Pendulum" and "Batteries and Bulbs") focused on how correctly the student answered the questions. The following outline summarizes the guidelines used to score these items.

Outline for Scoring of Open-Ended Science Items

Score	Definition
4	This score indicated a correct, detailed answer. (Only one of the scoring guides included this score point.)
3	This score indicated a correct answer.
2	This score indicated an answer that was correct to a point but either contained some misinformation or was too general.
1	This score indicated an incorrect response to the question.
0, 7, 8, 9	These scores were given to responses that were, respectively, blank, indecipherable, or off-task (not relevant), or contained a statement to the effect that the student did not know how to do the task.

Civics and U.S. History

The first part of the scoring standard for the open-ended civics item ("Presidential Responsibilities") included in the main assessment asked readers to distinguish between correct and incorrect responses to the initial part of the task, in which students were asked to name the current president. The second part of the task asked students to describe the president's responsibilities, and the accompanying guide defined the criteria for each score point, as shown below.

In contrast, the scoring rubrics for the two civics trend items ("Newspaper" and "Democracy") defined specific criteria for acceptable and unacceptable responses. Many types of acceptable responses were possible, and each type was given a separate score.

Outline for Scoring of Open-Ended Civics Item

Score	Definition
4	Elaborated. These responses provide a mix of specific examples and thoughtful discussion.
3	Adequate. These responses provide one or two examples of responsibilities with little discussion.
2	Minimal. These responses consist primarily of generalities or contain a list of information that contains errors.
1	Unacceptable. These responses digress from the topic, give incorrect information, or do not attempt to respond to the question.
9	Not Rated. No response or totally off-task.

As shown below, the scoring guides for the two open-ended U.S. history items included in the main assessment ("Settlers" and "Presidential Powers") focused on the accuracy and elaboration of students' responses to the questions. As with the civics item ("Presidential Responsibilities") previously described, the first part of the "Presidential Powers" task was scored dichotomously. Students were asked to state who was more powerful—the presidents of today or of Washington's era, and raters marked whether or not the student took a position. In the second part of the task, students were asked to support the position they stated.

Outline for Scoring of Open-Ended U.S. History Items

Score	Description
5	These responses contain several reasons supported by appropriate, specific examples. (Only the scoring guide for the "Settlers" item specified this score point.)
4	These responses contain at least two reasons with explanations and may also give a lengthy list with an explanation of at least one item.
3	These responses give a list of reasons without any explanation or one reason with an explanation. They contain no significant errors.
2	These responses provide only one correct reason, repeat a single point, or include incorrect or insignificant reasons.
1	These responses do not answer the question correctly or reiterate the question.
9	No response or totally off-task.

6.2.2 THE SCORING OPERATION

Overview of the Scoring Operation

For the main assessment, a group of eight persons scored the open-ended items for all subject areas. For the bridge assessment, three persons scored all the open-ended items in all subject areas. A majority of the readers had at least bachelor's degrees in education, English, or history. The two persons assigned to the scoring of the mathematics items at all three levels had at least a high-school education. The readers included men and women of various ages and racial, ethnic, and geographical backgrounds.

The NAEP scoring supervisor ctly monitored the scoring of the mathematics items and managed the operation of the other two groups. The scoring supervisor also reviewed discrepancies between readers in the scoring of responses to the bridge items. To facilitate the scoring process, the supervisor delegated the responsibility of reviewing scoring discrepancies between readers for the main assessment items to two of the best scorers in the group. However, the scoring supervisor was always available to consult with these individuals when they encountered responses that were particularly difficult to score.

Training: Mathematics

Because the mathematics items were scored as right, wrong, or omitted, lengthy training for scoring these items was unnecessary. In an orientation period, the readers were trained to follow the procedures for scoring the mathematics items and became familiar with the scoring guides, which listed the correct answers for the items in each of the blocks.

Training: Reading, Writing (Primary Trait), Civics, U.S. History, and Science Scoring

Before the training program started, the NAEP scoring supervisor worked with NAEP test development staff to prepare training sets (or sets of sample responses to accompany the scoring guides) and to refine the scoring guides for newly developed items.

For the main assessment, readers were trained on all the writing, reading, civics, and U.S history items at all three grade/age levels. Training involved explaining the item and its scoring guide and discussing responses that were representative of the various score points in the guide When this process was complete the readers scored and then discussed approximately 65 to 100 randomly selected "practice papers" for each item. The purpose of the training was to familiarize the group with the scoring guides and to reach a high level of agreement among the readers. When the group training had been completed, each reader scored all the open-ended items in each of nine bundles of booklets, after which a follow-up session was held to discuss responses that received a wide range of scores. Once the follow-up session was completed, the formal scoring process began. The initial training was completed in approximately four weeks.

The training program for the bridge assessment was carried out on all the items at one age class at a time, starting with age class 13, followed by age class 9, and ending with age class 17. (This order corresponds to the order in which the bridge assessments are conducted.) The training program followed the same procedures used for the main assessment scoring. In order to ensure continuity with the past scoring of the bridge items, at least half the sample papers in the training sets were taken from the 1984 training sets. The training program for each age level took approximately two weeks.

As a follow-up to the training program, notes on various items were compiled for the readers of each group for their reference and guidance throughout the scoring process. In addition, short training sessions were conducted when the scoring supervisor ascertained in reviewing discrepancies that certain items were causing difficulties for the scorers. The scoring supervisor also consulted with individual readers as the scoring progressed. When a reader's score was judged to be discrepant with that of another reader, the supervisor discussed the response and its score with that reader.

Training: Writing Mechanics Scoring

To prepare for the mechanics training, the NAEP scoring supervisor selected papers to be used in training. The training itself involved discussing the scoring guidelines and procedures and reviewing sample responses that had already been scored. The readers then practiced scoring other papers, and discussions were held when any discrepancies occurred. When the readers were comfortable with the guidelines, the actual scoring began. Several follow-up training sessions were held as problems arose.

Training: Holistic Scoring

The training for the holistic scoring of writing bridge items involved several stages. First, NAEP staff developed guidelines describing six levels of proficiency for each task. Then, NAEP staff and two chief readers—both of whom were experienced holistic readers—surveyed the pool of papers from the assessments and selected anchor papers, or papers representative of the six levels of proficiency. The guidelines were modified accordingly and criteria were established for distinguishing between top-half and bottom-half papers. A session was then held for the table leaders—who were also experienced holistic readers—to familiarize them with the guidelines and sample papers.

The training of the readers began with some discussion of the guidelines and the anchor papers and included several practice scorings of other papers to resolve discrepancies among readers. When all the readers were comfortable with the guidelines, they scored papers for an hour, after which they discussed additional anchor papers. Throughout the subsequent scoring, there were periodic discussions of papers t ensure that readers continued to adhere to the same standards.

Assignment of Work

The two groups of readers for the bridge assessment began scoring the age class 13 items in November 1987. These same two groups started scoring the age class 9 items in February 1988, and the age class 17 items in April. The readers for the main assessment started scoring in March. Each group of readers received the booklets in batches as they were received from the schools. Because of the spiral design, a reader would receive many, if not all, of the items at a particular age class as he or she scored a batch of booklets. In scoring the main assessment items, the scorers alternated through the three age classes so that they were continually exposed to responses from all age classes throughout the scoring.

6.2.3 RELIABILITY AND RESOLUTION

Trend Assessments

Twenty percent of the 1984 and 1986 responses to the open-ended reading items in the bridge assessments were retrieved, the scores were masked, and the responses were distributed to and rescored by the readers. This rescoring was performed concurrently with the scoring of the 1988 responses. Because of differences in the way that scoring guidelines were applied to open-ended reading and writing responses in 1988 and previous assessments, the 1988 results for the professionally scored items are not directly comparable to past results. Changes in percents correct for these items are not appropriate for inferences about changes in reading or writing achievement. (See Chapters 10 and 11 for further discussion.)

Main Assessments: Reading, Writing, U.S. History, Civics, and Science

Twenty percent of the items in the other subject areas were subjected to a reliability check, which entailed a scoring by a second reader. To prevent a second reader from being influenced by the first reader's scores, the first reader masked the scores in every fifth booklet in a batch. These booklets were passed along to a second reader. All discrepancies were then reviewed by the scoring supervisor or those designated by the scoring supervisor

Mathematics

Ten percent of the mathematics items were subjected to a correctness check in which a second scorer verified that the first scorer had correctly scored the items. If the second scorer found a mistake in scoring, he or she corrected it. To assess the reliability of each scorer, the second scorer kept count of the number of times he or she checked each of the other scorers and the number of times he or she had to correct a score. This procedure was followed because the mathematics items were scored as right, wrong, or omitted and because the scoring guides were exact as to the correct answers. Results of this correctness check showed that the first scorer was correct 99 percent of the time.

Two statistics were used to report reader reliability: the percent of *exact* agreement and the reliability coefficient. The percent of exact agreement is the percentage of times that the two readers agreed exactly in their ratings. The reliability coefficient is the intraclass correlation between readers.[2] The results for each age class are shown in Tables 6.2-3 and 6.2-4. The first column lists the number of responses analyzed; the second column lists the percentage of exact agreement between the first and second readers; and the third column is the reliability coefficient.

The reliability results generally show a high level of agreement between readers. The percentage of exact agreement among readers was at or above 70.7 percent for all but the trend holistic scoring, and several items showed agreement as high as 99 percent. The reliability coefficients were also high, ranging from .64 to .99.

The percentage of exact agreement between the first and second readers tended to be slightly lower in the holistic scoring than in the other types of scoring. However, the reliability coefficients (ranging from .65 to .83) did not differ substantially from those for the primary trait scoring and are generally as high or higher than those reported for other studies (Breland et al., 1987). Also, when agreement between adjacent score points was taken into consideration—that is, when readers did not differ by more than one score point on the 6-point scale—the percent of agreement for holistic scores ranged from 88 to 94 percent.

---

[2]The intraclass correlation coefficient (ICC) is calculated as

$$r(\text{ICC}) = \frac{\text{MSS} - \text{MSR}}{\text{MSS} + \text{MSR}(\text{K}-1)}$$

where MSS and MSR are the mean squares for subjects (ratees) and the mean square residual obtained from a one-way ANOVA, and K is the number of raters. (In the present application, K-1 = 1.) This provides a consistent (but biased) estimate of

$$\rho(\text{ICC}) = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_R^2}$$

and is therefore interpretable as the proportion of total variance due to differences among subjects. The error term contains the rater effects, which cannot be estimated separately because subjects were not rated by the same set of raters.

Table 6.2-3

Percentages of Exact Score Point Agreement and Intraclass Correlation Coefficients for Open-ended Items in the 1988 Main Samples

NAEP ID	Description	Grade 4/Age 9 N	Grade 4/Age 9 % Agree	Grade 4/Age 9 Rel.	Grade 8/Age 13 N	Grade 8/Age 13 % Agree	Grade 8/Age 13 Rel.	Grade 12/Age 17 N	Grade 12/Age 17 % Agree	Grade 12/Age 17 Rel.
Writing										
N000311	Rec. Opp. (15 Min.)				474	92.8	0.93	414	89.1	0.92
N000331	Rec. Opp. (30 Min.)				190	96.8	0.98	154	92.2	0.95
N000411	Food on Frontier (15 Min.)				484	87.6	0.86	415	88.0	0.88
N000551	Dissecting Frogs (7½ Min.)				636	91.8	0.91			
N000941	Radio Station (10 Min.)	709	94.9	0.96						
N000951	Radio Station (7½ Min.)				552	93.3	0.94			
N007711	Ghost Story (15 Min.)				450	92.2	0.91	370	92.2	0.91
N007721	Ghost Story (20 Min.)	181	89.0	0.85						
N007731	Ghost Story (30 Min.)				151	95.4	0.95	164	90.9	0.92
N007741	Ghost Story (10 Min.)	400	90.0	0.80						
N014741	Plants (10 Min.)	433	93.1	0.95						
N014821	Spaceship (20 Min.)	214	91.6	0.95						
N014841	Spaceship (10 Min.)	433	92.2	0.95						
N018051	Space Program (7½ Min.)							477	91.4	0.94
N021051	Bike Lane (7½ Min.)							576	85.9	0.89
W000141	Summary of Story (10 Min.)	649	88.9	0.87						
W000221	Favorite Animal (20 Min.)	213	91.6	0.95						
W000241	Favorite Animal (10 Min.)	420	89.8	0.91						
W000341	Three Wishes (10 Min.)	618	92.4	0.92						
W000411	Favorite Story (15 Min.)				669	93.0	0.89	584	90.1	0.91
W000511	TV Habits (15 Min.)				445	93.3	0.95	401	90.3	0.92
W000531	TV Habits (30 Min.)				225	90.7	0.93	203	95.1	0.97
W000611	Memorable Event (15 Min.)				608	87.3	0.89	565	88.1	0.91

Table 6.2-3 (continued)

Percentages of Exact Score Point Agreement and Intraclass Correlation Coefficients for Open-ended Items in the 1988 Main Samples

NAEP ID	Description	Grade 4/Age 9 N	% Agree	Rel.	Grade 8/Age 13 N	% Agree	Rel.	Grade 12/Age 17 N	% Agree	Rel.
Reading										
R000206	Dove and Ant	433	88.9	0.95						
R000807	Grandpa and Wind	437	93.4	0.92	470	97.5	0.95			
R002406	Small Fruits							516	97.7	0.96
N015905	High Tech Pizza							409	91.9	0.93
U.S. History										
H024901	Settlers to America				530	90.2	0.93			
H025002	Presidential Power Part 1							423	92.7	0.78
H025003	Presidential Power Part 2							403	87.6	0.92
Civics										
P018201	Presid. Resp. Part 1				623	99.5	0.97	574	99.7	0.97
P018202	Presid. Resp. Part 2				584	88.9	0.85	556	90.8	0.91

158

159

Table 6.2-4

Percentages of Exact Score Point Agreement and Intraclass Correlation Coefficients
for Open-ended Items in the 1988 Trend Samples

NAEP ID	Description	Grade 4/Age 9			Grade 8/Age 13			Grade 11/Age 17		
		N	% Agree	Rel.	N	% Agree	Rel.	N	% Agree	Rel.
Writing										
N000302	Recreation Opp.				335	85.4	0.82	293	90.8	0.93
N000371	Recreation Opp. (holistic)				290	56.2*	0.76	239	47.3*	0.66
N000402	Food on Frontier				299	79.9	0.69	260	93.1	0.86
N000471	Food on Frontier (holistic)				248	48.4*	0.71	253	49.4*	0.65
N000502	Dissecting Frogs				335	76.1	0.64			
N000602	XYZ Company	275	97.1	0.99	325	93.5	0.92			
N000902	Radio Station	309	93.5	0.95	316	87.0	0.89			
N001002	Appleby House	227	90.3	0.92	288	75.4	0.69	253	89.3	0.89
N007602	Flashlight	136	87.5	0.88						
N007608	Flashlight (holistic)	163	54.0*	0.83						
N014702	Plants	350	94.3	0.95						
N014802	Spaceship	306	91.8	0.95						
N014808	Spaceship (holistic)	236	52.1*	0.69						
N018002	Space Program							296	89.9	0.93
N019002	Job Application							286	92.3	0.92
N021002	Bike Lane							298	84.9	0.87

* Note: Primary trait scoring was on a 4-point scale; holistic scoring was on a 6-point scale. For the holistic scoring, percentages of agreement between adjacent score points—that is, when readers did not differ by more than one score point on the 6-point scale—were as follows:

		Age 9	Age 13	Age 17
N000371	Recreation Opp. (holistic)		93.8%	94.1%
N000471	Food on Frontier (holistic)		90.3%	88.1%
N007608	Flashlight (holistic)	93.9%		
N014808	Spaceship (holistic)	94.1%		

160

161

Table 6.2-4 (continued)

Percentages of Exact Score Point Agreement and Intraclass Correlation Coefficients
for Open-ended Items in the 1986 Trend Samples

NAEP ID	Description	Grade 4/Age 9 N	Grade 4/Age 9 % Agree	Grade 4/Age 9 Rel.	Grade 8/Age 13 N	Grade 8/Age 13 % Agree	Grade 8/Age 13 Rel.	Grade 11/Age 17 N	Grade 11/Age 17 % Agree	Grade 11/Age 17 Rel.
Reading										
N001507	Nuts	144	90.3	0.94	174	82.2	0.82	127	92.1	0.92
N001904	Charley				133	70.7	0.86	137	93.4	0.97
N002804	Bethune	163	92.6	0.94	169	76.3	0.69	146	93.8	0.92
N003104	Goods to Market	378	95.0	0.94	348	81.9	0.73	247	90.7	0.89
N003704	Web Life	115	94.7	0.91	126	74.6	0.68	120	90.8	0.93
N004303	Javelin				130	82.3	0.77	99	97.0	0.97
N004605	Jobs				149	85.2	0.91	114	93.0	0.93
N008905	Mother and Dog	135	89.6	0.71						
N015905	High Tech Pizza							98	96.9	0.96
N021301	Jacob							180	95.6	0.98
N021801	Eggplant I							162	96.3	0.98
N021802	Eggplant I (Sec. Trait)							131	95.4	0.95
N021805	Eggplant II							140	98.6	0.92
Civics										
P021001	Democracy							371	98.1	0.98
P021101	Newspaper Publishers				203	90.1	0.90	370	96.2	0.96
Science										
N430801	Pendulum							150	97.3	0.99
N437001	Battery/Bulb							155	99.4	0.99

162

163

Chapter 6.3

DATA TRANSCRIPTION SYSTEMS

Alfred M. Rogers

Educational Testing Service

The transcription of the student response data into machine-readable form was achieved through the use of three separate systems: scanning, loading, and resolution.

The student instruments were printed in a format that allowed the transcription of marked responses in the booklets to computer readable form on a magnetic tape by a programmable optical scanning machine. The first part of this chapter will describe the scanning equipment, the programs and data used by the machinery, and the ETS quality control standards and procedures.

A second procedure "loaded" the data records from the scanning output tape into an interactive computerized data entry and resolution system. This loading procedure validated each scanned data field, reformatted the data records to be compatible with the resolution system, and reported all problems for subsequent resolution. The second part of this chapter details the loading procedure.

A modified form of the data entry system developed for the 1986 assessment was used for resolution of the scanned data, entry of the documents rejected by the scanning machine, and entry of the questionnaire instruments. The third part of this chapter will provide an overview of this system, which is described by Rogers (1987).

Figure 6.3-1 is a schematic diagram that represents the flow of student-related assessment materials through the data transcription system Figure 6.3-2 similarly represents the flow of questionnaire materials through the system. The reader may refer to these diagrams for clarification of the relationships among the components of this system.

6.3.1 SCANNING

The student booklets were scanned on a National Computer System W201 scanning system. The scanner was controlled by a Hewlett Packard 1000 minicomputer. This system also included a disk drive for storage of the scanning programs, a tape drive for the output of scanned data records, and a printer for the periodic listing of individual record contents for quality control checking. The scanning programs used were specifically written for NAEP using the assembler language of the Hewlett Packard.

Figure 6.3-1

1983 NAEP Data Transcription System
(Part 1: Student-Related Materials)

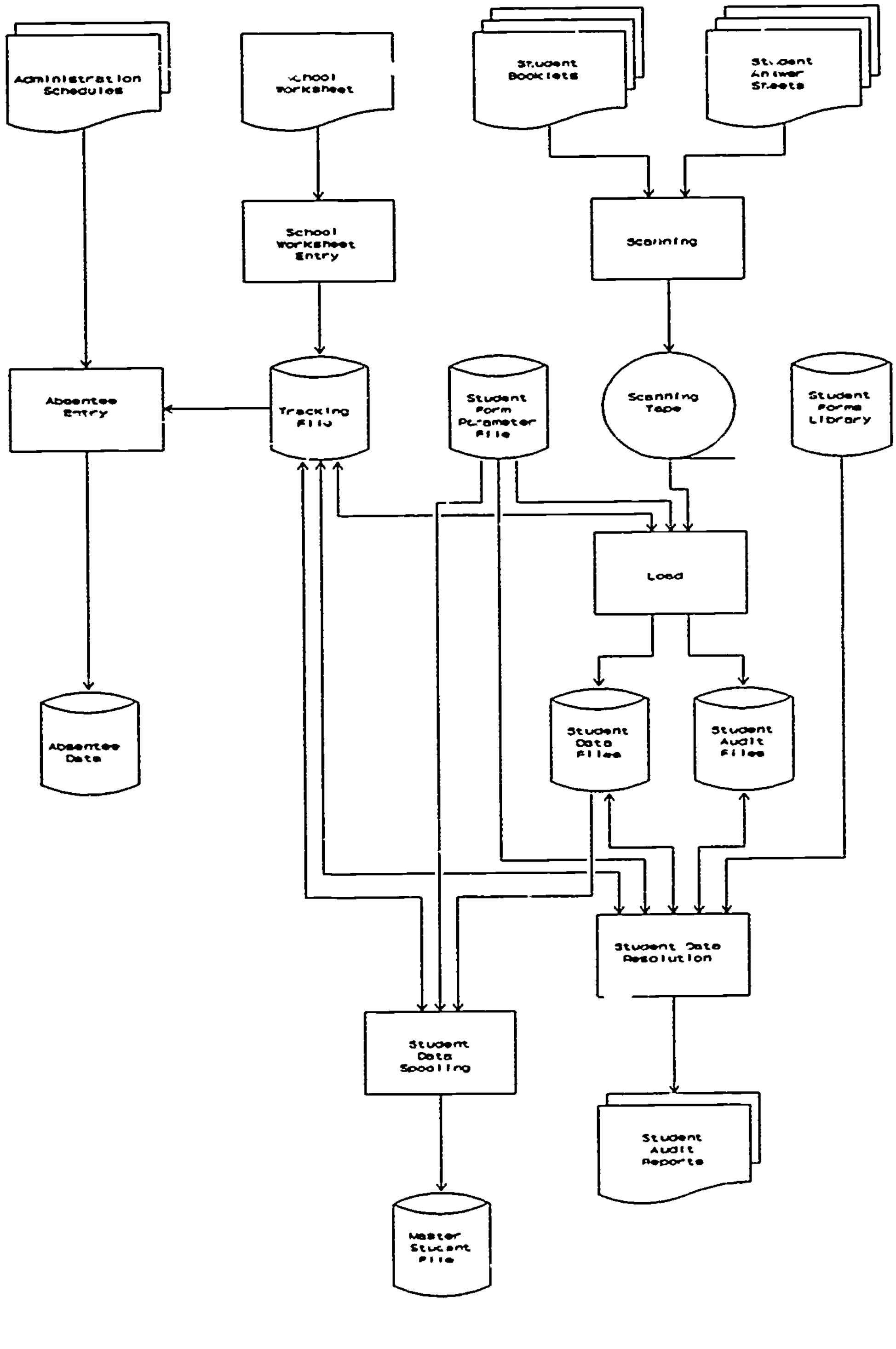
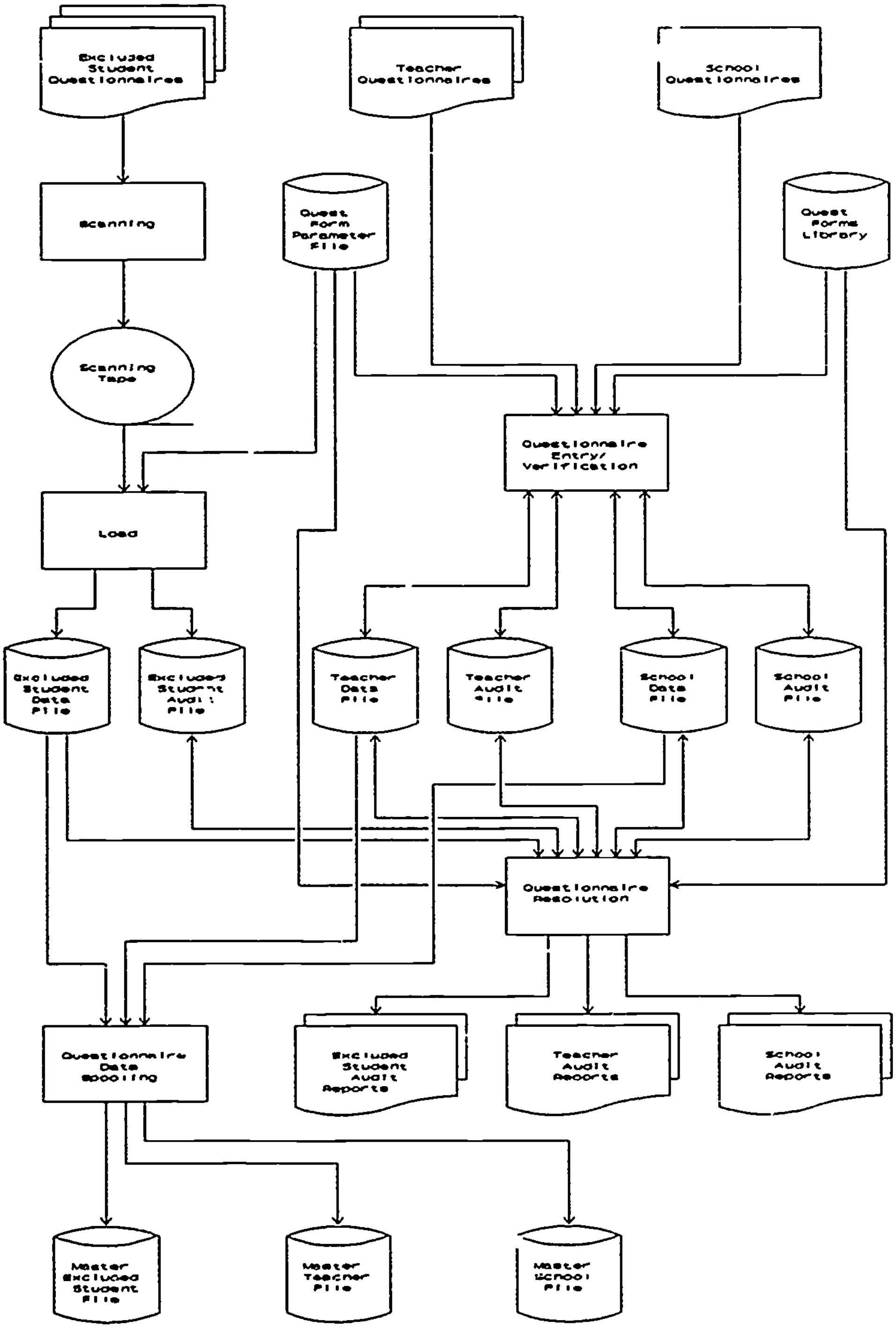


Figure 6.3-2

1988 NAEP Data Transcription System
(Part 2: Questionnaire Materials)



An optical scanner operates by sweeping a horizontally oscillating light beam across a vertically moving sheet and detecting reflections of the beam from pencil marks. The hardware logic of the scanner treats the page as a rectangular array of scannable areas, each of which is assigned a reflectance value from 0 to 15. This array of values is passed to the scanning program software, to be translated into response data.

After the first side of a sheet has been scanned, it is pushed through a loop that brings the other side of the sheet to face the scanning beam. A similar array of reflectance values is passed to the program that must then not only translate it into data, but decide whether to route this page to the output hopper and read in the next sheet or route it to the shunt hopper and stop processing.

The paper and inks used in producing scannable documents are required to have very low reflectances. A special set of marks are printed down one side of each page at equally spaced intervals to enable the scanning hardware to align each sheet and adjust the scanning rate to the movement of the sheet. These timing marks are printed using a highly reflective ink.

Each page of each item block has its own unique format in terms of the arrangement of the response and scoring ovals. The scanning program has to be able to identify a given page, determine which parts of the returned array to process, interpret the reflectance values, and transcribe them to data codes on the output record. Each page is printed with a set of marks next to the timing marks that are used by the program to identify it uniquely by block code and page number. The booklet covers are similarly identified according to booklet number.

The scanning program logic uses two sets of tables to control scanning processing. When a booklet cover is scanned, the program uses the booklet number and the first table to determine which blocks are to be processed. Each block code, in turn, is referred to the second table to determine the number, formats, and sequence of its constituent pages. By reading the booklet cover, the program "knows" which pages would follow and in what order.

The scanning program rejects a page if it is unreadable or out of sequence. A page is unreadable if the timing or identification marks have been corrupted by either tearing, improper trimming, or confusing stray pencil marks. If the unreadable page happens to be a booklet cover, the operator instructs the scanner to send the remaining pages of that booklet into the shunt hopper, places the pages perpendicularly on top of the output stack, and resumes processing with the next booklet. For any other page type, the operator instructs the program to substitute question marks for the data values on the unreadable page and proceed with the next page.

Pages out of sequence are generally attributable to collating errors in printing. When the program encounters this type of error, the operator directs the scanner to shunt the remaining pages of the booklet and places them perpendicularly on the output stack.

The scanning program writes three types of data records onto the magnetic tape. The first is a batch header record, containing information gridded onto the batch header sheet by receipt processing staff. The second is a data record containing all of the translated marked ovals from all pages within a booklet The third type is a dummy data record, serving as a place holder in the file for a booklet with an unreadable cover sheet. The origin code is a data field written in the same location on all records to distinguish them by type.

The batch header record precedes all data records for a given batch. As the scanning program processes the header sheet, it retains the batch identification code and initializes a sequence number or counter for that batch. The batch identification code and sequence number are written to each record; the batch header record always receives a sequence number of one, the first data record is assigned number two, and so forth. The scanning machine is directed to stamp the batch identification code and sequence number on each page of a booklet. This process greatly facilitates the location of individual pages within batches by resolution staff.

Each data record is formed by collecting the transcribed marked ovals from each page of a booklet, placing them into a buffer area within the program, and writing the buffer to tape when the last page of the booklet has been processed. Several options were considered in designing the format of the output data records. A format that requires a fixed column position for each item response value would be very large, because of the number of items in the assessment, and very sparse, because of the BIB spiral design. A format that has the response data strung out in contiguous fields across items and blocks is more consistent with the format of the data records in the NAEP data entry system, but would be difficult to check in listings for quality control. The format adopted for this assessment has fixed column positions for the booklet cover data fields and scorer identification codes. The response data starts at fixed positions for each block within an instrument, and the item responses are arranged in contiguous fields.

The data values from the booklet covers and scorer identification fields are coded as numeric data. Unmarked fields are coded as hyphens (-) except for the race/ethnicity, gender, grade, and birth date fields, which are returned as question marks (?) to alert processing staff of missing or uncoded critical data. Fields that have multiple marks are coded as asterisks (*). The data values for the item responses and scores are returned as alphabetic codes. The multiple-choice, single-response format items are assigned codes depending on the position of the response alternative; that is, the first choice is assigned the code "A", the second "B", and so forth. The circle-all-that-apply items are given as many data fields as response alternatives, the marked choices are coded as "A" and the unmarked choices as hyphens. The open-ended items have 10 ovals labeled from zero to nine; a marked zero is coded as "A", a marked one as "B", and so on up to "J". As with the cover data fields, unmarked responses are coded as hyphens and multiple marks as asterisks. The fields from unreadable pages are coded as question marks again as a flag for resolution staff to correct.

6.3.2 DATA LOADING SUBSYSTEM

Each magnetic tape produced by the scanning system contains data for one or more assessment sessions for one of the age groups. The data records on these tapes conform to a fixed format. These data now have to be edited for type and range of response, transformed to a compressed format compatible with the data entry system files, and loaded into the database for resolution processing. A procedure for accomplishing all of these tasks was designed and developed for this assessment.

The data records on the scanning output tape are ordered in the same sequence as the paper materials were processed by the scanner. A record for the batch header precedes all data records belonging to that batch; each set of records belonging to one batch are separated from the others by its batch header record. The origin code field on each record serves to distinguish the header records from the data records.

The processing of each batch begins with the identification of the header record. The batch identification number on this header record provides the link between the subsequent data records on the tape and the tracking file generated by the school worksheet entry program in the data entry system. The load program uses the batch identification number to locate and retrieve the processing information for that batch from the tracking file. The program then verifies that it had the correct batch by comparing the PSU, school, and session codes gridded on the header record with the same codes in the tracking record.

If a batch code can not be located in the tracking file, the program generates a new tracking record, using only the information contained on the header record, and records this condition on an error log file. If a batch code is located but the school or session codes do not agree, the program records this conflict in the error log and continues processing.

The batch header record also contains the date that the session materials were batched together, and the number of booklets batched by the receipt processor. This information is transferred to the tracking record for later processing and reporting.

The reading of a batch header record also initiates the generation of two new files in the entry system database: the data file and the audit file. As the program processes each record within a batch from the tape file, it writes the edited and reformatted data records to the data file and records all errors and special codes in the audit file. The data fields on an audit file record identify each data problem by the batch sequence number, booklet serial number, section or block code, field name or item number, and data value. The program generates a listing of the data problems after each batch has been processed, to be printed at the termination of the program.

As the program processes each data record, it first reads the booklet number and checks it against the batch session code for appropriate session type (main or bridge). Any mismatch is recorded in the error log and processing continues. The booklet number is then compared against the first two digits of the student identification number. If they disagree, because of improper gridding, a message is written to the error log and the booklet number is substituted for part of the student number The remaining booklet cover fields are then read and validated for range. The PSU and school codes must be identical to those on the tracking record; the range of grade codes is dependent on the age cohort being processed; and the range of birth dates is dependent upon the session type as well as the age cohort. All data values that are out of range are replaced with question marks and recorded on the audit file. All data fields that are read in as question marks or asterisks are also recorded in the audit file. The booklet cover data fields are written to a batch listing file that is printed at the end of load processing. This listing can be compared against the administration schedule to assist in resolving booklet cover data problems.

The scorer identification fields are processed at this point and certain checks are made. If a booklet contains any open-ended items, the first scorer field should be filled. If a booklet is part of the reliability sample, the second scorer field should be filled. The program has to determine from the booklet number whether the booklet contains any open-ended items. It then flags as erroneous any incomplete field that should be filled, or any nonblank field that should be blank and records the error in the audit file. Further, it remembers how many scorer fields are marked for later processing of the open-ended item scores.

The edited booklet cover and scorer identification fields are appended to the batch sequence number and transferred to an output buffer area within the program. As the program processes each block of data from the tape record, it appends the edited data fields to the data already in this buffer. The output data record in this "compressed" format, is thus made compatible with the NAEP data entry system.

The program is now prepared to cycle through the data areas corresponding to the item blocks. The task of translating, validating, and reporting errors for each data field in each block is performed by a subroutine that requires only the block identification code, the string of input data, and the number of scorers who gridded the appropriate identification fields for that block. This routine has access to an internal table that has, for each block, the number of fields to be processed, and, for each field the field type (alphabetic or numeric), the field width in the data record and the valid range of values. The routine then processes each field in sequence order, performing the necessary translation, validation, and reporting tasks.

The first of these tasks checks for the presence of hyphens, asterisks or question marks. Fields containing asterisks and question marks are recorded in the audit file and processing continues with the next field. No action is taken on hyphen-filled fields inasmuch as that code indicates a nonresponse. The field type code dictates whether numeric or alphabetic codes are to be output for a data field. The next step examines the type code and translates the input data from alphabetic to numeric if so indicated. The field is then validated for range of response, recording anything outside of that range to the audit file. The field type code is used by the program to make a further distinction among open-ended item scores and other numeric data fields. If the data field is an open-ended item, the routine uses the passed value of the number of scorers to determine whether a score should be marked. If no scorer codes are indicated and the item is marked, or a scorer code was gridded and the item is not marked, the disparity is noted in the audit file. The routine then looks ahead to the next field for a secondary scoring and compares its presence against the absence of a second scorer code, and vice versa, and again records a disparity in the audit file. Moving the translated and edited data field into the output buffer is the last task performed in this phase of processing.

The routine passes the edited data string back to the program, which then appends it to the current output buffer and sets up to process the next block within the booklet. The completed string of data is written to the data file, using the batch sequence number as the key for direct access by the entry system programs.

When the next batch header record or end of file is encountered, the program closes the data and audit files, generates an audit listing, and writes a count of the number of records processed to the message log. The program then updates the tracking record for that batch with the current date and time and the record count, and rewrites the record to the tracking file.

When the program encounters the end of the tape data file, it closes and rewinds the tape file, closes the tracking file, and transmits the message log, the audit listing, and the batch listing to the printer.

6.3.3 DATA ENTRY AND RESOLUTION SYSTEM

The data entry and resolution system is essentially the same system as that used in the 1986 assessment, modified to accommodate changes in the assessment design and data entry operations. The modified system must be able to process the materials from three age groups simultaneously, accommodate the separation and merging of scannable booklets and answer sheets, and permit the loading of excluded student questionnaire data from scanning tapes.

The system comprised separate programs for each main function (school worksheet entry, student data entry/resolution, and questionnaire entry/resolution). This separation permits the modification or enhancement of one component while allowing the others to operate. Access to these programs is controlled through a menu-type procedure written in the VAX command language and using screen control directives.

The use of batch identification codes instead of PSU/school/session codes and of batch sequence numbers for student identification codes as index keys for the tracking and data files, respectively, greatly facilitates the management of the system and correction of incorrectly gridded or keyed information.

Another addition to the batch data records is the data entry status codes. The records in a batch file are generated in one of two ways: the tape loading program or the manual entry of the booklets rejected by the scanner. The manually entered records have to undergo the two-step entry and verification processes. because of the high accuracy rate of the optical scanner, the loaded records are treated by the system as if they had undergone verification. The entry status code is used to distinguish between records that are undergoing manual processing and those that were loaded. The code on each record is tested and set by the different processes: entry, verification, loading, and resolution.

The form parameters, which control processing of each data entry screen, are maintained in a text library. Each set of parameters for each form are stored as a separate member or subfile within this library. This format permits easy extraction, modification, and replacement of parameter information as well as faster access by the entry programs. A set of programs was developed to facilitate the entry, documentation, and editing of the form parameter data.

Chapter 6.4

EDITING DATA

David S. Freund and Alfred M. Rogers

Educational Testing Service

The data editing process is divided into three separate steps: validation, identification, and correction. Validation ensures that each data value in the computer file is of the correct type, falls within a range or set of ranges of values, and is consistent with other data values. All invalid data values are identified and located in the raw data and either corrected or flagged as unresolvable in the computer file.

The errors uncovered by the editing process fall into two types: those made by the respondent (e.g., choosing two responses for a multiple-choice exercise requiring only one response) and those made by data entry. The validation process reports both types of error with no knowledge of their source. The identification process determines the type of each error. The data entry errors are, for the most part, correctable; the correct value can be determined from an examination of the information on the respondent's booklet or answer sheet. Errors made by the respondent, however, are difficult, if not impossible, to correct. If the intent of the respondent cannot be determined, the error must remain unresolved, but must be flagged in some way to prevent incorrect interpretation in the analysis and reporting procedures.

6.4.1 ABSENTEE DATA

As described in section 6.1.3, the absentee data (data for those students who were absent on the day of the assessment) were transcribed by the NAEP data entry system from the administration schedules. Validation of this data consisted of matching the school and session codes with those in the tracking file and checking that the sex, grade, and birth date codes were within the appropriate ranges for age cohort and session type. A further check performed on these files compared the number of absentee records within each session against the absentee count field on the corresponding tracking record.

The corrected file was again processed by the validation program to ensure that all errors had been fixed and that no new problems were created in the process. If further errors were uncovered, the cycle of identifying the records, correcting the errors, and validating the corrected file was repeated until no more errors were found. At this point, the absentee file was ready for transmittal to Westat for the estimation of sampling weights.

6.4.2 STUDENT DATA

The use of scannable materials, first introduced in the 1986 assessment in the form of scannable booklets and enhanced in 1988 to include scannable answer sheets, greatly improved the efficie y and accuracy of the transcription process by removing the possibility of human error. The scanning machinery was programmed to detect the marked responses in unique and fixed positions on each page; erroneous and out-of-range response codes could not be generated.

On the other hand, removing human intervention as a source of error also prevented the ex rcise of human judgment when more than one mark was detected for a single-response item. This would commonly happen when a student marked a second response without erasing the fir , or when a student misinterpreted the question as a "circle all that apply" ,pe response. Neither the human eye nor the scanning equipment can determine the student's intent in such a situation. However, the scanning program would also return a multiple-response code if the student had incompletely erased the first response or inadvertently made a stray mark on one of the ovals (situations in which the human eye could determine the intended response). Hence there were proportionally many more multiple-response codes produced by the scanning process than by the manual entry process.

Furthermore, collating errors in th printing of the booklets resulted in both missing and multiple pages, which the scanning program was unprepared to handle. A new code was used to designate responses to items from pages that were missing or otherwise unscannable.

Every multiple-response code and unscannable-page code had to be checked against the respondent's booklet or answer sheet and, where possibl , corrected by resolution operators. At the completion of resolution processing, all of the batch student data files were moved to a single master file in preparation for transfer to the IBM mainframe. A second validation was performed during this spooling process to catch errors that had slipped through the entry system undetected. An editing program was developed for applying corrections to this master file, using the same method as was used for the data entry program. This master file also served as the basis for preliminary descriptive data analyses and quality control checks.

6.4.3 QUESTIONNAIRE DATA

The data entry system was used for the entry of school teacher, and excluded student questionnaire data and served as the first ine of defense against bad data. As described above, all d ta values were validated for type and range as they were entered from the data terminal keyboard. Special codes assigned for multiple and indeterminate responses were recorded and reported via the audit trail. The indeterminate values were later corrected under the resolution process.

The questionnaire files received the same secondary validation processing as the student data. Special attention was given to the "circle all that apply"-type items to ensure consistency in the coding of responses. If a respondent circled one or more of the alternatives, those would be coded "1" while the rest would be coded "0"; if no alternatives were marked, yet the respondent had the opportunity to reply, all fields would be coded "0"; if no alternatives were marked and the respondent had not reached the item or was instructed to skip it, all fields would be coded as "no response."

6.4.4 PROFESSIONALLY SCORED ITEM DATA

The open-ended primary and secondary trait reading and writing items responses and open-ended mathematics, science, civics, and U.S. history item responses were read and scored prior to scanning processing. Their data values were subjected to the same editing procedures as the multiple-choice item responses. The open-ended holistic trait writing items, however, were not scored until after scanning and resolution processing. It was not feasible to enter so few scores for each booklet through the entry system, so these data were subjected to a separate entry and editing process.

The booklets that contained holistic writing items were batched and forwarded to ETS key entry systems where they were entered, verified, and transcribed to magnetic tape. The holistic scores and scorer ID numbers were recorded by the scorers on the back of the booklets. Twenty percent of the booklets were subjected to a second set of scores for use in calculating rater reliabilities. These scores and the student ID number and PSU and school codes from the front cover of the booklet were entered by the key entry operator. These tape files were loaded onto the IBM computer system where specially written validation programs performed thorough checks on the data values. When all of the items had been scored, entered, validated, and corrected, the data files were merged with the student database.

6.4.5 CONCLUSION

Before the NAEP data entry methodology was developed, the editing process for any data file proceeded in the same manner as for the absentee data and professionally scored computer items. The validation process was especially inefficient because it was performed after transcription and often by a second party who did not have immediate access to the respondent's booklet or answer sheet. Putting the validation mechanism at the point of entry removed most, if not all, of this inefficiency by informing the entry operator of a possible keying error while the respondent's booklet or answer sheet was accessible.

The editing process does not guarantee that all errors are removed from the data; only that the invalid, inconsistent, or otherwise unreasonable values have been at least identified, if not corrected. If a data value has been miskeyed during the entry process and meets the validation criteria, this error could persist through the editing process to the analysis stage without detection. The verification process detects most of these errors by comparing independent entries of the same data and reporting disagreements. The likelihood of an error surviving verification is thus very small, but still present. A quality control process must follow the entry and editing processes to ensure that the data values in a given record agree with the responses in the corresponding instrument.

Chapter 6.5

QUALITY CONTROL OF NAEP DATA ENTRY FOR 1988

John J. Ferris

Educational Testing Service

Great care was taken to control the NAEP data entry process and the quality of the data entered thereby. The result was an extremely high quality database, that is, one with extremely low error rates. It is of course necessary to establish the quality of any data that will be analyzed, since the most thorough and sophisticated analyses of bad data will yield nothing of value. As in past years, this NAEP database was found to be more than accurate enough to support as sensitive an analysis as may be desired. It is worth noting that the analyses done with the NAEP data are only intended to apply to groups of respondents, rather than to individuals, the fact that such analyses tend to be more tolerant of data errors further enhances the already high quality of the NAEP database.

The purpose of the analysis reported in this chapter was to assess the quality of the data resulting from the complete data entry system, from the actual instruments collected in the field to the final machine readable database used in the analyses. The process involved the selection of instruments at random from among those returned from the field and the comparison of these instruments, character by character, with their representations in the final database. In this way, we were able to measure the error rates in the data as well as the success of the data entry system.

Of course the observed error rate cannot be taken at face value. For example, the sample of school characteristics and policies questionnaires that happened to be selected for close inspection contained no errors at all. To conclude that the entire school characteristics questionnaire database is therefore error free would be an act of extreme optimism; we may simply have been lucky with this particular random sample. What is needed is an indication of how bad the true error rate might be given what we observed. Such an indication is provided by confidence limits. Confidence limits indicate how likely it is that a value will fall outside a specified range of values in a specified context or distribution. In our analysis, the specified range is an error rate between zero and some maximum value beyond which we are confident that the true error rate does not lie; the specified context or distribution turns out to be the cumulative binomial probability distribution. An example should demonstrate this technique:

> Let us say that 1,000 booklets were processed, each with 100 characters of data transcribed for a total of 100,000 characters. Let us say further that five of these characters were discovered to be in error in a random sample of 50 booklets that were completely checked; in other words, five errors were found in a

sample of 5,000 characters. The following expression may be used to establish the probability that the true error rate is .0025 or less, rather than the single-value estimate of the observed rate of one in a thousand (.001):

$$\sum_{j=0}^{5} \binom{5000}{j} \cdot .0025^{j} \cdot (1-.0025)^{(5000-j)} = .0147$$

This is the sum of the probability of finding five errors plus the probability of finding four errors plus. . . etc. . . plus the probability of finding zero errors in a sample of 5,000 with a true error rate of .0025, that is, the probability of finding five or fewer errors by chance when the true error rate is .0025. Notice that we did not use the size of the database in this expression. Actually, the assumption here is that our sample of 5,000 was drawn from a database that is infinite. The smaller the actual database is, the more confidence we can have in the observed error rate; had there been only 5,000 in the total database, our sample would have included all the data nd the observed error rate would have been the true error rate. The result of the above computation allows us to say, conservatively, that .0025 is an upper limit on the true error rate with 98.53 p ent (i.e., 1 - .0147) confidence; that is, we are quite sure th our true error rate is no larger than .0025.

The individual instruments are briefly discussed in the following sections and a summary table (Table 6.5-1) gives the upper 99.8 percent confidence limits for the error rates for each of the instruments as well as sampling rate information. The confidence limit of 99.8 percent was selected to make these results comparable to those of previous administrations when the same limit was used.

6.5.1 STUDENT DATA

In recent past assessments, only one each of the various booklets was sampled for this error rate analysis. Due to the complexity of the current assessment, a larger number of each booklet was examined. In all, over 300 booklets out of a total of about 120,000 were compared in detail with the final database. Across all scannable student data, only about 1 percent of the booklets or answer sheets could not be scanned and had to be keyed by hand; we did not attempt to sample this small group of booklets separately for quality analysis and relied instead on the bridges to 1984 to assess our keying operation, since these booklets (reading and writing booklets 51-56) were entirely keyed. In the past, keying error rates for instruments designed to be scanned have actually been somewhat better than keying error rates for instruments designed to be keyed. The summary table gives the error results across all three age classes; there were no noticeable differences among age classes.

Table 6.5-1

Summary of Quality Control Error Analysis for NAEP 1988 Data Entry

Instrument/Subsample	Entry type	Diff. Books	# Books Sampled	# Chars. Sampled	# Errors	Observed Error Rate	Upper 99.8% Confidence Limit
Student Data - Main	Scanned	97	217	29,172	3	.0001	.0004
Student Data - Reading and Writing Bridge	Keyed	18	54	9,201	13	.0014	.0030
Student Data - Other Bridges	Scanned	15	34	6,311	2	.0003	.0017
Excluded Student Questionnaire	Mixed	2	24	2,064	2	.0010	.0051
Teacher Questionnaire	Keyed	1	8	3,732	1	.0003	.0023
School Characteristics and Policies Questionnaire	Keyed	1	6	1,532	0	zero	.0041

6.5.2 EXCLUDED STUDENT QUESTIONNAIRE DATA

A total of 7,791 questionnaires was scanned in this assessment and this group was sampled at a rate of about .25 percent; the 850 questionnaires (about one-tenth of all the excluded student questionnaire data) from the fall bridge were keyed and these were sampled at double this rate. The few errors that were discovered in the scanned data were caused by the scanning machine. Respondents who change their answers do not always erase to the satisfaction of a scanner, and this caused an occasional misreading of a response.

6.5.3 TEACHER QUESTIONNAIRE DATA

There were 1,664 teacher questionnaires collected in this assessment. In the eight that were selected at random for a complete verification, one error was discovered.

6.5.4 SCHOOL CHARACTERISTICS AND POLICIES QUESTIONNAIRE DATA

There were 1,425 school characteristics and policies questionnaires collected in this assessment. No errors were found in the six questionnaires that were checked.

Chapter 6.6

CREATION OF THE 1988 NAEP DATABASE

David S. Freund and Alfred M. Rogers

Educational Testing Service

The data transcription and editing procedures described in Chapter 6.1 resulted in the generation of disk and tape files containing various assessment information. Before any analysis could begin, these files had to be brought together into a comprehensive, integrated database. Sampling weights were also required in order to make valid statistical inferences about the population from which the assessment sample was drawn.

This chapter describes the processes of extracting sample information for the derivation of sampling weights, and merging, or bringing together, the many transcription files into the NAEP database.

6.6.1 EXTRACTING SAMPLE DATA TO DERIVE WEIGHTS

For each grade/age cohort, up to four sets of weights were required to perform inferential analyses: school weights, excluded student weights, student weights, and teacher weights (age 17 did not include any teacher data). Because of the method of selecting teachers, sampling weights could not be assigned to teachers, but were instead assigned to students who were linked to participating teachers. (See Chapter 3 for more details.)

All of the sample information was extracted from the data files, edited, and transferred to tape files for shipment to Westat, where the weights were computed. (See Chapter 8 for details on computing weights.) The editing process included both the validation of the data values (verification that each data value falls within a range or set of ranges of values, and is consistent with other data values) and frequency distribution analyses containing counts of the number of students assessed for each session to be compared with tracking information from the data entry system.

The school sample information, such as PSU and school number, school type, and sampling description of community (SDOC), was available to Westat from the beginning of the assessment. No other information was required to compute school sample weights.

The excluded student sample information was extracted from the file of excluded student questionnaire data. This information included questionnaire serial number, PSU and school code, grade, gender, birth date, race/ethnicity, and a code indicating reason for exclusion. All data fields were taken from the front cover of each questionnaire, except for the exclusion code, which was derived from the response to item 2 ("Why is this student excluded from the NAEP survey?") of the questionnaire. A listing of the excluded student questionnaires that had not been received at ETS was included with the file for each grade/age cohort.

The student sample information came from two sources: the student database and the absentee file from the administration schedules. The assessed student sample information included booklet serial number, PSU and school code, grade, gender, birth date, and race/ethnicity. Since the absent students were not observed and not assigned an assessment booklet, the booklet serial number and race/ethnicity information were not part of the absentee data.

The absentee file had to be adjusted for makeup sessions. The field administration procedures required scheduling of makeup sessions if absentee rates exceeded certain limits. The students attending these makeup sessions were supposed to be originally sampled students who were absent for the regular sessions. Failure to remove the makeup students from the absentee file would have resulted in incorrect estimates of the number of students in those schools. The effect of these errors could have been particularly acute in the age 17 sample where absentee rates were high and many schools required makeup sessions.

The first step in the removal process was to identify the students in the student file who attended makeup sessions in each school. Then, for each school and session type (spiral or tape), the gender, grade, and birth dates of the makeup students were matched with those of the absentee students in the same school and session type. The absentees identified by perfect matches were removed from the absentee file: For each unmatched makeup student, a randomly selected absentee was removed from the file. This latter procedure was necessary only for the age [illegible] sample in only a few of the many schools that had makeup sessions.

The teacher sample information was extracted from the teacher questionnaire data file. The teacher identified up to ten students who participated in the assessment and met the proper criteria, i.e., for the grade 4/age 9 cohort, they were in the fourth grade and part of the focused-BIB reading assessment (booklets 8-14) and for the grade 8/age 13 cohort, they were in the eighth grade and took the focused-BIB writing assessment (booklets 1-7). (See Chapter 3.7 for more details.) The information used by Westat to produce student-based teacher weights included the PSU, school, teacher code, booklet number, birth date, race/ethnicity, and gender for each identified student in that teacher's class.

6.6.2 MERGING FILES INTO THE NAEP DATABASE

The transcription process resulted in the generation of up to five data files for each grade/age cohort: one file for each of the three questionnaires (no teacher questionnaire for age 17), the student response data file from the data entry system, and the student holistic writing scores from professional scoring and key entry. The process of deriving sample weights produced an additional four files (three at age 17) of sampling weights. Before data analyses could be performed, these files had to be integrated into a coherent and comprehensive database.

The database ultimately comprised up to four files per cohort: school, teacher, excluded student, and student files. The student file contained data from all student samples—the main assessment, the bridge to 1984, the bridge to 1986, and the civics bridge to 1976 and 1982. The school file could be linked to the other three files (student, excluded student, and student-based teacher) through the PSU and school codes. The student-based teacher file could be linked to a subset of the student main sample through the student booklet number, PSU, school, and teacher codes.

The school file was created by merging the school questionnaire file with the school weights file and with a file of school variables supplied by Westat which included demographic information about the schools that was originally collected by Quality Education Data, Inc. (QED). The PSU and school code were used as the matching criteria. Each record of the resulting file was formed by merging the weight information with the response data and the QED data. Since not all schools returned their questionnaires and/or were missing QED data, some of the output records contained only school identifying information and weight information.

The teacher file was generated from the teacher questionnaire file. Since the teacher weights were derived at the student level, no information had to be added to the questionnaire data.

The excluded student file was the result of merging the excluded student questionnaire file with the excluded student weights file. The booklet serial number was used as the matching criterion.

The student data were created in three steps, merging the student response data with the student weights, the student-based teacher weights, and professionally scored holistic writing item scores, in that order. In all three steps, the booklet serial number was used as the matching criterion. The merging of the professionally scored item data was a more complex procedure than the others, because only a subset of the student data records contained this data and for those records that did contain data, the item scores appeared in a different location in each booklet.

When the appropriate files had been merged, the database was ready for analysis. Any time that new data values, such as plausible values, were derived external to the database, they were added to the relevant files using the same matching procedures as described above. The public-use data tapes files were later generated from this database.

6.6.3 CREATING THE MASTER CATALOG

A critical part of any database is its processing control and descriptive information. A central repository of this information may be accessed by all analysis and reporting programs to provide correct parameters for processing the data fields as well as to provide consistent labeling to identify the results of the analyses. The NAEP master catalog file was designed and constructed to serve both of these purposes.

Each record of the master catalog contains the processing, labeling, classification, and location information for each data field in the NAEP database. The control parameters are used by the access routines in the analysis programs to define the manner in which the data values are to be transformed and processed.

All data fields have a 50-character label in the catalog describing the contents of the field and, where applicable, the source of the field. The data fields with discrete or categorical values (e.g., multiple choice items and professionally score items, but not weight fields) have additional label fields in the catalog containing 8- and 20-character labels for those values.

The classification area of the catalog record contains distinct fields corresponding to predefined classification categories for the data fields. For a given classification field, a nonblank value indicates the code within that classification category for the data field. This permits the collection of identically classified items or data fields by performing a selection process on one or more classification fields in the catalog.

According to the NAEP design, it is possible for item data fields to appear in more than one student sample and in more than one block within each sample. The location fields of the catalog record contain the age, block and, where applicable, the sequence within the block for each appearance of the data field. (Fields such as plausible values and weights would not contain sequence numbers since these fields are not pertinent to a given block.)

The master catalog file was constructed in parallel with the collection and transcription of the assessment data to be ready for use by analysis programs when the database was created. As new data fields were derived and added to the database, their corresponding descriptive and control information were entered into the catalog.

One of the most important uses of the master catalog was the control of the creation of the public-use data tapes files, codebooks, and file layouts. A synopsis of this process is presented in the next chapter.

Chapter 6.7

NAEP DATABASE PRODUCTS

Alfred M. Rogers

Educational Testing Service

The NAEP database described to this point serves primarily to support analysis and reporting activities that are directly related to the NAEP grant. This database has a singular structure and access methodology that is integrated with the NAEP analysis and reporting programs. One of the directives of the NAEP grant is to provide secondary researchers with a nonproprietary version of the database that is portable to any computer system. In the event of transfer of NAEP to another client, the grant further requires ETS to provide a full copy of the internal database in a format that may be installed on a different computer system.

In fulfillment of these requirements, ETS provides three sets of database products: the item information database, the restricted-use data files, and the public-use data files. The contents, format and usage of these products are documented in the publications listed under the appropriate sections below.

6.7.1 THE ITEM INFORMATION DATABASE

The NAEP m information database contains l of the descriptive, processing, ar ge information for every assessment item developed and used for NAEP since v. The primary unit of this database is the item. Each NAEP item is associated with different levels of information, including usage across years and age cohorts, subject area classifications, response category descriptors, and locations of response data on public-use data files.

The item information database is used for a variety of essential NAEP tasks: providing statistical information to aid in test construction, determining the usage of items across assessment years and ages for trend and cross-sectional analyses, labeling summary analyses and reports, and organizing items by subject area classifications for scaling analysis.

The creation, structure, and use of the NAEP item information database for all items used up to and including the 1988 assessment are fully documented in the NAEP publications, *A Guide to the NAEP Item Information Database* (Rogers, Barone, & Kline, 1990) and *A Primer for the NAEP Item Information Database* (Rogers, Kline, Barone, Mychajlowycz, & Forer, 1989).

The procedures used to create the 1988 version of the item information database are the same as those documented in the guide. The updated version of the guide also contains the learning area classification categories for the cognitive items.

6.7.2 THE RESTRICTED-USE DATA FILES

The restricted-use data files are for the exclusive internal use of the NAEP grantee. They contain a complete copy of the internal NAEP respondent database in a structured, documented, and portable format.

The internal database is maintained in a compressed format to conserve computing resources and to increase analysis efficiency. The access methods developed for this database locate data fields dynamically during the execution of analysis programs. The restricted-use data files, on the other hand, are "rectangular" in structure; each data field is in the same location on every record within a file. This static data definition, while not efficient from a computing resource standpoint, is much easier to document and is not dependent on any computing machinery, operating system, or data access method.

The restricted-use data files serve several critical purposes. They provide an archive for all respondent data collected and derived for NAEP since 1970. They ensure compatibility of usage by expressing this data in consistent, rectangular formats. Their portability greatly facilitates transition of the respondent database to future NAEP contractors. The accompanying data file layouts and codebooks provide a standardized, comprehensive reference source for NAEP staff.

The contents and formats of the NAEP restricted-use data files are documented in the NAEP publication *A Guide to the NAEP Restricted-use Data Files* (Rogers, Barone, & Kline, 1989).

The procedures used to create the restricted-use data files for the 1988 assessment are the same as those used to create the public-use data files. Since the public-use data file distribution package contains more products, the generation procedures will be described in the following section.

6.7.3 THE PUBLIC-USE DATA FILES

The public-use data files are designed to enable any researcher with an interest in the National Assessment database to perform secondary analysis on the same data as those used at ETS. They differ from the restricted-use data files in one important respect: all subregional identification information has been encrypted or excluded in order to maintain the confidentiality of the states, schools, and students who participated in the assessment.

The three elements of the distribution package are the data tapes, the printed documentation, and the microfiche copies of the assessment instruments. The complete set of files for each age cohort resides on a separate tape. Each tape contains, for each sample or instrument, the data file, a file of control statements that will generate an SPSS-X system file, a file of control statements that will generate a SAS system file, and a machine-readable catalog file containing control and descriptive information, intended for the user who does not use either SAS or SPSS-X. The printed documentation consists of four volumes: a guide to the use of the data files, and a set of data file layouts and codebooks for each of the three age cohorts (see *The NAEP 1988 Public-use Data Tapes Version 2.0 User Guide* [Rogers, Kline, Johnson, Mislevy, & Rust, 1990]).

The remainder of this section will discuss some of the issues raised during the creation of the data files and summarizes the procedures followed in generating the data files and related materials.

6.7.3.1 File Definition

The first issue addressed in the production of the public-use data files was the organization and format of the data files. The NAEP database consists of four data files for each grade/age cohort, corresponding to the three questionnaire instruments and the student database, incorporating the main sample and all five bridge samples. The logical relationship of the data files is a three-level hierarchy, with the six student and the excluded student samples at the lowest level; the teacher sample at the next level, with a linkage only to the main sample; and the school sample at the top, with direct linkages to all samples. A linkage may be viewed as a one-to-many mapping of the records within two files. For example, one school record can link to one or more records in the teacher file, and each of these teacher records can in turn link to one or more records in the main sample student file.

Two organization schemes were considered. The first scheme, using the concept of a static linkage, requires only seven files corresponding to the seven student samples at the lowest level of the hierarchy. All of the data from the higher-level samples would be appended to and repeated across as many of the lower-level records as dictated by the linkages. Using the previous example, each main sample record would be appended by its corresponding teacher record and school record. This scheme places no demand on the user to define the linkages since each data record is complete, but, because of its larger record size, requires substantially more computer storage space.

The second scheme, employing a dynamic linkage, requires these same seven samples, but without the appended teacher and school data. The teacher and school sample data would reside in their own files, with special data fields in all files to facilitate their linkage through program control. This approach is more economical in computer resource utilization but assumes a more sophisticated user. The potential for savings in computer storage and processing costs was the overriding consideration in choosing this scheme.

The teacher questionnaire for the 1988 assessment contained one section that provided a direct link to individual students in the main sample. The file generated for the teacher sample, therefore, was based on student-level data for those linked students, with the entire teacher questionnaire response data appended. The benefits gained by doing this are threefold: analysis of teacher response data can be properly performed at the student level and with the appropriate sample weights; the student-based teacher weight fields need not be present on the main sample student file, as in previous assessments, and the user is freed from programming linkages between the teacher and main sample student files.

6.7.3.2 Definition of the Variables

The selection and arrangement of data field, or variables, in each file was the next issue addressed. The initial step in this process was the generation of a file of descriptors of the variables for each data file to be created. Each of these LABELS files contained one record for each variable, each record containing the variable name, a short description of the variable, and processing control information to be used by later steps in the data generation process. This file could be edited for deletion of variables, modification of control parameters, or reordering of the variables within the file.

The first program in the processing stream, GENLYT, produced a printed layout for each file from the information in its corresponding LABELS file. These layouts were initially reviewed for the selection and ordering of the variables. The variables that were excluded from public-use data file processing fell primarily into two categories: nonapplicable and confidential.

The nonapplicable variables were found mostly in the student database. In the database used for analysis and reporting, the bridge samples were combined with the main sample. Therefore, many of the variables that applied to the main sample students did not apply to the bridge sample students, and vice versa. For example, the teacher code and the student-based teacher weights were used for the analysis of main sample data, but were not used at all in the design for the bridge samples.

The confidential variables included any descriptor or code that could be used to identify individual states, schools, or students in the NAEP sample. The PSU, school, teacher, and student identification codes used internally by ETS and Westat were "scrambled" according to specific algorithms to obtain new codes for use in linking the files together. These new codes were put on the tapes in lieu of the original codes.

Another confidentiality issue arose for an item for which students were asked to identify the state they had lived in four years prior to the assessment. A new variable was created using the student's response and current state residency information from the PSU code to determine whether the student had lived in the same state, the same region, or a different region.

The ordering of the variables within the data files followed a general trend of decreasing likelihood of usage. In this order of likelihood, identification information preceded weights, scores, and other derived variables, which were followed by the response data. The identification variables were generally those on the front covers of the instruments. The derived variables included the sampling weights, the IRT scale values, and the variables that were derived from the response data or other sources for the purpose of reporting. The response data variables were arranged according to their order in the instrument.

The data for the main sample posed an additional challenge with its multitude of booklet formats that had to be structured into a single, fixed format. The most convenient and economical solution was to arrange the "blocks" of item response data in order within subject areas. The responses to the common background questionnaire preceded all other blocks in the new record. The remaining blocks were grouped by subject area, each group consisting of the subject area background block followed by the cognitive blocks in numerical order. Each record from the input student data file was reformatted according to its booklet number; the data for its constituent blocks were moved into their assigned locations in the output record. The remaining data block areas contained blank fields, signifying that the data were missing by design.

In order to process and analyze the spiral sample data effectively, the user must also be able to determine, from a given booklet record, which blocks of item response data were present and their relative order in the instrument. This problem was remedied by the creation of a set of control variables, one for each block, which indicated not only the presence or absence of the block but its order in the instrument. These control variables were included with the derived variables.

6.7.3.3 Data Definition

To enable the data files to be processed on any computer system using any procedural or programming language, it was desirable that the data be expressed in numeric format. This was possible, but not without the adoption of certain conventions for reexpressing the data values.

As mentioned in Chapter 6.3, the responses to all multiple-choice items were transcribed and stored in the database using the letter codes printed in the instruments. This scheme afforded the advantage of saving storage space for items with ten or more response options, but at the expense of translating these codes into their numeric equivalents for analysis purposes. The response data fields for most of these items would require a simple alphabetic-to-numeric conversion. However, the data fields for items with ten or more response choices would require "expansion" before the conversion, since the numeric value would require two column positions. One of the processing control parameters on the LABELS file indicates whether or not the data field is to be expanded before conversion and output.

The ETS database contained special codes to indicate certain response conditions: "I don't know" response, multiple response, omitted response, not-reached response, and unresolvable response, which included out-of-range responses and responses that were missing due to errors in printing or processing. The primary trait scores for the reading essay and writing items included additional special codes for ratings of "illegible," "off task," and nonrateable by the scorers. All of these codes had to be reexpressed in a consistent numeric format.

The follo.ing convention was adopted and used in the designation of these codes: The "I do t know" and nonrateable response codes were always converted to 7; the omi d response codes were converted to 8; the "not reached" response codes were converted to 9; the multiple response codes were converted to 0; the "illegible" codes were converted to 5; and the "off task" codes were converted to 6. The out-of-range and missing responses were coded as blank fields, corresponding to the "missing by design" designation.

This coding scheme created conflicts for those multiple choice items that had seven or more valid response options as well as the "I don't know" response and for those open-ended items whose primary trait scoring guide had five or more categories. These data fields were also expanded to accommodate the valid response values and the s ecial codes. In these cases, the special codes were "extended" to fill the output data field: The "I don't know" and nonrateable codes were extended from 7 to 77, etc.

The numeric variables on the tape files were classified into two categories: continuous and discrete. The continuous variables include the weights, IRT values, identification codes, and item responses where counts or percentages were requested. The discrete variables include those items for which each numeric value corresponds to a response category. The designation of "discrete" also includes those derived variables to which numeric classification categories have been assigned. The open-ended items wei treated as a special subset of the discrete variables and were assigned to a separate category to facilitate their identification in the documentation.

6.7.3.4 Data File Layouts

The data file layouts, as mentioned above, were the first user product to be generated in the public-use data files process. The generation program, GENLYT, used a LABELS file as input and produced a printable fi . The LAYOUT file is little more than a formatted listing of the LABELS file.

Each line of the LAYOUT file contains the following information for a single data field: sequence number, field name, output column position, field widt., number of decimal places, data type, value range, key or correct response value, a d a short description of the field. The sequence number of each field is implied from its order on the LABELS file. The field name is an 8-character label for the field that is to be used consistently by all public-use data files materials to refer to that field on that file. The output column position is the relative location of the beginning of that field on each record for that file, using bytes or characters as the unit of measure. The field width indica es the number of columns used in representing the data values for a field. If the field contains continuous numeric data, the value under the number of decimal places entry indicates how many places to shift the decimal point before processing data values.

The data type category uses three codes to designate the nature of the data in the field. Continuous numeric data are coded "C"; discrete numeric data are coded "D"; open-ended item data are coded "O." Additionally, the discrete numeric fields that include "I don't know" response codes are coded "DI" and the open-ended items that include nonrateable response codes are coded "OI." If the field type is discrete numeric, the value range is listed as the minimum and maximum permitted values separated by a hyphen to indicate range. If the field is a response to a scorable item, the correct option value, or key, is printed. A range of correct options was indicated for those professionally scored items that were treated with cutoff scoring for IRT scaling. Finally, each variable was further identified by a 50-character descriptor.

6.7.3.5 Data File Catalogs

The LABELS file contains sufficient descriptive information for generating a brief layout of the data file. However, to generate a complete codebook document, substantially more information about the data is required The CATALOG file provides most of this information.

The CATALOG file is created by the CATGEN program from the LABELS file and the 1988 master catalog file. Each record on the LABELS file generates a CATALOG record by first retrieving the master catalog record corresponding to the field name. The master catalog record contains usage, classification, and response code information, prefixed by the positional information from the LABELS file: field sequence number, output column position, and field width.

The information for the response codes, also referred to as "foils," consists of the valid data values for the discrete numeric fiel[illegible] a 20-character description of each. The CATGEN program uses additional control information from the LABELS file to determine if extra foils should be generated and saved with each CATALOG record. The first flag controls generation of the "I don't know" or nonrateable foil; the second flag regulates omitted or "not reached" foil generation; and the third flag denotes the possibility of multiple responses for that field and sets up an appropriate foil. All of these control parameters, including the expansion flag, may be altered in the LABELS file by use of a text editor, in order to control the generation of data or descriptive information for any given field.

The LABELS file supplies control information for many of the subsequent public-use data processing steps. The CATALOG file provides detailed information for those and other steps.

6.7.3.6 Codebooks

The data file codebook is a printed document containing complete descriptive information for each data field. Most of this information originates from the CATALOG file; the remaining data came from two other files: the COUNTS file and the IRT parameters file.

Each data field receives at least one line of descriptive information in the codebook. If the data type is continuous numeric, no more detail is given. If the variable is discrete numeric, the codebook lists the foil codes, foil labels, and frequencies of each value in the data file. Additionally, if the field represents an item used in IRT scaling, the codebook lists the parameters used by the scaling program.

The frequency counts are not available on the catalog file, but must be generated from the data. The GENFREQ program creates the COUNTS file using the field name to locate the variable in the database, and the foil values to validate the range of data values for each field. This program also serves as a check on the completeness of the foils in the CATALOG file, as it flags any data values not represented by a foil value and label.

The IRT parameter file is linked to the CATALOG file through the field name. Printing of the IRT parameters is governed by a control flag in the classification section of the CATALOG record.

The LAYOUT and CODEBOOK files are written by their respective generation programs to print-image disk data files. Draft copies are printed and distributed for review before the production copy is generated. The production copy is printed on an IBM 3800 printer that uses laser-imaging technology to produce high-quality, reproducible documentation.

6.7.3.7 Control Statement Files for Statistical Packages

An additional requirement of the NAEP grant is to provide, for each public-use data file, a file of control statements each for the SAS and SPSS-X statistical systems that will convert the raw data file into the system data file for that package. Two separate programs, GENSAS and GENSPX, generate these control files using the CATALOG file as input.

Each of the control files contains separate sections for variable definition, variable labeling, missing value declaration, value labeling, and creation of scored variables from the cognitive items. The variable definition section describes the locations of the fields, by name, in the file, and, if applicable, the number of decimal places or type of data. The variable label identifies each field with a 50-character description. The missing value section identifies values of those variables that are to be treated as missing and excluded from analyses. The value labels correspond to the foils in the CATALOG file. The code values and their descriptors are listed for each discrete numeric variable. The scoring section is provided to permit the user to generate item score variables in addition to the item response variables.

Each of the code generation programs combines three steps into one complex procedure. As each CATALOG file record is read, it is broken into several component records according to the information to be used in each of the resultant sections. These record fragments are tagged with the field sequence number and a section sequence code. They are then sorted by section code and sequence number. Finally, the reorganized information is output in a structured format dictated by the syntax of the processing language.

The generation of the system files accomplishes the testing of these control statement files. The system files are saved for use by NAEP staff. These control statement files are included on the distributed data tape to permit users with access to SAS and/or SPSS-X to create their own system files.

6.7.3.8 Machine-readable Catalog Files

For those NAEP data users who have neither SAS nor SPSS-X capabilities, yet require processing control information in a computer-readable format, the distribution tape also contains machine-readable catalog files. Each machine-readable catalog record contains processing control information, IRT parameters and foil codes and labels.

PART II

The Analysis of 1988 NAEP Data

Chapter 7

OVERVIEW OF PART II: THE ANALYSIS OF 1988 NAEP DATA[1]

Rebecca Zwick

Educational Testing Service

In 1988, NAEP conducted major assessments of reading, writing, civics, and U.S. history. In each of these areas, the analyses included trend results providing links to previous assessments, as well as cross-sectional results for the 1988 assessment year, providing detailed information about student proficiency for grades 4, 8, and 12.

Another major component of the analyses of the 1988 data was the continuing research into the 1986 NAEP reading anomaly—the unexpectedly low reading proficiency results for ages 9 and 17 that spurred a three-year investigation. For purposes of investigating the anomaly further, the 1988 assessment included samples of students who were assessed with 1984 instruments and procedures, as well as samples who were assessed with 1986 booklets and methods. The study based on these samples, documented in *The Effect of Changes in the National Assessment: Disentangling the NAEP 1985-86 Reading Anomaly* (Beaton & Zwick, 1990), showed that seemingly minor changes in assessment technology had a substantial effect on estimated reading proficiency in 1986. Because the 1986 reading booklets that were administered as part of this study also contained mathematics and science blocks, these data were scaled as well.

Finally, geography was assessed for grade 12/age 17 only in a special study, co-sponsored by the National Geographic Society.

7.1 SAMPLES OF STUDENTS

The samples of students included in the 1988 NAEP assessment are listed and described in detail in Chapter 1. Only a brief description of the types of 1988 samples is given here. The 1988 samples were of two general types: bridge samples, the purpose of which was to provide links to earlier assessments, and main NAEP samples, which were based on a common set of assessment procedures, including winter and spring administration times and calendar-year age definitions.

The 1988 bridge assessments consisted of a bridge to 1984 in reading and writing, a bridge to 1986 in reading, mathematics, science, and U.S. history, and a bridge to 1976 and 1982 in civics. The 1988 main NAEP samples fell into

[1]Robert Mislevy and Norma Norris provided helpful comments on this chapter.

three categories: focused-BIB, intercorrelation, and special studies. The focused-BIB design provides for booklets that include three blocks of items in a single subject area, as well as background items. Focused-BIB assessments were conducted for reading, writing, civics, and U.S. history, and, for grade 12/age 17, geography. The intercorrelation samples received booklets that included more than one subject area to allow researchers to investigate the association between proficiencies in different subjects. The intercorrelation booklets included blocks of reading, civics, and U.S. history items at all three grade/age levels, as well as geography items at grade 12/age 17. Some additional booklets were included in the 1988 main assessment for special studies. The long writing booklets at all grade/age levels and the document literacy booklets at the two upper grade/age levels fall into this category.

7.2 ANALYSIS STEPS

The analysis methods described in the following chapters are not identical across subject areas. Procedures depend on whether data are dichotomous or ordinal and whether links across age groups or across assessments are required. Nevertheless, certain asic procedures are common to most or all of the analyses described in the following chapters; these are summarized here.

7.2.1 Item Analysis

The first analysis step in each subject area was to conduct item analyses within each grade/age cohort and within major reporting categories. These preliminary analyses had multiple purposes: to check the number of respondents, the scoring of items, and the coding of background data; to investigate the difficulty level of items and their ability to distinguish between students of high and low proficiency; to check for speededness; and to call attention to items that may have had popular but incorrect response options (indicating possible flaws in wording or scoring).

For each NAEP background item, the unweighted and weighted percent of students who gave each response were examined, along with the percent of students who omitted the item and the percent who did not reach the item. The number of respondents was also tabulated. Each block of dichotomously scored cognitive items was subjected to item analysis routines that yielded, for each item, the number of respondents, the percent of students who selected the correct response and each incorrect response, the percent who omitted the item, the percent who did not reach the item, and the correlation between the item score and the block score. In addition, summary statistics were computed for each block, including the reliability (internal consistency).

7.2.2 Analyses of Differential Item Functioning Across Age, Gender, and Racial/ethnic Categories

For subject areas that yielded dichotomous item responses, graphical techniques that are available through the NAEP's modification of the BILOG computer program (Mislevy & Bock, 1982) were used to determine whether it was reasonable to assume a common item response function across age, gender, and racial/ethnic categories. (In the present context, an item response function is the regression of a dichotomous item response on an unobserved proficiency variable. In NAEP, this nonlinear regression is assumed to take the three-parameter logistic form described in Chapter 9.) The NAEP BILOG program produces plots that show the estimated item response function for a particular sample (say, the three age classes combined). In addition, BILOG can plot expected proportions correct for specified subsamples (say, each of the three age classes) at several points along the proficiency scale (see Mislevy & Bock, 1989 for further discussion). The expected proportions correct can then be examined to determine whether departures from the common item response function are large or systematic. The same method can be used to check for differential item functioning across gender and racial/ethnic groups. Items that functioned differently across groups were reviewed to determine whether they should be deleted. In the case of items that function differently across age groups, another option is to estimate separate item response functions for each age level (e.g., see section 10.3).

7.2.3 Scaling

Unidimensional scales based on item response theory were derived for reading, writing, civics, mathematics, science, and geography. The NAEP methods use random draws ("plausible values") from estimated proficiency distributions to compute subpopulation statistics. Chapter 9 describes in detail the theoretical underpinnings of NAEP's scaling methods and the required estimation procedures. Only the basic analysis steps are outlined here.

For developing scales in the dichotomously scored subject areas (all areas except writing), the steps were as follows:

1) Use NAEP's version of the BILOG program[2] (Mislevy & Bock, 1982) to estimate the parameters of the item response functions on an arbitrary scale, assuming the three-parameter logistic model.

2) Use the M-GROUP program (Sheehan, 1985), which implements the method of Mislevy (see Chapter 9 or Mislevy, in press) to estimate proficiency distributions for each student on an arbitrary scale,

[2]NAEP BILOG allows students in each of the three age classes to be designated as distinct populations. This is important because, in NAEP, item sampling is not random across age classes. In this situation, age class membership must be taken into account to obtain consistent item parameter estimates via marginal maximum likelihood (see Mislevy & Sheehan, 1989).

based on these item parameter estimates and the student's responses to cognitive items and background questions.

3) Determine the appropriate metric for reporting the results and transform the results as needed.

4) Use random draws from these proficiency distributions ("plausible values" in NAEP terminology) for computing the statistics of interest, such as means for demographic groups.

In the case of the writing assessment, which yielded ordinal scores, another scaling model, the average response method (ARM) was applied. The basic steps to be applied were as follows:

1) Estimate the means on the writing exercises and intercorrelations among the exercises.

2) Use linear regression theory to impute a proficiency distribution for each student, based on these estimated means and correlations and on the student's responses to the writing exercises and background questions. Proficiency in this case is defined as the expected score on the entire set of writing exercises, given the responses to a subset of these exercises.

3) Use random draws from these proficiency distributions for computing the statistics of interest, such as means for demographic groups.

As explained in Chapter 9, the plausible values obtained through the IRT and ARM approaches are not optimal estimates of individual proficiency; instead, they serve as intermediate values to be used in estimating subpopulation characteristics. Under the assumptions of the scaling models, these subpopulation estimates will be consistent, which would not be true of subpopulation estimates obtained by aggregating optimal estimates of individual proficiency.

7.2.4 Scale Anchoring

Scale anchoring is a process that NAEP has used, beginning with the 1984 reading scale, to improve the utility of proficiency scale results by providing a criterion-referenced interpretation of selected scale levels. In this way, NAEP can further its goals of describing what students know and can do and stimulating debate about whether these levels of performance are satisfactory.

In NAEP's scale anchoring process, the first step is to choose four to five scale points to be anchored. For each point, items are then evaluated as potential anchor items, based on the percent of correct responses among students with proficiency levels at that point, as well as the corresponding percent for the next lower anchor point. For the anchoring of the 1988 U.S. history and civics scales, an item was considered to anchor at a particular point if (1) the percent of students with proficiency levels approximately equal to that point (i.e., within a 25-point interval centered at the point) who answered correctly was at least 65, (2) the percent of students with proficiency levels approximately equal to the *next lower anchor point* who answered correctly was less than 50, and (3) the difference between the percents in (1) and (2) was at least 30. (Of course, conditions (2) and (3) did not apply to the lowest anchor point.) After the items that anchored at each point were determined, subject area experts chose from among these the items that best characterized each point and developed descriptions of the anchored proficiency levels. The descriptions provide information about the types of skills that are possessed by a large proportion of students at that anchor point, but are not possessed by most students at lower levels. The percents of students at or above each anchor level are given in NAEP subject-area reports, along with the exemplar items and scale-point descriptions.

For the 1988 reading trend scale and the mathematics and science scales, which had already been anchored in the past, previously established anchor points were used. The process for developing these points was similar to that described above, but the anchoring criteria differed somewhat (see Beaton, 1987a and Johnson, 1988 for information specific to the mathematics scales and Yamamoto, 1988 for information specific to the science scales).

7.3 OVERVIEW OF CHAPTERS 8 THROUGH 15

The remaining chapters in Part II of this report are as follows:

Chapter 8: The 1988 National Assessment used a stratified multistage probability sampling design that provided for sampling certain subpopulations at higher rates (see Chapter 3). Because probabilities of selection are not the same for all assessed students, sampling weights must be used in the analysis of NAEP data. Also, in NAEP's complex sample, observations are not independent. As a result, conventional formulas for estimating the sampling variance of statistics are inappropriate. Chapter 8 describes the weighting procedures and methods for estimating sampling variance that are necessitated by NAEP's sample design. Further detail on sampling and weighting procedures is provided in *The 1988 National Assessment of Educational Progress—Sampling and Weighting Procedures, Final Report* (Rust, Bethel, Burke, & Hansen, 1990), a report prepared by Westat, Inc., the NAEP subcontractor in charge of sampling.

Chapter 9: A major NAEP innovation introduced by ETS is the reporting of subject-area results in terms of proficiency scales. Scaling methods can be used to summarize results even when students answer different subsets of items. For purposes of summarizing dichotomous item responses, NAEP developed scaling techniques that have their roots in item response theory and in the theories of imputation of missing data. For application to ordinal data, such as scores on the NAEP writing essays or responses to NAEP background items, NAEP developed the average response method (ARM), which is based on regression theory and imputation techniques. The ARM uses a multiple linear regression approach to estimate a student's score on a complete set of items, given responses to a subset of items. Chapter 9 describes these two scaling techniques, the underlying theory, and the application of these methods to 1988 NAEP data. Chapter 9 also includes a discussion of the advantages achieved as a result of the adoption in 1988 of the focused-BIB design. Administering three blocks of items in a single subject area produced more precise estimates of individual student proficiency than those that could be obtained in 1984 and 1986. As explained in the chapter, this greater precision reduces potential biases in the results of secondary analyses of the NAEP data. The final section of Chapter 9 gives an overview of the NAEP scales that were developed for the 1988 assessment.

Chapter 10: Two main components of the 1988 reading analysis are described in this chapter. First, the reading trend results for the years 1971 through 1984 were extended to include 1988 at ages 9, 13, and 17. The results of the reading trend analysis, which include the percents of students at or above the reading scale anchor points established in 1984, are reported in *The Reading Report Card, 1971-88: Trends from the Nation's Report Card* (Mullis & Jenkins, 1990). In addition, a detailed cross-sectional analysis of reading for grades 4, 8, and 12 in 1988 was conducted, including a study of the association between reading proficiency and student background variables. At grade 4, background information and data on instructional methods were collected from teachers and the relation of these variables to reading proficiency was examined. The cross-sectional analyses are reported in *Learning to Read in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Langer, Applebee, Mullis, & Foertsch, 1990).

Chapter 11: Like the reading analysis, the writing analysis consisted of two main components. The writing trend results, which provide a link to 1984 for grades 4, 8, and 11, are reported in *The Writing Report Card, 1984-88: Findings from the Nation's Report Card* (Applebee, Langer, Mullis, & Jenkins, 1990). A detailed cross-sectional analysis of writing for grades 4, 8, and 12 in 1988 was also conducted, including an examination of the association of writing skills with instructional techniques, student background variables, and the amount of time allocated for completion of the exercises. For grade 8, teacher data were collected and their association with writing proficiency was analyzed. The cross-sectional results are reported in *Learning to Write in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Applebee, Langer, Mullis, Jenkins, & Foertsch, 1990).

Chapter 12: The trend and cross-sectional analyses of the civics data are detailed in Chapter 12. The results of the trend analysis, which provided links to the 1975-76 and 1981-82 assessments for ages 13 and 17, are reported in *The 1988 Civics Report Card: Trends in Achievement from 1976 to 1988 at Ages 13 and 17 and Achievement in 1988 at Grades 4, 8, and 12* (Anderson, Jenkins, Leming, MacDonald, Mullis, Turner, & Wooster, 1990). A detailed cross-sectional analysis of civics for grades 4, 8, and 12 in 1988 was also conducted, including an examination of the association of civics knowledge with instructional techniques and student background variables. The cross-sectional results, which also include the percents of students at or above four anchor points that were determined in 1988, also appear in *The 1988 Civics Report Card*.

Chapter 13: Cross-sectional analyses for U.S. history, analogous to those described for the preceding subject areas, were conducted for grades 4, 8, and 12. The outcome of these analyses, which include scale anchoring results, are reported in *The U.S. History Report Card: The Achievement of Fourth-, Eighth-, and Twelfth-grade Students in 1988 and Trends from 1986 to 1988 in the Factual Knowledge of High-school Juniors* (Hammack, Hartoonian, Howe, Jenkins, Levstik, MacDonald, Mullis, & Owen, 1990). The 1988 U.S. history assessment also included a bridge sample linking the 1988 results to those of a special study of U.S. history conducted for grade 11 in 1986. For grade 11, trend results were obtained in terms of item percents correct, and were included in the 1990 report along with the cross-sectional results.

Chapter 14: NAEP assessed geography for the first time in 1988. Data were collected from students who were 17 years old or in grade 12. Results of the cross-sectional analyses for grade 12 which include the outcome of scale anchoring, are reported in *The Geography Learning of High-school Seniors* (Allen, Bettis, Kurfman, MacDonald, Mullis, & Salter, 1990).

Chapter 15: As noted earlier, mathematics and science items were included in the 1986 booklets that were administered in 1988 to allow an in-depth study of the 1986 reading anomaly. Therefore, a small-scale study of mathematics and science trend, including the derivation of scale anchoring results, was possible at ages 9, 13, and 17. Chapter 15 describes the methods used to link the 1988 results to those for 1978, 1982, and 1986. Because the mathematics and science analyses were conducted for the purpose of illuminating the reading anomaly, the analysis results, along with further detail on the analysis techniques, appear in Yamamoto's (1990) chapter in *The Effect of Changes in the National Assessment: Disentangling the NAEP 1985-86 Reading Anomaly* rather than in a subject-area report.

Chapter 8

WEIGHTING PROCEDURES AND ESTIMATION OF SAMPLING VARIANCE[1]

Eugene G. Johnson

Educational Testing Service

Keith F. Rust and Morris H. Hansen

Westat, Inc.

As was the case in previous assessments, the 1988 National Assessment used a complex sample design with the goal to obtain a sample from which estimates of population and subpopulation characteristics could be obtained with reasonably high precision (as measured by low sampling variability). At the same time, it was necessary that the sample be economically and operationally feasible to obtain. The resulting sample had certain properties that had to be taken into account in the proper analysis of the data from the assessment.

The 1988 NAEP sample was obtained through a stratified multistage probability sampling design that included provisions for sampling certain subpopulations at higher rates (see Chapter 3). To account for the differential probabilities of selection, and to allow for adjustments for nonresponse, each student was assigned a sampling weight. Section 8.1 discusses the procedures used to derive these sampling weights.

Another consequence of the NAEP sample design is its effect on the estimation of sampling variability. Because of the effects of cluster selection (students within schools, schools within primary sampling units) and because of the effects of certain adjustments to the sampling weights (nonresponse adjustment and poststratification), observations made on different students cannot be assumed to be independent of one another. As a result, ordinary formulas for the estimation of the variance of sample statistics, based on assumptions of independence, will tend to underestimate the true sampling variability. Section 8.2 discusses the jackknifing technique used by NAEP to estimate sampling variability. (The estimation of variability due to imperfect measurement of individual proficiency is discussed in Chapter 9.)

The jackknifing technique provides good quality estimates of sampling variability but requires considerable computations. Section 8.3 suggests the

---

[1]The statistical programming for this chapter was provided by David Freund, Bruce Kaplan and Lee Ann Held of Educational Testing Service, and Dalia Kahane of Westat, Inc.

use of design effects, combined with conventional variance estimation formulas, as a simple approximation to sampling variability. The technique is demonstrated for three types of statistics: simple item-level percent-correct statistics, average proficiency scores, and simple regression coefficients.

Yet another effect of the multistage NAEP sampling scheme is a reduction of the degrees of freedom of variance estimates, as compared with directly drawing independent samples of students (or of schools) without clustering them. The degrees of freedom of a variance estimator provide information about its stability: the higher the degrees of freedom, the lower the variability of the estimator. In a simple random sample, the degrees of freedom of a variance estimate depend upon the number of subjects and on the distribution of the variable under consideration. In the NAEP design, the degrees of freedom are primarily a function of the number of primary sampling units and the number of strata in the design, rather than the number of subjects, and the distribution of the variable under consideration has less impact. Section 8.4 discusses the degrees of freedom for NAEP jackknife variance estimates.

Since the sample design determines the derivation of the sampling weights and the estimation of sampling variability, it will be helpful to note the key features of the 1988 NAEP sample design. A description of the design appears in Chapter 3.

The 1988 sample was a multistage probability sample consisting of four stages of selection. The first stage of selection, the primary sampling units (PSUs), consisted of counties or groups of counties. The second stage of selection consisted of elementary and secondary schools. The assignment of sessions to sampled schools comprised the third stage of sampling, and the fourth stage involved the selection of students within schools and their assignment to sessions. The probabilities of selection of the first-stage sampling units were proportional to measures of their size, while the probability for subsequent stages of selection were such that the overall probabilities of selection of students were approximately uniform, with exceptions for certain subpopulations that were oversampled by design. For the main assessment, schools with relatively high concentrations of Black and/or Hispanic students were deliberately sampled at twice the normal rate to obtain larger samples of respondents from those subpopulations, in order to increase the precision in the estimation of the characteristics of these subpopulations. Students from schools with smaller numbers of eligibles received lower probabilities of selection, as a means of enhancing the cost efficiency of the sample.

The 1988 main assessment includes three student cohorts: students who were *either* in the fourth grade or 9 years old; students who were *either* in the eighth grade or 13 years old; and students who were *either* in the twelfth grade *or* 17 years old. The main assessment represents two overlapping samples. The first sample represents students of specified grades (who could be of any age). The second sample represents students of specified ages (who could be of any grade). Students were age-eligible if they were born in the appropriate calendar year (1978, 1974, or 1970). The main assessment of all grade/age levels was conducted in the winter and spring of 1988 and the sample design was such that the students assessed in the winter and the students assessed in the spring constitute two representative samples of the population.

The full 1988 assessment also includes a number of additional samples designed to determine the possible effects of changes in age definitions, time of testing, and mode of administration (elimination of the audiotape used for pacing the items), and to provide links to the results from previous assessments. Because the purpose of these studies was to provide the statistical linkage between the 1988 data and data from previous assessments, they are referred to as bridge (or trend) studies.

The full 1988 NAEP assessment thus includes a number of different samples from several populations. Each of these samples has its own set of weights that are to be used to produce estimates about the characteristics of the population addressed by the sample (the target population). The various samples and their target populations are as follows:

The Main Samples of Students. These samples, one for each of the three grade/age combinations, were drawn in the winter and spring, use the new age definitions, and consist of all students assessed in the main assessment. The target population for each of these samples consists of all students who are in the specified grade/age combination who were deemed assessable by their school.

Civics Bridge to 1976 and 1982. This bridge (trend) sample addresses the subject area of civics and consists of samples comparable to past assessments of citizenship and social studies and so uses pre-1986 definitions of age and time of testing. Since trend data have been traditionally collected only by age, grade sampling was unnecessary. The civics bridge sample consists of one booklet for age 13 and one booklet for age 17. Respondents to each booklet constitute a representative sample of the population of all students of that age. Because there were no reusable civics items from previous assessments of 9-year-olds, an age 9 sample was not needed.

Bridge to 1984. This bridge consists of trend samples comparable to the 1984 main assessment and addresses the subject areas of reading and writing. The samples were collected by grade and age for grade 4/age 9, grade 8/age 13, and grade 11/age 17, using the age definitions and time of testing from 1984. Six assessment booklets were administered at each grade/age. The respondents to the combined set of assigned booklets at a given grade/age constitute a representative sample of the population of students who are of the specified grade *or* of the specified age. The respondents to any one of the booklets also constitute a representative sample.

Bridge to 1986, Ages 9 and 13. This bridge consists of samples for ages 9 and 13 comparable to those used for the measurement of trends in 1986. The samples were collected by age only and used the same age definitions and time of testing as in 1984 and in the 1986 bridge to 1984. Three assessment booklets were administered to each age group and the respondents to any one of the three booklets assigned to a given age constitute a representative sample of the population of all students of that age.

Bridges to 1986, Grade 11/Age 17. These bridge (trend) samples, a U.S. history bridge and a reading, mathematics, and science bridge (see Chapter 1 for further details), are samples of grade 11/age 17 students comparable to the 1986 main assessment sample and were selected and administered using the 1986 age definition and time of testing. Since the age definition and time of testing also correspond to those used in samples from the 1984 and earlier assessments, the students in these bridge samples are comparable to the students from these earlier assessments. (However, the performance results are not directly comparable because the earlier assessments had paced audiotape administrations.) Seven assessment booklets were administered to grade 11/age 17 students. The administration of these booklets was nonpaced. The respondents to the combined set of seven booklets comprise a representative sample of the grade 11/age 17 population, as do the respondents to any one of the booklets.

For purposes of sampling and weighting, the assessment samples are categorized as "tape" or "spiral" according to whether or not paced audiotapes are to be used in the administration:

1) *Tape samples* are bridge samples that require audiotape pacing in the assessment (the civics bridge and the age 9 and 13 bridges to 1986). For these samples, all students within a particular assessment session receive the same booklet and are paced through at least part of the booklet with an audiotape. These assessment sessions are accordingly referred to as tape sessions. The students assigned to each distinct booklet of the tape samples are treated as a separate sample of the population for weighting.

2) *Spiral samples* are all main assessment samples and the remaining bridge samples. For these samples, no audiotape pacing was employed and the assessment booklets presented to a particular sample are spiraled through each assessment session (that is, the booklets are systematically interspersed and assigned for testing in that order). These assessment sessions are referred to as spiral sessions. The combined set of all students assigned to any of the booklets spiraled together is treated as a sample of the population for weighting.

8.1 DERIVATION OF THE SAMPLE WEIGHTS

As indicated previously, NAEP uses differential sampling rates, deliberately oversampling certain subpopulations to obtain larger samples of respondents from those subgroups, thereby enhancing the precision of estimates of characteristics of these oversampled subgroups. As a result of the oversampling, these subpopulations, corresponding to students from schools with high concentrations of Black and/or Hispanic students, are overrepresented in the sample. Lower sampling rates were introduced also for very small schools (those schools with only 1 to 19 eligible students). This was done in an approximately optimum manner as a means of reducing variances per unit of cost. Appropriate estimation of population ch racteristics must take disproportionate representation into account. This is accomplished by assigning a weight to each respondent, where the weights properly account for the sample design and reflect the appropriate proportional representation of the various types of individuals in the population.

The weighting procedures for 1988 included computing the student's base weight, the reciprocal of the probability that the student was invited to a particular session. These base weights were adjusted for nonresponse and then subjected to a trimming algorithm to reduce a few excessively large weights. The weights were further adjusted by a poststratification procedure in an effort to reduce the sampling error and certain potential biases of estimates relating to student populations corresponding to several subgroups of the total population. Poststratification was performed by adjusting the weights of the sampled students so that the resulting estimates of the total number of students in a number of specified subgroups of the population corresponded to population totals based on information from the Current Population Survey and Census Bureau estimates of the population. The subpopulations were defined in terms of race, ethnicity, geographic region, age, and grade.

The following sections provide an overview of the procedures used to derive the sampling weights. Further details in the derivation of these weights can be found in *1988 National Assessment of Educational Progress—Sampling and Weighting Procedures, Final Report*, (Rust, Bethel, Burke, & Hansen, 1990).

8.1.1 Student Base Weight

The base weight assigned to a student is the reciprocal of the probability that the student was invited to a particular type of assessment session, that is, a main assessment session or a particular bridge assessment session. That probability is the product of four factors:

1) the probability that the PSU was selected;

2) the conditional probability, given the PSU, that the school was selected;

3) the conditional probability, given the sample of schools in a PSU, that the school was allocated the specified type of session (this component is needed only for the bridge samples); and

4) the conditional probability, given the school, that the student was invited to the specified type of session.

Thus, the base weight for a student may be expressed as the product

$$W_B = PSUWT \cdot SCHWT \cdot SESSWT \cdot STUSCHW$$

where PSUWT, SCHWT, SESSWT, and STUSCHW are, respectively, the reciprocals of the preceding probabilities. The SESSWT term was included only for the bridge samples.

The season-specific base weight for a student is

$$W_{BS} = W_B \cdot SSUBWGT$$

where SSUBWGT is the reciprocal of the probability that the school attended by the student was selected for assessment in the particular (winter or spring) season. Each school had a probability of 0.5 of being allocated to winter or spring. In the case of those certainty PSUs that were paired, with one member assigned at random to each season (see Chapter 3), a ratio adjustment was made to weight the given pair member to the size of the pair, based on total population. The adjustment factor for students from the larger PSU, SSUBWGT, is somewhat less than 2.0, while for those from the smaller member, SSUBWGT is greater than 2.0.

The base weight for a student in a bridge sample is

$$W_{BB} = W_B \cdot WGTBRDG$$

where WGTBRDG is the reciprocal of the probability of selecting the student's PSU into the bridge sample, given that the PSU was selected for the main samples.

Table B-1 in Appendix B shows the distribution of base weights for each of the separate sessions conducted as part of the 1988 assessment. The variations in probabilities of selection, and consequently of weights, were introduced by design, either to increase the effectiveness of the sample in achieving its goals of reporting for various subpopulations, or to achieve increased efficiency per unit of cost.

8.1.2 Adjustment of Base Weights for Nonresponse

The base weight for a selected student was adjusted by four nonresponse factors. One of these was to adjust for noncooperating schools, while the second (used only in the case of bridge samples) was needed to adjust for allocated sessions that occasionally were not conducted. The third adjustment was needed to account for those few cases where, either inadvertently or on the insistence of the school, only students in the modal grade were given a chance of inclusion in the sample. The fourth adjustment was needed to adjust for students who were (or should have been) invited to the assessment but did not appear either in the scheduled session or a makeup session. Thus, the nonresponse adjusted weight for a student is of the form

$$W' = W_B \cdot SCHNRF \cdot SESNRF \cdot AOENRF \cdot STUNRF$$

where the nonresponse adjustment factors SCHNRF, SESNRF, AOENRF, and STUNRF are computed as described below.

The season-specific base weight was similarly adjusted for the same four types of nonresponse.

8.1.2.1 School Nonresponse Adjustment (SCHNRF)

The school nonresponse adjustment was intended to compensate for school nonresponse occurring before session assignment. These factors were computed separately within a PSU (except in a few cases where PSUs from similar strata were combined to give a more stable adjustment factor).

The school nonresponse adjustment factor in PSU h, $SCHNRF_h$, is given by

$$SCHNRF_h = \frac{\sum_{i \in A_h} SCHWT_{hi} \cdot G_{hi}}{\sum_{i \in B_h} SCHWT_{hi} \cdot G_{hi}}$$

where

$SCHWT_{hi}$	=	the school weight for school i in PSU h;
G_{hi}	=	the estimated number of grade/age-eligible students in school i in PSU h based on QED data (for sessions involving only age-eligible students, the number of age eligibles in each school was used);
set A_h		consists of the *original* sample of schools (cooperating and noncooperating schools, but not substitutes); and
set B_h		consists of all schools cooperating at the time of session allocation (including schools that were substituted for noncooperating schools).

For a substitute school, $SCHWT_{hi}$ is defined as the school weight of the originally selected school, while the value of G is taken from the substitute school itself. In those cases where PSUs were combined, the value of PSUWT (or PSUWT $\cdot$ WCTBRDG in the case of bridge samples) was included in the numerator and denominator of the school nonresponse adjustment factor.

Table B-2 in Appendix B shows the distribution of school nonresponse adjustment factors for each of the 1988 assessment sessions.

8.1.2.2 Session Nonresponse Adjustment (SESNRF)

The session nonresponse adjustment was intended to compensate for school nonresponse occurring in a few PSUs after session assignment in the bridge samples. These factors were computed separately within a PSU, except in cases where PSUs from similar strata were combined to give a more stable adjustment factor.

In PSU h, the session nonresponse adjustment factor $SESNRF_h$ was given by

$$SESNRF_h = \frac{\sum_{i \in B_h} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot G_{hi}}{\sum_{i \in C_h} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot G_{hi}}$$

where

$SCHWT_{hi}$ = the school weight for school i in PSU (or group of PSUs) h;

$SCHNRF_{hi}$ = the school nonresponse adjustment for school i in PSU h;

$SESSWT_{hi}$ = the session allocation weight for school i in PSU h;

G_{hi} = the estimated number of grade/age-eligible students in school i in PSU h in the case of spiral bridge sessions, and the estimated number of age-eligible students in the case of the tape sessions, to which only age eligibles were invited (the values of G_{hi} were based on QED data);

set B_h consists of all in-scope schools allocated to a particular type of session in PSU (or group of PSUs) h that were to be participating at the time of session allocation; and

set C_h consists of all schools allocated to the session type in PSU (or group of PSUs) h that ultimately participated.

In those cases where PSUs were combined, the value of PSUWT • WGTBRDG was included in both the numerator and denominator of SESNRF. Table B-3 in

Appendix B shows the distribution of the session nonresponse adjustment factor for each of the 1988 bridge sample sessions.

8.1.2.3 Age-only Eligibles Nonresponse Adjustment (AOENRF)

In a few schools in which assessments took place, only those students in the modal grade were listed for sampling (see Chapter 3), even though there was definite or very strong evidence that other eligible students were enrolled. Thus, an adjustment factor was needed to account for the fact that, although students eligible by age alone (age-only eligibles) were almost certainly enrolled in these schools, they were not given a chance of inclusion in the sample. These factors were calculated separately by PSU.

The school-level age-only eligibles nonresponse adjustment factor in PSU h, $AOENRF_h$, is given for students not in the modal grade by

$$AOENRF_h = \frac{\sum_{i \in C_h} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AO_{hi}}{\sum_{i \in D_h} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AO_{hi}}$$

where

$SCHWT_{hi}$ = the school weight for school i in PSU h;

$SCHNRF_{hi}$ = the school nonresponse adjustment for school i in PSU h;

$SESSWT_{hi}$ = the session allocation weight for school i in PSU h (bridge samples only);

$SESNRF_{hi}$ = the session nonresponse adjustment for school i in PSU h (bridge samples only);

AO_{hi} = the estimated number of age-only eligible students in PSU h, school i, based on PQ data;

set C_h consists of all schools allocated to the particular session type in PSU h that ultimately participated; and

set D_h consists of all schools allocated to the particular session type in PSU h, that could be reasonably supposed to have included age-only eligible students in the assessment, if any, or that had no age-only eligible students.

The value of $AOENRF_h$ for students in the modal grade is given as 1.0, since they were not subject to this component of nonresponse. Table B-4 in

Appendix B shows the distribution of the age-only eligible nonresponse adjustment factor for each of the 1988 assessment sessions.

8.1.2.4 Student Nonresponse Adjustment (STUNRF)

Student nonresponse adjustment factors were completed separately for spiral sessions and for each of the tape sessions within each PSU.

For spiral sessions, the student nonresponse adjustment was made separately for two classes of students in PSU h by age class: those in or above the modal grade for their age, and those below. This differentiation acknowledges likely differences between students in the two classes, both in their assessed abilities and in their likelihood of nonresponse. For some sessions in some PSUs, these two classes were combined, since one or both was too small to form the basis for an adjustment factor. For each class c in PSU h, the student nonresponse adjustment factor $STUNRF_{hc}$ is computed by

$$STUNRF_{hc} = \frac{\sum_{A'_{hc}} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AOENRF_{hi} \cdot STUSCHW_{hij}}{\sum_{B'_{hc}} SCHWT_{hi} \cdot SCHNRF_{hi} \cdot SESSWT_{hi} \cdot SESNRF_{hi} \cdot AOENRF_{hi} \cdot STUSCHW_{hij}}$$

where

$SCHWT_{hi}$	=	the school weight for school i in PSU h;
$SCHNRF_{hi}$	=	the school nonresponse adjustment factor for school i in PSU h;
$SESSWT_{hi}$	=	the session allocation weight for spiral sessions in school i in PSU h (bridge samples only);
$SESNRF_{hi}$	=	the session nonresponse adjustment factor for spiral sessions in school i in PSU h (bridge samples only);
$AOENRF_{hi}$	=	the age-only eligibles nonresponse adjustment factor in PSU h, school i;
$STUSCHW_{hij}$	=	the within-school student weight for student j in school i in PSU h;
Set A'_{hc}		consists of the students in class c in school i in PSU h who were invited to the session; and
Set B'_{hc}		consists of the students in class c in school i in PSU h who were assessed in the session.

The student nonresponse adjustment for tape sessions was similar, except that the adjustment was computed within a PSU for each tape booklet across all students originally invited to the assessment for that booklet.

Table B-5 in Appendix B shows the distribution of student nonresponse adjustment factors for each of the 1988 assessment sessions.

8.1.3 Trimming of Weights

In a number of cases, students were assigned relatively large weights. One cause of large weights was underestimation of the number of eligible students in some schools leading to inappropriately low probabilities of selection for those schools. A second major cause is the presence of large schools (high schools in particular) in PSUs with small selection probabilities. In such cases, the maximum permissible within-school sampling rate (determined by the maximum sample size allowed per school—see Chapter 3) could well be smaller than the desired overall within-PSU sampling rate for students. Large weights arose also because very small schools were, by design, sampled with low probabilities. Other large weights arose as the result of high levels of nonresponse coupled with low to moderate probabilities of selection, and the compounding of nonresponse adjustments at various levels.

Students with notable large weights have an unusually large impact on estimates such as weighted means. Since, under some simplifying assumptions the variability in weights contributes to the variance of an overall estimate by an approximate factor $1 + V^2$, where V^2 is the relative variance of the weights, an occasional unusually large weight is likely to produce large sampling variances of the statistics of interest, especially when the large weights are associated with students with atypical performance characteristics.

To reduce this problem, a procedure of trimming a few of the more extreme weights to values somewhat closer to the mean weight was applied. This trimming can increase the accuracy of the resulting survey estimates, substantially reducing V^2 and hence the sampling variance while introducing a small bias. The trimming algorithm was identical to that used in the 1984 and 1986 assessments and had the effect, approximately, of trimming the weight of any school that contributed more than a specified proportion, ζ, to the estimated variance of the estimated number of students eligible for assessment. The trimming was done separately for the spiral assessment and for each tape booklet in each of the bridge samples. In each case, the value of the proportion ζ was chosen to be 10/K, where K was the number of schools in which a specified assessment was conducted. The number of schools where weights were trimmed was small, being between 0 and 5 in each of the samples. Table B-6 in Appendix B shows the distribution of trimming factors for each of the 1988 assessment sessions. From the table it is seen that the most extreme trimming factors applied were of the order of 0.5 to 1. While we have not extensively examined the potential magnitude of bias that might be introduced from such trimming, based on the available evidence it seems reasonable to conclude that such bias would be quite small and that the reduction of variance would be large enough to result in a reduction in the mean square error.

8.1.4 Poststratification

As in most sample surveys, the respondent weights are random variables that are subject to sampling variability. Even if there were no nonresponse the respondent weights would at best provide unbiased estimates of the various subgroup proportions. However, since unbiasedness refers to average performance over a conceptually infinite number of replications of the sampling, it is unlikely that any given estimate, based on the achieved sample, will exactly equal the population value. Furthermore, the respondent weights have been adjusted for nonresponse and a few extreme weights have been reduced in size.

To reduce the mean squared error of estimates using the sampling weights, these weights were further adjusted so that estimated population totals for a number of specified subgroups of the population, based on the sum of weights of students of the specified type, were the same as presumably better estimates based on composites of estimates from the 1985 and 1986 Current Population Survey and 1988 population projections made by the Census Bureau. This adjustment, called poststratification, is intended especially to reduce the mean squared error of estimates relating to student populations that span several subgroups of the population, and thus to reduce the variance of measures of changes over time for such student populations.

8.1.4.1 1988 Poststratification Procedures

The poststratification in 1988 was done for each grade/age and separately for each of the spiral assessments and each of the tape assessments. Within each grade/age and assessment type group, poststratification adjustment cells were defined in terms of race, ethnicity, and NAEP region as shown in Table 8-1.

Table 8-1
Major Subgroups for Poststratification in 1988

Subgroup	Race	Ethnicity	Region*
1	White	Non-Hispanic	NE
2	White	Non-Hispanic	SE
3	White	Non-Hispanic	Central
4	White	Non-Hispanic	West
5	Any	Hispanic	Any
6	Black	Non-Hispanic	Any
7	Other	Non-Hispanic	Any

*Regions are the same as for stratification and reporting (see Chapter 3), except that all of Virginia is included in the southeast region for poststratification purposes

The result is seven poststratification cells for each tape session. For the assessments involving both age and grade eligible students, each of the seven subgroups was further divided into two or three eligibility classes. For age classes 9 and 13 and for the grade 11/age 17 bridge sample, three eligibility classes were used:

a) students eligible by both age and grade;
b) students eligible by age only;
c) students eligible by grade only.

For the grade 12/age 17 main assessment sample, the 7 subgroups were each divided into two subclasses:

a) students eligible by grade (of any age);
b) students eligible by age only.

This variation in the procedure from that used for the other age classes and for the grade 11/age 17 bridge was adopted because the independent estimates of the numbers of students in the population did not provide consistent data on the numbers of twelfth grade students eligible only by grade (see Rust et al., 1990, for further details).

Thus, there were 7, 14, or 21 cells for poststratification. The poststratified weight for each student within a particular cell was the student's base weight, with adjustments for nonresponse and trimming, times a poststratification factor. For each cell, the poststratification factor is a ratio whose denominator is the sum of the weights (after adjustments for nonresponse and trimming) of assessed and excluded students, and whose numerator is an adjusted estimate of the total number of students in the population who are members of the cell. This estimated total was a composite based on the October 1985 and 1986 Current Population Surveys and 1988 population projections. Table B-7 in Appendix B shows the distribution of poststratification factors for each of the 1988 assessments.

8.1.4.2 Differences From Earlier Procedures

The poststratification procedures used in 1988 were derived using an approach similar to those used in 1984 and 1986, but with major variations. To make the differences clear, the 1984 and the 1986 procedures will be described.

The same poststratification procedures were used for both the 1984 and 1986 assessments. For the spiral assessments, 13 subgroups were defined in terms of race, ethnicity, census region and community size (SDOC) as shown in Table 8-2. Each of the 13 subgroups was further divided into three classes:

a) students eligible by both age and grade;
b) students eligible by age only;
c) students eligible by grade only.

Table 8-2
Major Subgroups for Poststratification in 1986 and 1984

Subgroup	Race	Ethnicity	Region	SDOC*
1	White	NonHispanic	NE	1, 2
2	White	NonHispanic	NE	3, 4, 5
3	White	NonHispanic	SE, Central	1, 2
4	White	NonHispanic	SE, Central	3
5	White	NonHispanic	SE, Central	4, 5
6	White	NonHispanic	West	1, 2
7	White	NonHispanic	West	3, 4, 5
8	Any	Hispanic	NE, SE, Central	Any
9	Any	Hispanic	West	Any
10	Black	NonHispanic	NE	Any
11	Black	NonHispanic	SE	Any
12	Black	NonHispanic	Central, West	Any
13	Other	NonHispanic	Any	Any

*SDOC (Sample Description of Community) categories. 1—Big City, 2—Fringe of Big City, 3--Medium City; 4—Small Place; and 5—Extreme Rural.

The division of the sample by major subgroup and grade/age eligibility class resulted in 39 poststratification cells for each age class. The final weight for a student was the product of the base weight (as adjusted for nonresponse and after trimming) and a poststratif'· tion factor whose denominator was the sum of those weights for the c l to which the student belongs and whose numerator was an adjusted estimate of the total number of students in the cell. This adjusted estimate was a composite of estimates from the NAEP sample and an independent estimate based on projections based on Current Population Survey estimates and Census projections. The adjusted estimate was a weighted mean of the two estimates, the weights being inversely proportional to the approximate variances of the NAEP and the independent estimates.

The sample of students in each of the tape assessments was much smaller than the sample for the spiral assessments. Consequently, some subgroups in Table 8-2 were collapsed for poststratification as follows:

1, 2	6, 7
3	8, 9
4	10, 11, 12
5	13

Furthermore, to improve comparability with earlier assessments, there was no subdivision into age and grade eligibility classes, so that there were eight poststratification cells for each age class.

The 1988 poststratification procedures thus differ from those used for the 1984 and the 1986 assessments in three ways:

1) The 1988 poststrata totals incorporate current Census Bureau monthly population estimates by single years of age by race/ethnicity groups. Such monthly estimates were not available at the time of the poststratification of the 1984 and 1986 weights. Furthermore, the use of these estimates eliminated the need to derive year-to-year retention factors for age 17 students, as had been done in the previous years. This resulted from the fact that the estimates of in-school eligibles were obtained using data relating only to the particular grade and age in question, rather than incorporating projections from younger ages and lower grades, as was done in 1984 and 1986.

2) For the spiral assessments, the number of cells used in poststratification was reduced from the 39 cells used in 1986 and 1984 to the 14 or 21 cells used in 1988. For the tape assessments, the number of cells was reduced from eight to seven. The poststrata used for 1988 vary substantially in mean performance level and yet are large enough to produce reasonably stable poststratification factors. The reduction in the number of cells was made to increase the stability of the poststratification factors in an effort to reduce the sampling variance.

3) The 1988 poststrata totals were derived solely from Current Population Survey data and Census Bureau population projections and, in contrast to the method used in previous years, placed less reliance on the data from the 1988 NAEP samples. NAEP data were still used to determine the proportion of students eligible by both age and grade, for the spiral samples (other than for grade 12).

The 1988 procedure was adopted in order to speed up the production of the weights, since poststrata totals based only on Current Population Survey and Census data can be derived well in advance of the weighting of the data.

Appendix E describes the revisions made in the poststratification weights for the main samples for 1984 for grade 4/age 9 and for grade 8/age 13, also to improve the accuracy of estimates for 1984 and the trend measurements from 1984 to subsequent years.

8.1.4.3 A Measure of the Effect of Changes in Poststratification Procedures

It is clearly important to ascertain the impact of these changes in poststratification on the estimates of subgroup proficiencies. In particular, it is important to establish that the measurement of trend in subgroup proficiencies is affected in a minimal way by this revision in procedures. The approach used to ascertain the effect of the change in poststratification procedures was to reweight the 1986 samples according to the new procedures and then compare the results with the previous results. (This approach is considerably more cost and time efficient than the alternative approach of reweighting the 1978 data according to the 1986 procedures.)

Tables 8-3, 8-4, and 8-5 show the result when the age-eligible students in the trend samples of the 1986 assessment of reading are reweighted using the new poststratification factors. The first two columns in each table compare the new procedure with the old in terms of the estimated relative frequencies by race/ethnicity, region, parental education, and grade. The last two columns compare the two procedures in terms of the mean reading proficiencies for those subgroups.

An examination of these tables shows that the effect of changing the poststratification procedure on mean proficiency estimates is slight: in most cases, the difference between the proficiency estimates based on the two procedures is less than one standard error (of the mean proficiency based on the old method) and in every case the difference is less than 1.25 standard errors. The differences between estimates based on the two poststratification methods are well within the fluctuations to be expected by chance in either of the individual estimates.

We note that the standard errors of the difference between the original and revised estimates are likely to be relatively small, because of the high degree of correlation between the two sets of estimates. However, the important aspects of the change in the method are the sizes of the resulting differences in estimates, relative to the precision of the estimates themselves, as discussed above.

8.1.5 The Final Student-Weight: The Full-Sample Weight and the Season-Specific Final Weight

The final weight assigned to a student is the student full-sample weight. This weight is the student's base weight after the application of the various adjustments described above. The student full-sample weight was used to derive all estimates of population and subpopulation characteristics that have been presented in the various NAEP reports, including simple estimates such as the proportion of students of a specified type who would respond in a certain way to an item and more complex estimates such as mean proficiency levels.

In addition to the full-sample weight, a student season-specific weight was also derived. The season-specific weight is to be used for analyses that are based on data from the main NAEP assessments from either the winter administration or the spring administration alone. The season-specific weight is the student's season-specific base weight adjusted for nonresponse, subjected to the trimming algorithm, and then poststratified. The winter and spring season-specific weights were separately poststratified to the same poststrata totals as were used for the full-sample weights. The distributions of the season-specific poststratification factors are summarized in Table B-7 in Appendix B.

Table 8-3

Effect of Change in Poststratification Procedures on Relative Frequencies and Mean Reading Proficiencies, Age 9, 1986

	Relative Frequencies		Mean Reading Proficiencies	
	New Procedure	Old Procedure	New Procedure	Old Procedure
Observed Race/Ethnicity				
White	76.0%(1.0)	76.5%(1.1)	214.7(1.5)	214.9(1.3)
Black	15.5%(0.5)	14.9%(0.5)	186.4(1.6)	185.0(1.6)
Hispanic	6.0%(1.1)	6.2%(1.1)	189.0(2.9)	189.8(3.3)
Other	2.4%(0.5)	2.5%(0.5)	204.7(6.2)!	203.7(6.6)!
Region				
Northeast	20.7%(1.1)	21.1%(1.1)	212.0(3.0)	212.3(2.7)
Southeast	25.9%(2.0)	22.5%(4.7)	205.2(3.2)	202.5(2.7)!
Central	26.2%(0.9)	28.6%(4.0)	211.7(2.5)	212.9(2.7)
West	27.2%(1.6)	27.7%(1.6)	206.0(3.1)	206.5(3.0)
Grade				
< Modal Grade	34.2%(1.7)	33.9%(1.7)	188.3(1.2)	189.4(1.4)
at Modal Grade	65.5%(1.7)	65.8%(1.7)	218.9(1.3)	218.5(1.2)
> Modal Grade	0.3%(0.1)	0.3%(0.1)	238.2(8.8)!	241.9(11.3)!
Parental Education				
Not Graduated H S	4.3%(0.4)	4.2%(0.4)	190.1(2.9)	189.5(2.8)
Graduated H S	16.0%(0.8)	16.4%(0.7)	201.5(1.4)	202.2(1.9)
Post H S	44.7%(1.2)	44.4%(1.2)	219.2(1.4)	219.0(1.3)
Total			208.5(1.3)	208.6(1.2)

Note: Standard errors in parentheses

! Interpret with caution—the sampling error cannot be accurately estimated, since the coefficient of variation of the estimated total number of students in the subpopulation exceeds 20 percent.

Table 8-4

Effect of Change in Poststratification Procedures on Relative Frequencies and Mean Reading Proficiencies, Age 13, 1986

	Relative Frequencies		Mean Reading Proficiencies	
	New Procedure	Old Procedure	New Procedure	Old Procedure
Observed Race/Ethnicity				
White	77.3%(0.9)	76.8%(1.0)	260.3(0.9)	258.8(1.2)
Black	14.4%(0.8)	14.4%(0.9)	239.2(1.9)	239.3(1.6)
Hispanic	6.1%(1.0)	6.6%(1.1)	242.1(2.6)	242.2(3.1)
Other	2.2%(0.3)	2.2%(0.3)	262.3(3.6)	263.9(4.1)
Region				
Northeast	23.9%(1.6)	22.4%(1.6)	259.6(2.2)	258.7(2.1)
Southeast	23.9%(1.9)	24.7%(5.8)	254.3(1.6)	254.8(1.6)!
Central	25.6%(0.6)	24.9%(5.0)	254.6(1.3)	250.8(3.6)
West	26.7%(1.4)	28.0%(1.5)	256.1(1.8)	256.0(1.7)
Grade				
< Modal Grade	32.3%(1.6)	32.7%(2.1)	239.3(1.4)	238.4(1.4)
at Modal Grade	67.3%(1.6)	66.8%(2.1)	264.1(1.0)	263.0(0.9)
> Modal Grade	0.5%(0.1)	0.5%(0.1)	279.5(6.5)!	275.8(6.0)!
Parental Education				
Not Graduated H S	7.3%(0.5)	7.8%(1.0)	245.4(2.2)	244.2(2.9)
Graduated H S	29.6%(1.3)	30.5%(1.2)	249.8(1.2)	249.3(1.1)
Post H S	54.0%(2.0)	52.3%(2.1)	263.7(1.0)	262.7(0.9)
Total			256.2(0.8)	255.0(1.0)

Note: Standard errors in parentheses

! Interpret with caution—the sampling error cannot be accurately estimated, since the coefficient of variation of the estimated total number of students in the subpopulation exceeds 20 percent.

Table 8-5

Effect of Change in Poststratification Procedures on Relative Frequencies and Mean Reading Proficiencies, Age 17, 1986

	Relative Frequencies		Mean Reading Proficiencies	
	New Procedure	Old Procedure	New Procedure	Old Procedure
Observed Race/Ethnicity				
White	76.6%(0.4)	78.0%(0.4)	290.9(0.9)	291.4(0.9)
Black	14.6%(0.2)	13.5%(0.2)	264.9(1.3)	265.0(1.2)
Hispanic	6.4%(0.2)	6.2%(0.2)	266.3(2.4)	267.5(2.1)
Other	2.4%(0.3)	2.4%(0.3)	274.1(4.1)	276.0(4.4)
Region				
Northeast	25.4%(1.2)	23.8%(0.3)	291.2(2.0)	293.1(2.0)
Southeast	24.0%(0.6)	21.2%(1.4)	280.0(1.0)	279.4(1.0)
Central	26.1%(0.6)	28.4%(1.5)	287.1(2.1)	288.1(2.1)
West	24.5%(0.9)	26.5%(0.5)	281.7(1.4)	282.7(1.5)
Grade				
< Modal Grade	24.9%(0.6)	21.8%(0.6)	258.0(0.9)	257.7(1.0)
at Modal Grade	65.8%(0.4)	70.3%(0.4)	293.1(0.8)	293.1(0.8)
> Modal Grade	9.3%(0.6)	7.9%(0.5)	301.2(2.0)	301.0(2.1)
Parental Education				
Not Graduated H S	9.3%(0.5)	8.9%(0.6)	265.0(1.1)	266.3(1.4)
Graduated H S	27.8%(0.9)	27.7%(0.8)	274.9(0.8)	275.9(0.8)
Post H S	58.9%(1.3)	59.4%(1.2)	295.3(0.8)	295.8(0.9)
Total			285.1(0.8)	286.0(0.9)

Note: Standard errors in parentheses

The effects of all of the adjustments to the base weights are summarized in Table B-8 in Appendix B, which shows the distribution of the single factor given as the product of SCHNRF, SESNRF, AOENRF, STUNRF, the trimming factor, and the poststratification factor, for each of the assessment components. The distributions of the final student weights are given in Table B-9 in Appendix B.

As indicated earlier, under some simplifying assumptions the factor $1 + V^2$ indicates the approximate relative increase in variance of estimates resulting from the variability in the weights. The factor $1 + V^2$ for each sample is readily derivable from Table B-9 by adding 1 to the square of the ratio of the standard deviation to the mean weight. These factors, resulting from the combined effect of the variations in weights introduced by design and from other causes, as discussed above, are generally about 1.3 for the main samples. They vary from 1.12 to 1.53 for the various bridge samples.

8.1.6 Other Weights

In addition to the weights for the assessed students, weights were also derived for excluded students and for the students whose teachers participated in the teacher survey.

Weights for excluded students. Excluded students are students who were designated by the schools as unable to complete the assessment because they were non-English speaking, mildly mentally retarded (educable), or functionally disabled. Since the same grade and age eligibility definitions apply, no distinction is made between students excluded from the bridge assessments and students excluded from the main assessments for grade 4/age 9 and grade 8/age 13. However, since the age and grade eligibility definitions differ for the oldest age class. the excluded students from the grade 11/age 17 bridge assessments (with an October-September age definition and modal grade of 11) are treated as separate from the excluded students from the grade 12/age 17 main assessment (with a calendar-year age definition and modal grade of 12).

As in the case of the weights for the assessed students, the excluded student weights were constructed from components reflecting the probability of selection, correction for nonresponse, weight trimming, and poststratification. Further details on the derivation of the excluded student weights can be found in Rust et al. (1990). The distributions of the base weights, the various weight components and their composite, and the final student weight are given in Table B-10 in Appendix B for each of the four samples.

Weights for students in the teacher survey. The reading teachers of every grade 4 student who was assessed for reading in the main assessment and the writing teachers of every grade 8 student who was assessed for writing in the main assessment were identified within each school. Up to seven of these teachers within each school were selected to complete the teacher questionnaire; in schools with more than seven such teachers, a sample of five teachers of students assessed for reading or writing (as appropriate) in the main assessment was selected. Every selected teacher was provided a list of all his or her students (up to a maximum of 10) who had been assessed for reading or writing (as appropriate) in the main assessment. If more than 10 students fit the criteria, a random sample of 10 such students was provided. The selected teachers were asked to complete a detailed questionnaire about the capabilities of each selected student and the kinds of reading or writing instruction the student received.

These data can be analyzed using the teacher-student weights. It is important to note that the teacher-student weights are appropriate for use in estimating the number or percent of students in the total population who have various characteristics; they are *not* appropriate for use in estimating the number of teachers or the number of teachers with various characteristics. They are supplied only for a subsample of students and are appropriate for use in analyses involving the teacher-characteristics of students (for example, to estimate the proportion of fourth-grade students whose reading teachers have at least a master's degree).

The teacher-student weights are based on the final student weights of those students in the main assessment who are linked to a completed teacher questionnaire. The teacher-student weight is the nonresponse-adjusted student weight further adjusted for the probability that the student's teacher was selected as well as for nonresponse on the part of the teachers. Additionally, the teacher-student weights were subjected to the trimming algorithm and poststratification adjustments. Further details on the construction of these weights appear in Rust et al. (1990). The distributions of the teacher subsampling adjustment, the teacher nonresponse adjustment, the trimming factor, the poststratification factor, and their composite, together with that of the final teacher-student weights, are summarized in Table B-11 in Appendix B.

Finally, in addition to these weights, which were used to derive all estimates of population and subpopulation characteristics, other sets of weights, called jackknife replicate weights, were derived to facilitate the estimation of sampling variability by the jackknife variance estimation technique. These weights and the jackknife estimator are discussed in the next section.

8.2 PROCEDURES USED BY NAEP TO ESTIMATE SAMPLING VARIABILITY

A major source of uncertainty in the estimation of the value in the population of a variable of interest exists because information about the variable is obtained on only a sample from the population. To reflect this fact, it is important to attach to any statistic (e.g., a mean) an estimate of the sampling variability to be expected for that statistic. Estimates of sampling variability provide information about how much the value of a given statistic would be likely to change if the statistic had been based on another, equivalent, sample of individuals drawn in exactly the same manner as the achieved sample.

Another important source of variability is that due to imprecision in the measurement of individual proficiencies. In NAEP, proficiencies in subject areas are summarized through item response theory (IRT) or average response method (ARM) models, but not in the way that these models are used in standard applications where each person responds to enough items to allow for precise estimation of that person's proficiency. In NAEP, each individual responds to relatively few items so that individual proficiency values are not well determined. Consequently, the variance of any statistic based on proficiency values has a component due to the imprecision in the measurement of the proficiencies of the sampled individuals in addition to a component measuring sampling variability. The estimation of the component of variability due to measurement imprecision and its effect on the total variability of statistics based on proficiency values are discussed in Chapter 9.

The estimation of the sampling variability of any statistic must take into account the sample design. In particular, because of the effects of cluster selection (students within schools, schools within PSUs) and because of effects of nonresponse and poststratification adjustments, observations made on different students cannot be assumed to be independent of each other (and are, in fact, generally positively correlated). Furthermore, to account for the differential probabilities of selection (and the various adjustments), each student has an associated sampling weight, which should be used in the computation of any statistic and which is itself subject to sampling variability. Ignoring the special characteristics of the sample design and treating the data as if the observations were independent and identically distributed, will generally produce underestimates of the true sampling variability.

The proper estimation of the sampling variability of a statistic based on the NAEP data is complicated and requires techniques beyond those commonly available in standard statistical packages. Fortunately, the *jackknife* procedure (see, e.g., Wolter, 1985; Frankel, 1971) provides good quality estimates of the sampling variability of most statistics, at the expense of increased computation, and can be used in concert with standard statistical packages to obtain a proper estimate of sampling variability.

The jackknife procedure used by NAEP has a number of properties that make it particularly suited for the analysis of NAEP data. When properly applied, a jackknife estimate of the variability of a linear estimator (such as a total) will be the same as the standard textbook variance estimate specified for the sample design (if the first-stage units were sampled with replacement and approximately so otherwise). Additionally, if the finite sampling corrections for the first stage units can be ignored, the jackknife produces asymptotically consistent variance estimates for statistics such as ratios, regression estimates or weighted means and for any other nonlinear statistic that can be expressed as a smooth function of estimated totals of one or more variables (Krewski & Rao, 1981).

Through the creation of student replicate weights (defined below), the jackknife procedure allows the measurement of variability attributable to the use of poststratification and other weight adjustment factors that are dependent upon the observed sample data. Once these replicate weights are derived, it is a straightforward matter to obtain the jackknife variance estimate of any statistic.

The jackknife procedure in this application involves first defining pairs of first-stage sampling units or of appropriate aggregates of them in a manner that models the design as one in which two first-stage units are drawn with replacement per stratum in a manner reflective of the actual design. (The term first-stage sampling unit is used here in an effort to avoid the confusion that may arise because it has been convenient and traditional to retain the term PSU for large geographic sample units, many of which are included in the sample with certainty.) This requires different approaches for the certainty and for the noncertainty strata. For the main sample, the 60 noncertainty PSUs were formed into 30 pairs of PSUs, where the pairs were composed of PSUs from adjacent strata within each subuniverse (thus the strata were relatively similar on the characteristics of proportion minority population, population change between 1970 and 1980, and the proportions of urban and farm populations). Whereas the actual sample design was to select one PSU with probability proportional to size from each of 60 strata, for variance estimation purposes, the design is regarded as calling for the selection of two PSUs with probability proportional to size with replacement from each of 30 strata. This procedure likely gives a small positive bias to estimates of sampling error. The certainty PSUs are treated as strata for variance estimation purposes (which they in fact are), with the largest certainty PSU being treated as two strata (this is appropriate since stratified selection was used within PSUs), and all the other 33 PSUs being treated as a single stratum each. The schools in each of the 35 strata so defined were the first-stage sampling units, and were assigned within each strata to one of two half-samples, with equal probability, and systematically. For those certainty PSUs with both winter and spring assessed schools, each half-sample was composed of half winter schools and half spring schools, so as not to reflect inappropriately between season differences as a component of sampling error. For the noncertainty pairs, one PSU was assigned at random to each half-sample.

This procedure thus gives 65 half-sample pairs: 30 from noncertainty PSUs and 35 from certainty PSUs. About one-third of the U.S. population lives within the 35 certainty PSUs and thus the total number of half-sample pairs from this sector can be reduced somewhat, without appreciably reducing the effective degrees of freedom available for variance estimation (see section 8.4). Thus each of the 26 pairs formed by the 26 smallest certainty PSUs was combined with another pair from the same group. Thus these PSUs constitute 13 half-sample pairs. For each such pair, a given half-sample contains half of the schools (and hence their students) from each of two PSUs. The end result was 52 jackknife pairs of half-samples—two containing students from the largest certainty PSU, seven containing students each from a single large certainty PSU, 13 containing students each from two smaller certainty PSUs, with the schools from each PSU split between the half-samples, and 30 containing students each from two noncertainty PSUs, one PSU per half-sample

Similar pairings and half-samples were defined for the bridge assessments and for the season-specific administrations of the main assessment, but these correspond to somewhat different clusters of PSUs and, since fewer PSUs were involved, smaller numbers of pairs were defined in each case. Further information on the construction of jackknife pairs can be found in Rust et al.(1990).

NAEP's jackknife variance estimator is designed for the situation where the first-stage units, or appropriate aggregates of them, are paired within strata. It estimates the sampling variability of any statistic as the sum of components of variability that may be attributed to each of the jackknife pairs. The variance attributed to a particular jackknife pair is measured by estimating how much the value of the statistic would change if the information embodied in the jackknife pair were to be changed. This is done by the computation of a quantity t_i called a pseudoreplicate, which is associated with the i^{th} jackknife pair, and which is an estimate of the statistic of interest t based on an altered sample. Specifically, the i^{th} pseudoreplicate of the statistic t is created by randomly designating the half-sample members of the pair as first and second, eliminating the data from the first half-sample of the pair, replacing the lost information with that from the second half-sample of the pair (so that the second half-sample is included twice), repoststratifying the weights, and then reestimating the statistic for the pseudoreplicates based on this altered set of data.

The component of the sampling variability attributable to a jackknife pair is estimated as the squared difference between the value of the statistic for the complete sample and the pseudoreplicate associated with the pair. The estimated sample variance of the statistic t is the sum of M squared differences (where M is the number of jackknife pairs defined):

$$\hat{Var}(t) = \sum_{i=1}^{M} (t_i - t)^2 \quad .$$

The statistic for the pseudoreplicate associated with a given jackknife pair is the original statistic for the pseudoreplicate recomputed using an altered set of weights, referred to as the student replicate weights. The student replicate weight, SRWTi, for the i^{th} pair of first-stage units is computed as follows:

1) Let W_B be the nonresponse adjusted base weight of a student, where W_B accounts for the probabilities of selection and nonresponse but does not include poststratification adjustments.

2) Let W_{Bi} be the nonresponse adjusted replicate base weight formed by replacing the second member of the jackknife pair by the first, specifically:

$$W_{Bi} = \begin{cases} 0 & \text{if the student is in the first set of first-stage units in jackknife pair } i \\ JF * W_B & \text{if the student is in the second set of first-stage units in jackknife pair } i \\ W_B & \text{if the student is in neither of the first-stage units in jackknife pair } i \end{cases}$$

where JF is a constant multiplier (usually equal to 2) designed to maintain certain population totals.

3) Then the student replicate weight for the jackknife pair i is obtained by applying the poststratification adjustments to the weights W_{Bi} in the associated pseudoreplicate.

The poststratification adjustments are i computed for each jackknife replicate to reflect more completely the total effect of replacing one member of a jackknife pair with the other. (Nonresponse adjustments are not recomputed since these are generally performed within the PSU level and therefore their effect is appropriately reflected in the variance estimate.)

This estimation technique was used by NAEP to estimate all sampling errors presented in the various reports. A similar procedure was followed to estimate the sampling variability for statistics based on any of the bridge samples. The only difference was in the number of jackknife pairs (and hence replicate weights) used.

A further discussion of the variance estimation procedure used by NAEP including a discussion of alternative jackknife estimators that were also considered appears in Johnson (1989).

We noted above (as discussed in Chapter 9) that a separate estimate of the contribution to variance due to the imprecision in the measure of individual proficiencies is made and added to the jackknife estimate of variance. That variance component could have been appropriately reflected in the jackknife variance estimates simply by separately applying the IRT computations to each jackknife replicate. Because of the heavier IRT computational load, this was not done. Less work was involved by the simple procedure of making separate estimates of this component to be added to the jackknife variance estimates. Also, a separate measure of this component of variance is then available, which would not be so if it were reflected in the jackknife variance estimate.

8.3 APPROXIMATING THE SAMPLING VARIANCE USING DESIGN EFFECTS

In practical terms, the major expenditure of resources in the computation of a jackknife variance estimate occurs in the preparation of estimates for each of the pseudoreplicates. In the 1988 assessment, this implies that the statistic of interest has to be recomputed up to 53 times, once for the overall estimate t, and once for each of the up to 52 pseudoreplicates t_i. Because this is a considerable increase in the amount of computation required, relative to a conventional variance estimate, it is of interest to see how much the jackknife variance estimates differ from their less computationally intensive, simple random sampling based, analogues. For this purpose, we will compare the jackknife variance estimates with the conventional estimates for three commonly computed statistics: item proportion-correct statistics, mean proficiency estimates for subgroups, and simple regression coefficients.

The comparison of the conventional and the jackknife methods of variance estimation will be in terms of a statistic called the *design effect*, which was developed by Kish (1965) and extended by Kish and Frankel (1974). The design effect for a statistic is the ratio of the actual variance of the statistic (taking the sample design into account) over the conventional variance estimate based on a simple random sample with the same number of elements. The design effect is the inflation factor to be applied to the conventional variance estimate in order to adjust error estimates based on simple random sampling assumptions to account approximately for the effect of the sample design. The value of the design effect depends on the type of statistic computed and the variables considered in a particular analysis as well as the combined clustering, stratification, and weighting effects occurring among sampled elements. Generally, the design effects for statistics from complex samples such as NAEP are greater than one, because variances based on simple random sampling assumptions tend to provide underestimates.

8.3.1 Design Effects for Proportion-correct Statistics

As an example of the distribution of design effects to be expected from NAEP data, we first consider the design effect for the statistic P, the estimated proportion of a specified subgroup of the population who would correctly respond to a given assessment item. The proportion-correct statistic is the weighted mean of the responses to the item of the assessed individuals who belong to the subgroup, where an individual's response is either 1=correct or 0=incorrect. The design effect for the proportion-correct statistic P is of the form

$$deff(P) = [Var_{JK}(P)]/[P(1 - P)/N].$$

In the above, N is the total number of individuals in the subgroup responding to the item, $Var_{JK}(P)$ is the jackknife variance of P, and $P(1 - P)/N$ is the conventional variance estimate of P. (Although the estimate $P(1 - P)/N$ has the same form as the simple random sampling estimator of the variance of a proportion correct, the use of sample weights in the estimation of P reflects the appropriate distribution of the population.)

The distributions of estimated design effects across items for oportions correct by grade and by demographic subgroup within grade across all cognitive reading items presented in the 1988 main assessment of reading are indicated in Tables 8-6 through 8-8.

Table 8-6 addresses the distributions of the design effects for the 81 multiple-choice cognitive reading items presented in 1988 to grade 4 students. These distributions are shown for the population as a whole (Total) as well as for a variety of demographic subgroups: gender; race/ethnicity (White, Black, Hispanic, other); age (less than modal age, modal age, greater than modal age); region (Northeast, Southeast, Central, West); size and type of community (rural, low metropolitan, high metropolitan, big city, urban fringe, medium city, small place); parental education (at most high school, graduated high school, post-high school, graduated college, unknown); and type of school (public, nonpublic). For each of these groupings of fourth-grade students, Table 8-6 provides the lower quartile (LoQ), median, upper quartile (HiQ) and maximum design effect as well as the mean design effect for the 81 multiple-choice items presented to fourth-grade students.

Equivalent information on the distributions of design effects for the 99 multiple-choice cognitive reading items presented to grade 8 students appears as Table 8-7. The 108 multiple-choice cognitive reading items presented to grade 11 students are addressed by Table 8-8.

The particular demographic variables shown (gender, race/ethnicity, age, region, parental education, and size and type of community) were selected because (1) they are major variables in NAEP reports and (2) they reflect different types of divisions of the population that might have different levels of sampling variability.

The tables show that the design effects are predominantly larger than 1, indicating that standard variance estimation formulas will be generally too small, sometimes markedly so. Although the distributions of design effects appear somewhat different for certain subgroups of the population, they are, perhaps, similar enough (at least within a grade) to select an overall composite value that is adequate for most purposes. In choosing a composite design effect, some consideration must be made about the relative consequences of overestimating the variance as opposed to underestimating the variance For example, adopting the position that an overestimate of the variance is as severe an error as an underestimate leads to using a composite that is near to the center of the distributions of the design effects. Possible composites of this type are the mean and median design effects across the combined distribution of all design effects. In the current data, the mean design effects are 1.36, 1.31, and 1.36 respectively for grades 4, 8, and 12. These are close to, but greater than, the median design effects: 1.29, 1.26, 1.30.

Alternatively, one can adopt the position that it is a graver error to underestimate the variability of a statistic than to overestimate it. For example, Johnson and King (1987) examine estimation of variances using design effects (among other techniques) under the assumption that the consequences of an underestimate are three times as severe as those of an overestimate of the

Table 8-6

Distributions of Design Effects Across Items
by Demographic Subgroup
for Proportion-correct Statistics
for the Cognitive Reading Items Given in 1988

Grade 4*

Group	LoQ	Median	HiQ	Max	Mean
TOTAL	1.20	1.39	1.60	2.35	1.44
MALE	1.15	1.32	1.52	2.50	1.38
FEMALE	1.05	1.24	1.45	2.44	1.27
WHITE	1.05	1.26	1.54	2.08	1.29
BLACK	1.05	1.18	1.40	2.50	1.25
HISPANIC	1.11	1.37	1.59	2.32	1.38
OTHER	1.04	1.17	1.41	1.96	1.19
< MODAL	0.76	1.07	1.29	2.12	1.03
AT MODAL	1.00	1.31	1.48	1.96	1.27
> MODAL	1.17	1.35	1.61	2.27	1.41
NE	1.14	1.41	1.89	2.68	1.52
SE	1.13	1.52	1.94	3.48	1.63
CENTRAL	0.99	1.21	1.67	3.00	1.31
WEST	0.96	1.29	1.69	3.93	1.40
RURAL	0.69	0.99	1.30	3.08	1.09
LOW MET	1.13	1.52	1.88	4.06	1.59
HI MET	0.96	1.19	1.51	2.86	1.24
BIG CITY	1.51	1.96	2.43	3.43	2.03
FRINGE	0.99	1.24	1.82	3.30	1.41
MED CITY	1.36	1.74	2.09	3.22	1.76
SMALL PL	1.07	1.34	1.63	2.15	1.37
< HS	0.95	1.19	1.43	2.26	1.21
GRAD HS	1.01	1.20	1.36	1.70	1.18
HS +	0.99	1.14	1.43	1.84	1.19
GRADCOL	1.02	1.16	1.42	1.90	1.22
IDK	1.11	1.23	1.44	2.53	1.29
PUBLIC	1.24	1.40	1.68	2.46	1.47
NON-PUB	1.[illegible]	1.24	1.53	2.25	1.32

* Distributions are based on 81 multiple-choice items.

Table 8-7

Distributions of Design Effects Across Items
by Demographic Subgroup
for Proportion-correct Statistics
for the Cognitive Reading Items Given in 1988

Grade 8*

Group	LoQ	Median	HiQ	Max	Mean
TOTAL	1.17	1.36	1.57	2.24	1.38
MALE	1.11	1.30	1.57	2.55	1.36
FEMALE	1.06	1.23	1.36	1.94	1.24
WHITE	1.18	1.38	1.55	2.57	1.39
BLACK	0.98	1.25	1.50	2.46	1.28
HISPANIC	1.05	1.26	1.56	2.57	1.32
OTHER	0.97	1.17	1.35	2.50	1.20
< MODAL	0.69	1.03	1.29	2.38	1.01
AT MODAL	1.11	1.30	1.52	1.95	1.32
> MODAL	1.02	1.21	1.37	1.94	1.21
NE	0.96	1.22	1.52	2.89	1.30
SE	0.92	1.32	1.63	2.85	1.35
CENTRAL	1.16	1.61	2.07	3.61	1.67
WEST	0.93	1.25	1.69	3.70	1.40
RURAL	0.84	1.18	1.67	4.15	1.32
LOW MET	0.88	1.15	1.50	2.95	1.22
HI MET	0.98	1.38	1.72	2.98	1.41
BIG CITY	1.19	1.51	1.88	4.34	1.59
FRINGE	1.06	1.28	1.59	2.37	1.34
MED CITY	0.94	1.22	1.58	3.14	1.30
SMALL PL	1.03	1.32	1.65	2.66	1.34
< HS	1.10	1.26	1.42	2.19	1.26
GRAD HS	1.03	1.24	1.45	1.82	1.24
HS +	1.00	1.17	1.40	2.06	1.21
GRADCOL	1.01	1.20	1.47	2.30	1.25
IDK	1.07	1.21	1.39	1.99	1.23
PUBLIC	1.13	1.35	1.58	2.29	1.38
NON-PUB	0.95	1.24	1.48	2.32	1.24

* Distributions are based on 99 multiple-choice items.

Table 8-8

Distributions of Design Effects Across Items
by Demographic Subgroup
for Proportion-correct Statistics
for the Cognitive Reading Items Given in 1988

Grade 12*

Group	LoQ	Median	HiQ	Max	Mean
TOTAL	1.21	1.42	1.60	2.76	1.42
MALE	1.16	1.36	1.56	2.34	1.36
FEMALE	1.13	1.32	1.53	2.29	1.37
WHITE	1.11	1.35	1.49	2.31	1.32
BLACK	1.15	1.38	1.63	2.66	1.42
HISPANIC	1.18	1.39	1.62	3.72	1.44
OTHER	1.04	1.21	1.49	2.08	1.28
< MODAL	0.70	1.01	1.33	2.27	1.03
AT MODAL	1.19	1.41	1.66	2.67	1.41
> MODAL	1.04	1.23	1.40	2.26	1.25
NE	1.06	1.26	1.68	3.20	1.39
SE	0.87	1.10	1.57	3.29	1.28
CENTRAL	0.79	1.13	1.49	4.07	1.23
WEST	1.06	1.31	1.72	3.61	1.42
RURAL	0.70	1.05	1.57	3.51	1.18
LOW MET	1.02	1.38	1.69	5.65	1.45
HI MET	1.00	1.39	1.85	3.82	1.49
BIG CITY	1.06	1.31	1.75	3.73	1.49
FRINGE	0.73	1.07	1.43	2.26	1.13
MED CITY	0.94	1.24	1.90	3.60	1.41
SMALL PL	1.08	1.42	1.81	2.96	1.44
< HS	1.11	1.30	1.57	2.73	1.35
GRAD HS	1.10	1.26	1.45	2.17	1.30
HS +	1.06	1.23	1.39	2.17	1.25
GRADCOL	1.12	1.32	1.55	2.12	1.35
IDK	1.13	1.41	1.80	4.83	1.57
PUBLIC	1.21	1.41	1.58	2.33	1.41
NON-PUB	1.09	1.45	1.84	2.75	1.50

* Distributions are based on 108 multiple-choice items.

same magnitude. Adopting a loss function that is a weighted sum of absolute values of the deviations of predicted from actual, with underestimates receiving three times the weight of overestimates, produces the upper quartile of the design effects as the composite value. This assumes that the distribution of design effects is roughly independent of the jackknife estimates of variance, so that the size of a design effect does not depend on the size of the variance. The values of this composite, for grades 4, 8, and 12, respectively, are 1.59, 1.54, and 1.59.

8.3.2 Design Effects for Subgroup Mean Proficiency Scores

Since most of the analyses conducted by NAEP are based on the results of scaling models that summarize performance of students across a learning area, we next consider the design effects to be expected for analyses based on these scale scores. For reasons given in Chapter 9, NAEP provides each individual with a set of "plausible values," each of which is a random draw from the distribution of the potential scale scores for that individual. Since our current interest is on the effect of the sampling design on estimation and inference, we will restrict our attention to a single measure of an individual's proficiency, the first plausible value of the individual's scale score.

A key statistic of interest is the estimated mean proficiency of a subgroup of the population. An estimate of the subgroup mean proficiency is the weighted mean of the first plausible values of proficiency of the sampled individuals who belong to the subpopulation of interest. Let $\bar{Y}$ be the weighted mean of the plausible values of the sampled members of the subpopulation. The conventional estimate of the variance of $\bar{Y}$ is

$$\mathrm{Var}_{con}(\bar{Y}) = [\sum_{i=1}^{N} w_i \cdot (y_i - \bar{Y})^2] / [N \cdot W_+]$$

where N is the total number of sampled individuals in the subpopulation for which proficiency scores are available, w_i is the weight of the i^{th} individual, y_i is a plausible value from the distribution of potential scale scores for that individual, and W_+ is the sum of the weights across the N individuals.

The design effect for the subgroup mean proficiency estimate is

$$\mathrm{deff}(\bar{Y}) = \mathrm{Var}_{JK}(\bar{Y}) / \mathrm{Var}_{con}(\bar{Y})$$

where $\mathrm{Var}_{JK}(\bar{Y})$ is the jackknife variance of $\bar{Y}$. (As has been pointed out previously, $\mathrm{Var}_{JK}(\bar{Y})$ as computed does not measure the variability of $\bar{Y}$ due to imprecision in the measurement of the proficiencies of the sampled individuals. The estimation of this very important source of variability is discussed in Chapter 9.)

Values of the average design effects for subgroup mean proficiencies are displayed, by grade, in Table 8-9. The subpopulations considered in this

Table 8-9

Average Design Effects
by Demographic Subgroup and Grade
for Mean Reading Proficiency Scores*

Group	Grade 4	Grade 8	Grade 12
TOTAL	3.08	3.10	2.45
MALE	2.46	2.43	1.79
FEMALE	2.02	2.30	2.51
WHITE	2.88	3.72	2.51
BLACK	1.57	2.32	1.44
HISPANIC	1.41	1.39	1.43
OTHER	1.23	1.14	1.03
< MODAL	0.92	0.79	0.94
AT MODAL	2.23	2.31	2.84
> MODAL	2.43	1.88	1.32
NE	4.40	2.15	4.09
SE	3.83	2.48	1.08
CENTRAL	3.51	8.35	2.15
WEST	3.15	1.85	1.96
RURAL	2.14	3.34	1.83
LOW MET	2.42	2.43	2.57
HI MET	3.05	3.10	4.11
BIG CITY	6.32	4.20	2.35
FRINGE	3.58	3.11	1.85
MED CITY	5.29	2.19	3.24
SMALL PL	3.90	3.69	3.04
< HS	1.09	1.25	1.31
GRAD HS	1.51	2.28	0.84
HS +	0.93	1.20	1.09
GRADCOL	1.59	1.78	2.10
IDK	1.61	0.96	2.68
PUBLIC	3.41	3.08	2.23
NON-PUB	3.61	2.36	4.16

* Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of reading proficiency.

table are identical with those considered for the design effects of proportion-correct statistics. It is interesting to note that the design effects for subgroup mean proficiency estimates are noticeably larger than the design effects for the proportion-correct statistics. This suggests a larger effect of clustering for mean proficiency scores (which are averages of continuous variables) than for proportion-correct statistics (which are averages of binary variables). Perhaps more important, this increase in design effect may arise because of the use of BIB spiraling. Thus, for a percent correct for an item, the number of students taking the item within a first-stage sampling unit is less than half the number that contribute to the average proficiency score for a subject like mathematics or science. About 7/3 as many students within a first-stage sampling unit are assessed on a subject as on an item, and contribute to an average proficiency score for the subject. Consequently, the effective cluster size (and design effect) for an average proficiency is likely to be larger—although not as large as it likely would be if all students assessed in a subject were given all items for that subject.

As was the case for the design effects of the proportion-correct statistics, three candidates for an overall composite design effect for subgroup mean proficiencies are the mean, the median, and the upper quartile of the distribution of design effects. The values of these potential composite values, by grade, are as follows:

	Grade 4	Grade 8	Grade 12
Mean	2.70	2.54	2.18
Median	2.45	2.32	2.13
Upper Quartile	3.56	3.10	2.65

Each of these values is roughly twice the equivalent value for the design effects of proportion-correct statistics.

We note that the $Var_{(con)}\overline{Y}$ as defined above is an estimate of S^2/N where S^2 represents the unit variance for a simple random sample for the population of students from which the sample is also drawn. This is an appropriate estimate of the increase in variance over simple random sampling from that population. However, the computer packages used for estimating the variance may not reflect the weights in estimating the unit variance, as given above, but instead may provide an estimate of a unit variance of the form

$$\sum^{N} (y_i - \overline{y})^2/N^2.$$

In this case, the unweighted estimate of unit variance would be appropriate for the denominator of a design effect measure of the increase in variance over the unit variance as estimates by the computer package. If there is no correlation between the w_i and y_i, there would be little difference between the two.

8.3.3 Design Effects for Simple Regression Coefficients

Table 8-10 shows the design effects for simple regression coefficients from the weighted regression of the first plausible value of reading proficiency on each of the individual conditioning variables. (The codings of the conditioning variables appear in Appendix ¯.) The conventional standard errors were obtained from a standard weighted regression in which the weights were scaled to add to the total sample size, thereby more closely representing the true amount of information available in the construction of the standard errors relative to the standard errors that would have been produced if the weights were left in their original metric. We see from Table 8-10 that this scaling of the weights does not go nearly far enough in terms of appropriately representing the true sampling variability of the regression estimates.

The mean, median, and upper quartile of the distribution of the design effects for regression coefficients are, by grade:

	Grade 4	Grade 8	Grade 12
Mean	2.40	2.02	1.91
Median	1.78	1.64	1.59
Upper Quartile	3.31	2.56	2.14

The mean, median, and upper quartiles of the distribution of design effects for regression coefficients are distinctly larger than those of the proportion-correct statistics, but are somewhat smaller than those of the subgroup mean proficiencies. This results accords with the conjecture made by Kish and Frankel (1974) that the design effects for complex statistics (such as regression coefficients) tend to be smaller than the corresponding design effects for means of the same variable.

8.4 THE DEGREES OF FREEDOM OF THE VARIANCE ESTIMATE

It is important to have an indication of the number of degrees of freedom to attribute to the jackknife variance estimator $\hat{Var}(t)$. The degrees of freedom of a variance estimator provide information on the stability of that estimator: the higher the number of degrees of freedom, the lower the variability of the estimator. In practical terms, the number of degrees of freedom of the variance estimator corresponds to the number of residual degrees of freedom that can be assumed for inferential procedures.

Since the jackknife procedure estimates the sampling variability of the statistic by assessing the effect of change in the sample at the paired first-stage sampling unit (FSSU) level, the number of degrees of freedom of the variance estimator $\hat{Var}(t)$ will be at most equal to M, the number of FSSU pairs. The maximum number of degrees of freedom equals the number of independent pieces of information used to gererate the variance. In the case of data from the main assessments, the pieces of information are the 52 squared differences $(t_i - t)^2$, each supplying at most one degree of freedom (regardless of how many individuals were sampled within any FSSU).

Table. 8-10

Average Design Effects by Grade
for Simple Regression Coefficients Based on Reading Proficiency Scores*

Coefficient	Grade 4	Grade 8	Grade 12
FEMALE	1.37	1.87	1.86
BLACK	1.32	1.84	1.30
HISPANIC	1.24	1.71	1.48
ASIAN	2.00	0.51	1.14
HI METRO	3.29	3.08	4.19
OTHER STOC	6.02	2.83	3.70
SOUTHEAST	3.94	2.96	1.49
CENTRAL	3.30	7.81	2.13
WEST	3.58	2.59	2.12
H.S.	1.07	1.43	1.20
>H.S.	0.71	0.93	0.95
GRAD COLLEGE	1.46	1.51	1.93
PARED MISSING	1.32	1.49	3.57
3-4 ITEMS	0.75	1.26	1.26
>4 ITEMS	1.21	1.5[illegible]	1.96
TV	1.70	1.50	1.02
TV SQUARED	1.65	1.41	1.00
LANG MINORITY	1.31	1.89	1.92
NO HW GIVEN	2.22	1.69	2.50
HOMEWORK	2.24	2.47	2.50
TIME HW	2.05	1.59	1.79
% LUNCH	6.41	3.05	2.14
NO LUNCH	6.98	3.42	3.61
MINORITY SCH	3.54	3.33	2.69
INTEGRATED SCH	4.72	3.14	3.76
MODAL AGE	1.65	1.02	1.74
> MODAL AGE	1.88	1.05	1.72
NONPUBLIC	3.33	1.96	4.15
HOME HELP	1.69	1.24	1.47
MULTIPAR	0.84	1.05	1.07
MOM HOME	1.21	2.09	1.34
MOM WORK	2.15	1.57	1.36
SOM READ	1.43	1.93	1.26
LOT READ	1.66	1.19	1.61
LOW ABSENT	—	1.45	1.59
GRADES	—	1.45	1.17
COL PREP	—	—	1.51
VOC-TECH	—	—	1.13
2YR-COL	—	—	1.40
4YR-COL	—	—	1.13
HRS WKED	—	—	1.22
ADV ENGLISH	—	—	1.28
REM ENGLISH	—	—	2.58

---

* Regressions are based on the first plausible values of reading proficiency.

The number of degrees of freedom of the sample variance estimator can be strictly less than the number of FSSU pairs. For example, suppose that the statistic t is a mean for some subgroup and no members of that subgroup can come from either FSSU in the i^{th} FSSU pair. (Examples of such a subgroup are any PSU-level partitioning of the population, such as region.) In this instance, neither member of the FSSU pair i directly contributes to the estimate of t, so that the pseudoreplicate t_i would nearly equal the statistic t. If the replicate weights used to generate t_i had not received poststratification adjustments, the resulting pseudoreplicate t_i would be identical to the overall estimate t so that $(t_i - t)^2 = 0$. In this case, such a FSSU pair would impart no information to the variability of the statistic t and thus contribute zero degrees of freedom to the variance. However, since the replicate weights have received poststratification adjustments, the component $(t_i - t)^2$ is measuring the effect of the poststratification on the estimate. While being nonzero, such a component will tend to be much smaller in magnitude than the squared difference $(t_k - t)^2$ for any PSU pair k that does contribute to the estimate of t (see Rust, 1985).

The squared difference $(t_i - t)^2$ estimates σ_i^2, say, the contribution to the sampling variance of the statistic t which can be attributed to the i^{th} FSSU pair and $\hat{Var}(t)$ estimates the sum of the contributions across all pairs.

$$\sum_{i=1}^{M} \sigma_i^2 .$$

If the σ_i^2 vary widely, as when a few of the σ_i^2 are markedly larger than the remainder, as in the above case, then $\hat{Var}(t)$ is predominantly estimating the sum of these larger components, which dominate the remaining terms. The effective degrees of freedom of $\hat{Var}(t)$ in this case will be nearer to the number of dominant terms.

One way to estimate how many degrees of freedom to attribute to the jackknife variance estimate of a statistic t is to match estimates of the first two moments of $\hat{Var}(t)$ to those of a chi-square random variable (Satterthwaite, 1941). If the t_i are normally distributed, the effective number of degrees of freedom using this approximation is

$$df_{eff} = \frac{\left[\sum_{i=1}^{M} (t_i - t)^2 \right]^2}{\sum_{i=1}^{M} (t_i - t)^4} .$$

However, this approximation will overestimate the effective degrees of freedom if the distribution of the differences $t_i - t$ has positive kurtosis (Cochran, 1977, p. 96). A more direct way of assessing the effective degrees of freedom of a variance estimate is possible when a number of independent replicates of the estimate are available. In such a situation, the ordered values of the replicate estimates of $\hat{Var}(t)$ are compared with the expected values of the order statistics of various chi-square distributions. The aim is to find the chi-square distribution that most closely matches the empirical distribution. If the distribution of $\hat{Var}(t)$ is approximately chi-square, a quantile-quantile plot (Chambers, Cleveland, Kleiner, & Tukey, 1983, section 6.2) of the ordered replicate estimates versus the expected values of the order statistics from the best fitting chi-square distribution will be close to a straight line through the origin. The degrees of freedom for that best fitting chi-square distribution are taken to be the effective degrees of freedom of $\hat{Var}(t)$.

This procedure will be followed to estimate the effective degrees of freedom for the variance estimates of the subpopulation mean proficiencies and of the regression coefficients. A slightly different procedure will be used for the effective degrees of freedom of the variances of weighted proportion-correct statistics.

It is possible to estimate the number of degrees of freedom to attribute to the jackknife variance estimates of the weighted proportion-correct statistics by considering the distribution of design effects for a given set of items in a population or subpopulations (such as males or total) under the assumptions that the individual design effects are all estimating the same, underlying, design effect D and that the variance estimates of all weighted proportion-correct statistics have the same degrees of freedom, f. Specifically, assume that the jackknife variance estimate, V_j, of the j^{th} weighted proportion-correct statistic, P_j, is distributed like the random variable $(\sigma_j^2 / f)\, X_f^2$, where X_f^2 is a chi-square random variable with f degrees of freedom and σ_j^2 is the expected value of V_j. Further assume that the expected value of the conventional variance estimate is σ_j^2/D, where D is the underlying design effect. Then, for a sufficiently large sample size, so that the conventional variance estimate can be taken to be σ_j^2/D, the design effect of P_j will be approximately distributed like the constant (D/f) times a chi-square random variable with f degrees of freedom. If the underlying design effect D and the degrees of freedom f are the same for all P_j, then the distribution of the estimated design effects of the proportions correct across the set of items will be approximately distributed like a multiple times a chi-square random variable with f effective degrees of freedom.

Under these assumptions, the value of the effective degrees of freedom for the design effects of a given subpopulation (such as males in the fourth grade) is found by comparing the ordered values of the design effects for all of the proportion-correct statistics for that group with the expected values of the order statistics of a sample of the same size of chi-square random variables with f degrees of freedom. The comparison is repeated for all integral degrees of freedom f between 1 and M (where M=52 for the main assessment). The value of f that minimizes the residual mean-square of the least-squares line through the origin predicting th ordered design effects from the chi-square order statistics is the effective degrees of freedom for the set of design effects.

Table 8-11 shows the result of this estimation of the effective degrees of freedom of the design effects, and hence the jackknife variance estimates, for the weighted proportion-correct statistics. The quality of the

Table 8-11

Effective Degrees of Freedom for the Design Effects of the Proportion-correct Statistics

Group	Grade 4	Grade 8	Grade 12
TOTAL	41	37	40
MALE	36	32	35
FEMALE	27	47	40
WHITE	31	36	41
BLACK	25	23	30
HISPANIC	27	23	30
OTHER	38	26	33
< MODAL	16	13	13
AT MODAL	36	46	41
> MODAL	38	48	40
NE	17	15	11
SE	12	11	9
CENTRAL	11	12	10
WEST	13	11	13
RURAL	10	10	7
LOW MET	12	15	13
HI MET	15	14	13
BIG CITY	19	13	13
FRINGE	13	25	11
MED CITY	15	11	9
SMALL PL	21	18	18
< HS	22	35	32
GRAD HS	39	38	36
HS +	34	35	43
GRADCOL	42	30	44
IDK	34	43	14
PUBLIC	38	33	42
NON-PUB	21	24	14

approximation to the distribution of the design effects is quite good: In every case, the residual mean squared error from the prediction of the ordered values of the design effects from the order statistics from a chi-square distribution with the effective degrees of freedom was no more than 10 percent of the variance of the actual values and was within 5 percent in all but two of the 84 cases.

The numbers in Table 8-11 show that the effective degrees of freedom of the jackknife variance estimates are indeed no larger than the number of FSSU pairs, and are, in fact, markedly smaller in some cases. There appear to be two major groupings:

1) Variance estimates for weighted proportion statistics corresponding to subpopulations that can appear in any PSU in the sample. These subpopulations include: total, the two genders, two of the four race/ethnicity groups (White and Other), students at or above modal age, the five levels of parental education (except for grade 12), and public schools. The variance estimates for these subpopulations have the highest number of effective degrees of freedom. The estimated range is between 26 and 48, and the average is about 38. (Some exceptions are Black and Hispanic students, age less than modal age for grade, less than high-school education, nonpublic school students, and parent's education unknown. Note that while these classes can appear in any PSU, they tend to vary widely among PSUs; or may also be based on a very small sample of students. The estimated effective degrees of freedom for these range from 13 to 38, and average about 23.)

2) Variance estimates corresponding to subpopulations that can only appear in certain parts of the sample; such as the 4 regions and the 7 size and type of community designations. These tend to have the smallest effective degrees of freedom. They range between 7 and 15, and average about 13.

It is also interesting to determine how many degrees of freedom to attribute to the variance estimates for the subpopulation mean proficiencies and for the regression coefficients. Tables 8-12 and 8-13 provide this information. In constructing these tables, a replication procedure was used rather than that used for Table 8-11 since, instead of having a large number of design effects (one for each item) from which to build a distribution, we have only a single jackknife variance for each of the subpopulation means and regression coefficients. In order to produce a distribution of variances for each of the coefficients, the full sample of respondents was split into seven interpenetrating subsamples based on the assessment booklet taken. (Recall that the main assessment of reading consisted of seven booklets at each grade/age that were spiraled throughout the sample.) Thus, all students receiving the first reading assessment booklet were placed in the first subsample; all students receiving the second booklet were placed in the second subsample; and so on. The result is seven samples, each of which include all of the PSUs in the full sample while including roughly 1/7 of the students within each of the PSUs. Because of the reduced student sample size, the jackknife variance estimate based on one of these subsamples will be somewhat

Table 8-12
Effective Degrees of Freedom for the Design Effects
for Mean Reading Proficiency Scores*

Group	Grade 4	Grade 8	Grade 12
TOTAL	46	41	49
MALE	35	26	52
FEMALE	21	52	47
WHITE	22	45	52
BLACK	22	32	26
HISPANIC	52	14	26
OTHER	16	31	23
< MODAL	16	7	12
AT MODAL	23	49	52
> MODAL	29	26	39
NE	18	15	22
SE	16	27	10
CENTRAL	10	11	11
WEST	10	7	9
RURAL	52	9	6
LOW MET	12	12	21
HI MET	23	21	8
BIG CITY	39	29	11
FRINGE	15	31	15
MED CITY	27	16	20
SMALL PL	14	14	52
< HS	52	52	22
GRAD HS	48	34	52
HS +	35	28	33
GRADCOL	49	52	52
IDK	27	37	5
PUBLIC	22	50	35
NON-PUB	51	52	46

---

* Design effects are based on the first plausible values of reading proficiency.

Table 8-13

Effective Degrees of Freedom for the Design Effects
for Simple Regression Coefficients Based on Reading Proficiency Scores*

Coefficient	Grade 4	Grade 8	Grade 12
FEMALE	33	52	32
BLACK	28	52	31
HISPANIC	52	18	15
ASIAN	7	20	6
HI METRO	23	20	10
OTHER STOC	52	50	16
SOUTHEAST	16	33	30
CENTRAL	12	13	17
WEST	13	18	22
H.S.	34	35	52
>H.S.	20	47	35
GRAD COLLEGE	52	52	18
PARED MISSING	38	15	3
3-4 ITEMS	52	48	37
>4 ITEMS	44	52	40
TV	13	40	23
TV SQUARED	21	52	12
LANG MINORITY	28	52	52
NO HW GIVEN	52	11	11
HOMEWORK	52	13	13
TIME HW	40	28	37
% LUNCH	13	21	52
NO LUNCH	38	37	20
MINORITY SCH	13	28	9
INTEGRATED SCH	14	52	39
MODAL AGE	17	30	52
> MODAL AGE	20	25	40
NONPUBLIC	24	30	52
HOME HELP	52	52	26
MULTIPAR	29	31	43
MOM HOME	25	21	15
MOM WORK	52	29	36
SOM READ	52	52	41
LOT READ	42	52	39
LOW ABSENT	—	24	52
GRADES	—	52	26
COL PREP	—	—	24
VOC-TECH	—	—	22
2YR-COL	—	—	17
4YR-COL	—	—	42
HRS WKED	—	—	14
ADV ENGLISH	—	—	18
REM ENGLISH	—	—	3

* Regressions are based on the first plausible values of reading proficiency.

larger than the est'mator from the full sample. Because the effects of clustering have been reduced, the design effects for these replicate variance estimates will be reduced. However, since the same number of jackknife pairs a involved, the replicate variance estimates may reasonably have approximately the same number of degrees of freedom as did the estimate from the full sample.

For each of the seven subsamples, jackknife variances, conventional variances, and design effects were estimated for each of the subpopulation means and regression coefficients. For each subpopulation mean and regression coefficient, the best fitting chi-square distribution to the distribution of the seven design effect estimates was found, in the same manner analogous as that used for the design effects of the proportion-correct statistics. The resulting effective degrees of freedom appear in Table 8-12 for the subpopulation means and in Table 8-13 for the regression coefficients.

We see again that the effective degrees of freedom of the variance estimates tends to be smaller than the number of FSSU pairs, and that this is true for each of the subpopulation means and regression coefficients. Unfortunately, the most striking characteristic of these estimates of effective degrees of freedom is their variability: the effective degrees of freedom of a subpopulation for one grade can differ considerably from the equivalent estimate for a different grade. (An extreme example is the rural subpopulation mean, which has effective deg... of freedom, for grades 4, 8 and 12, of 52, 9, and 6.) This variability may be reflecting the instability of the degrees of freedom estimates since each estimate is based on only 7 values. The degrees of freedom estimates for the proportion-correct statistics are much more stable since these are based on at least 81 values.

The effective degrees of freedom for the NAEP jackknife variance estimates are markedly smaller than the degrees of freedom attributed to the corresponding error estimates from conventional techniques. This fact affects inferential procedures since significance tests based on the conventional degrees of freedom will be too liberal (and confidence intervals will be too small). Fortunately, for the usual significance levels, the effect of using the effective degrees of freedom rather than the conventional values is generally moderate: a t statistic significant at the α=5% level assuming infinite degrees of freedom (essentially the conventional estimate) is significant at the α=6% level for 20 effective degrees of freedom, the α=7% level for 10 effective degrees of freedom, and the α=10% level for five effective degrees of freedom.

For practical purposes, the impact of the reduced degrees of freedom on inferential techniques can be largely accounted for by (1) using a moderate number (say 25) of degrees of freedom for all inferences of items that can appear in all PSUs, and (2) using a smaller number (say 10) for the remaining items. Certainly one should be cautious about barely significant results for subgroups that are highly clustered in the population.

Chapter 9

SCALING PROCEDURES[1]

Robert J. Mislevy

Educational Testing Service

A key innovation in NAEP during the ETS tenure is scale-score reporting. With scaling methods, the performance of a sample of students in a subject area or subarea can be summarized on a single scale even when different students have been administered different exercises. Similar procedures can be used to summarize responses to sets of related background questions. This chapter presents an overview of the scaling methodologies employed in the analyses of the 1988 NAEP surveys:

- Section 9.1 briefly discusses the perspective on scaling from which the procedures were conceived and applied.
- Section 9.2 reviews the "plausible values" methodology used in NAEP scale-score analyses.
- Section 9.3 describes how plausible values are used in subsequent analyses.
- Section 9.4 lists the scale-score analyses carried out on the 1988 data.

Details of scaling procedures for specific subject areas are presented in Chapters 10 through 15.

9.1 SCALING IN NAEP

NAEP reports were originally envisaged some 20 years ago as simple lists of percents correct to individual survey items, in the population as a whole and in subpopulations of particular interest. It soon became apparent, however, that major features of the detailed results from hundreds of items could not be effectively communicated without some kind of summarization The Education Commission of the States, the NAEP contractor at that time, introduced averages of percents correct for sets of items, which summarized patterns of performance in groups of related exercises and made it possible to compare more easily the general performances of subpopulations—as long as

[1]The contributions of Albert Beaton, Eugene Johnson, David Freund, Bruce Kaplan, Jennifer Nelson, Kathleen Sheehan, Minhwei Wang, and Rebecca Zwick to this chapter are gratefully acknowledged.

those subpopulations had been presented a common subset of items. Because this approach limits comparisons to percents correct on specific sets of items, it provides no simple way to report trends over time as the item pool evolves. Moreover, it gives no information about the distributions of skills among students in targeted subpopulations.

These limitations can be overcome by the use of response scaling methods. If several items require similar skills, the regularities observed in response patterns can often be exploited to characterize both respondents and items in terms of a relatively small number of variables. When combined through appropriate mathematical formulas, these variables capture the dominant features of the data. Using the scale, it becomes possible to talk about distributions of proficiency in a population or subpopulation, and to estimate the relationships between proficiency and background variables.

Early work on scaling is attributed to Thurstone, but the more recent development of item response theory (IRT) has been particularly influential on measurement practice (Hambleton, 1989). IRT and a newly developed procedure called the average response method (ARM), both of which are reviewed in section 9.2, are the two scaling procedures ETS has introduced in NAEP reporting.

We hasten to point out that any procedure of aggregation, from a simple average to a complex multidimensional scaling model, highlights certain patterns at the expense of other potentially interesting patterns that may reside within the data. In a very real sense, every single item in a NAEP survey is of interest in its own right, and can provide useful information about what young Americans know and can do. The choice of an aggregation procedure must be driven by a conception of just which patterns are salient for a particular purpose. The procedure that is optimal for one purpose may be poorly suited for another.

The relatively high levels of aggregation found in ETS/NAEP reports such as *The Reading Report Card: Progress Toward Excellence in Our Schools* (NAEP, 1985), for example, are well suited to high-level discussions of trends and policy implications. They average over, and therefore are *not* keyed to, the microanalysis of performance at the level of specific skills, as might be desired by educational psychologists; they do not reveal popular student misconceptions or erroneous rules, as might be of interest to classroom teachers in a subject area. For studying specific skills, one might prefer the precision of a latent class model for more highly specified skills. For studying misconceptions, detailed discussions of results for individual items might be more appropriate. By no means do the scale-score methods we employ as a reporting vehicle exhaust the potential of NAEP data; neither do they prevent other researchers from carrying out alternative analyses from different perspectives. Indeed, NAEP public-use data tapes, which contain the original responses of all surveyed students, were created expressly to encourage and facilitate such analyses.

A reporting scale in the 1988 NAEP survey, then, simply summarizes performance on a collection of educational tasks in much the same way that the Consumer Price Index (CPI) summarizes the total cost of a market basket of products. The two indices exhibit some of the same useful features and limitations. Just as the CPI composite represents average American spending patterns, the items in a NAEP survey were specified by independent consensual process to tap a "market basket of skills." Just as changes in the CPI reflect at a glance changes in the cost of goods in general, changes in NAEP scale-score distributions reflect changes in proficiency as averaged over the items in the pool. But understanding just how and why the CPI changes requires deeper analyses, into specific components of the market basket, when the CPI goes up, some of the components will have gone up by greater rates than others, while some may have even dropped in price. The NAEP scale depends similarly on the balance of items of varying types and topics in the survey, and reflects only an average over the varying patterns among them. NAEP first attempts to carry out scaling in subject areas in which similar patterns can be expected over items; then, within each scaling area, highlights meaningful departures from general trends in several ways, including the following:

1) Explicitly discussing countertrends or comparisons that can be identified with one or a few items. This is analogous to reporting that the Consumer Price Index jumped 5 percent, but noting that the increase was mainly due to a change in OPEC oil prices.

2) Supplementing scale-score distributional results with more detailed breakdowns in terms of percents correct for groups of related items.

3) Carrying out scaling in separate subareas within the subject area, in which it is anticipated that trends or comparisons may differ because of different curricular emphases over time or across schools. As was done in the 1986 NAEP surveys of mathematics and science, these subscale results were supplemented by a subject area average. This is comparable to calculating price changes in separate market baskets for food, transportation, energy, and so on, and reporting these individually along with the overall average.

9.2 NAEP SCALING METHODOLOGY

The paragraphs that follow review the scaling models employed in the analyses of 1988 NAEP data, and the "plausible values" methodology that allows such models to be used with NAEP's sparse item-sampling design. The reader is referred to Mislevy (1988a, in press) for an introduction to plausible values methods and a comparison with standard psychometric analyses, to Mislevy and Sheehan (1987) and Beaton and Johnson (1987, 1990) for additional information on how the models are used in NAEP, and to Rubin (1987) for the theoretical underpinnings of the approach.

9.2.1 The Scaling Models

Two types of scaling models were used by NAEP in the 1988 assessment. The three-parameter logistic (3PL) model from item response theory (IRT; e.g., Lord, 1980) was used for the subject areas of reading, civics, U.S. history, geography, mathematics, and science. The average response method (ARM, Beaton & Johnson, 1987, 1990), developed by NAEP for the 1984 assessment, was used for the subject area of writing. Both are "latent variable" models, quantifying respondents' tendencies to provide responses in a given direction (e.g., correct answers to items in a subject area; positive responses on attitude questions; higher rather than lower ratings in written essays), as a function of a parameter that is not directly observed.

The three-parameter logistic (3PL) IRT model. The fundamental equation of the 3PL model is the probability that a person whose proficiency is characterized by the *unobservable* variable θ will respond correctly to item j.

$$P(x_j=1|\theta,a_j,b_j,c_j) = c_j + (1-c_j)/\{1+\exp[-1.7a_j(\theta-b_j)]\}$$
$$= P_j(\theta), \qquad (9.1)$$

where

- x_j is the response to item j, 1 if correct and 0 if not;
- a_j, where $a_j>0$, is the slope parameter of item j, characterizing its sensitivity to proficiency;
- b_j is the threshold parameter of item j, characterizing its difficulty; and
- c_j, where $0\le c_j<1$, is the lower asymptote parameter of item j, reflecting the chances of a correct response from students of very low proficiency. In 1988 NAEP analyses, c parameters were estimated for multiple-choice items, but were fixed at zero for open-ended items.

For the purposes of reporting item parameter estimates and other intermediary estimates, the linear indeterminacy apparent in (9.1) may be resolved by an arbitrary choice of the origin and unit size in a given scale. This was done for the reading scale in 1984 by standardizing the combined grade 4/age 9, grade 8/age 13, and grade 11/age 17 samples. To aid interpretation, final published results are reported on scales that are transformed linearly from the θ scale in ways related to the 0-to-500 reading proficiency scale developed in the 1984 NAEP assessment of reading (Beaton, 1987a). These transformations are described in the corresponding subject area chapters in this report.

Under the usual IRT assumption of local independence, the probability of a vector $x=(x_1,\ldots,x_n)$ of responses to n items is simply the product of terms based on (9.1):

$$P(x|\theta,\underline{a},\underline{b},\underline{c}) = \prod_{j}^{n} [P_j(\theta)]^{x_j} [1-P_j(\theta)]^{1-x_j} . \qquad (9.2)$$

It is typically also assumed that response probabilities are conditionally independent of background variables, say y, given θ, or

$$P(x|\theta,\underline{a},\underline{b},\underline{c},y) = P(x|\theta,\underline{a},\underline{b},\underline{c}) .$$

(Checks on the degree to which these assumptions are met in NAEP data, and ways that meaningful departures are handled, are described below.)

After x has been observed, (9.2) can be viewed as a likelihood function, and provides a basis for inference about θ or about item parameters. In NAEP, estimates of item parameters were obtained with a modified version of Mislevy and Bock's (1982) BILOG computer program, then treated as known in subsequent calculations. Once items have been calibrated in this manner, a likelihood function for θ is induced by a vector of responses to any subset of calibrated items, thus allowing θ-based inference from matrix samples.

Conditional independence is a mathematical assumption, not a necessary fact of nature. Even though the IRT models are employed in NAEP only to summarize average performance, a number of checks are made to detect serious violations of conditional dependence, and, when warranted, remedial efforts are made to mitigate its effects on inferences. These include the following:

1) Checks on relative item operating characteristics among distinct gender and ethnicity groups (i.e., differential item functioning, or DIF, analyses). Some degree of relative differences are to be expected, of course, and modestly varying profiles among groups will exist beyond the differences conveyed by their differing θ distributions. The intent at this stage is to detect and eliminate items that operate differentially for identifiable reasons that are unrelated to the skills intended to be measured in the subject area.

2) When a scale extends over age groups, evidence is sought of different operating characteristics over ages. When such effects are found, an item in question is represented by different item parameters in different age groups.

3) When a scale extends over time, evidence is similarly sought as to whether an item's relative operating characteristics have changed over time, over and above differences that can be accounted for by

changes in the overall θ distribution. Studies of NAEP reading data (Beaton & Zwick, 1990) suggest these effects are small in adjacent assessments when assessment forms are held constant, but they too can be taken into account by allowing item parameters to vary over time (see Bock, Muraki, & Pfeiffenberger, 1988). The variation in item parameters is relative to other items in the scale; general trends in item parameters are equivalent to changes in θ distribution.

Item-level factor analyses have diminished in importance as our perspective of the role of IRT in NAEP has evolved. The assumption that performance in a scaling area is driven by a single unidimensional variable is unarguably incorrect in detail. But our use of the model is not theoretical, but data analytic; interpretation of results is not trait-referenced, but domain-referenced. Scaling areas are determined *a priori* by considerations of content and politics, as collections of items for which overall performance is deemed to be of interest. The IRT summary is not expected to capture all meaningful variation in item response data, but to reflect distributions of overall proficiency—to summarize the main patterns in item percents-correct in the populations and subpopulations of interest. Using a unidimensional IRT model when the true model is multidimensional captures these overall patterns even though it over- or under-estimates the covariances among responses to items in pairs. For inferences based on overall proficiency, violations of the model with respect to dimensionality are less serious than violations in the shapes of the marginal response curves—hence our greater attention to routine checks of item-fit residuals for every item in every calibration run than to factor analytic results.

Once it is accepted that the model cannot be strictly correct, attention focuses on violations that distort the most important inferences that are to be drawn. Estimated proficiency distributions and item parameters, for example, correctly capture groups' overall performances, but may over- or under-predict certain items. The first type of information is very important in NAEP; the second is less important. Note however that this imparts an importance to the mix of items that would not be present if the IRT model were true, since the balance of items presented determines the nature of average proficiency. Item mixes are identical for subpopulations within a given assessment, so overall proficiency automatically has the same meaning for them. Item mixes are *not* necessarily the same for different grade/age groups or different time points, so the more careful checks listed above as 2) and 3) are required to maintain (as well as possible) meaning across these linkages.

In all NAEP IRT analyses, missing responses at the end of each block a student was presented were considered "not-reached," and treated as if they had not been presented to the respondent. Missing responses before the last observed response in a block were considered intentional omissions, and treated as fractionally correct at the value of the reciprocal of the number of response alternatives. These conventions are discussed by Mislevy and Wu (1988). With regard to the handling of not-reached items, they find that ignoring not-reached items introduces slight biases into item parameter estimation to the degree that (1) not-reached items are present and (2) speed is correlated with ability. With regard to omissions, they find that the method described above provides consistent limited-information likelihood estimates of item and ability parameters under the assumption that respondents omit only if they can do no better than responding randomly.

The local independence assumption embodied in (9.2) implies that item response probabilities depend only on θ and the specified item parameters—not on the position of the item in the booklet, on the content of items around an item of interest, or on test-administration timing conditions. These effects are certainly present in any application, however. The practical question is whether the IRT probabilities obtained via (9.2) are "close enough" to be robust with respect to (1) the context in which the data are to be collected and (2) the inferences that are to be drawn. For example, experience with adaptive testing has shown using the same item parameters regardless of when an item is administered does not materially bias estimates of the proficiencies of individual examinees. Our experience with the 1986 NAEP reading anomaly, has shown, however, that for measuring small changes over time, changes in item context and speededness conditions lead to unacceptably large random error components. These can be avoided by presenting items used to measure change in identical test forms, with identical timings and administration conditions. Thus we do *not* maintain that the item parameter estimates obtained in any particular booklet configuration are appropriate for other conceivable configurations, and the parameter estimates are context-bound. (For this reason, we prefer common population equating to common item equating whenever equivalent random samples are available for linking.) A given assessment block, it will be recalled, can appear as the first, second, or third block of a booklet. The appearances are balanced, so that any differences in item parameters due to block position are averaged out.

The average response method (ARM) model. The basic equation of the ARM is an average of item responses:

$$\theta = a'x \ . \tag{9.3}$$

Here a is a vector of constants, specified so as to provide a meaningful summary of performance. Weights of 1/n for an n-item test, for example, yield simply an average score; weights given by the k^{th} eigenvector of the covariance matrix for x yield the k^{th} component score. If a respondent responded to all items, then an ARM score would be directly calculable via (9.3) without error. Typically, however, a given NAEP respondent receives only a subset of the items in an ARM scale, so that his or her ARM θ is not observed directly.

9.2.2 An Overview of Plausible Values Methodology

Item response theory was developed in context of measuring individual examinees' abilities. In that setting, each individual is administered enough items (often 100 or more) to permit precise estimation of his or her θ, as a maximum likelihood estimate $\hat{\theta}$, for example. Because the uncertainty associated with each θ is negligible, the distribution of θ, or the joint distribution of θ with other variables, can then be approximated using individuals' $\hat{\theta}$ values as if they were θ values.

This approach breaks down in the assessment setting when, in order to provide broader content coverage in limited testing time, each respondent is administered relatively few items in a scaling area. The problem is that the uncertainty associated with individual θs is too large to ignore, and the features of the $\hat{\theta}$ distribution can be seriously biased as estimates of the θ distribution. (The failure of this approach was verified in early analyses of the 1984 NAEP reading survey; see Wingersky, Kaplan, & Beaton, 1987.) "Plausible values" were developed as a way to estimate key population features consistently, and approximate others no worse than standard IRT procedures would. A detailed development of plausible values methodology is given in Mislevy (1988a, in press). Along with theoretical justifications, that paper presents comparisons with standard procedures, discussions of biases that arise in some secondary analyses, and numerical examples. The following paragraphs give a brief overview of the plausible values approach, focusing on its implementation in the 1988 NAEP analyses.

Let $\underline{y}$ represent the responses of all sampled examinees to background and attitude questions, along with design variables such as school membership. If IRT or ARM θ values were available for all sampled examinees, it would be possible to compute a statistic $t(\underline{\theta},\underline{y})$—such as a subpopulation sample mean, a sample percentile point, or a sample regression coefficient—to estimate a corresponding population quantity T. A function $U(\underline{\theta},\underline{y})$—e.g., a jackknife estimate—would be used to gauge sampling uncertainty, as the variance of t around T in repeated samples from the population.

Because the 3PL model and the ARM are latent variable models, however, θ values are not observed even for sampled students. To overcome this problem, we follow Rubin (1987) by thinking of $\underline{\theta}$ as "missing data" and approximate $t(\underline{\theta},\underline{y})$ by its expectation given $(\underline{x},\underline{y})$, the data that actually were observed, as follows:

$$
\begin{aligned}
t^*(\underline{x},\underline{y}) &= E[t(\underline{\theta},\underline{y})|\underline{x},\underline{y}] \\
&= \int t(\underline{\theta},\underline{y})\, p(\underline{\theta}|\underline{x},\underline{y})\, d\underline{\theta} \quad . \qquad (9.4)
\end{aligned}
$$

It is possible to approximate t^* using random draws from the conditional distributions $p(\theta|x_i,y_i)$ of each sampled student i. These values are referred to as "imputations" in the sampling literature, and "plausible values" in NAEP. The value of θ for any respondent that would enter into the computation of t is thus replaced by a randomly selected value from the conditional distribution for θ given his or her responses to cognitive items (x_i) and background items (y_i). Rubin (1987) proposes that this process be carried out several times—"multiple imputations"—so that the uncertainty associated with imputation can be quantified. The average of the results of, say, K estimates of t, each computed from a different set of plausible values, is a Monte Ca approximation of (9.4); the variance among them, B, reflects uncertainty due to not observing θ, and must be added to the estimated expectation of $U(\underline{\theta},\underline{y})$, which reflects uncertainty due to testing only a sample of students from the population. Section 9.3 explains how plausible values are used in subsequent analyses.

It cannot be emphasized too strongly that plausible values are *not* test scores for *individuals* in the usual sense.

Plausible values are offered only as intermediary computations for calculating integrals of the form of equation 9.4, in order to estimate *population* characteristics. When the underlying model is correctly specified, plausible values will provide consistent estimates of population characteristics, even though they are not generally unbiased estimates of the proficiencies of the individuals with whom they are associated. The key idea lies in a contrast between plausible values and the more familiar θ estimates of educational measurement that are in some sense optimal for each examinee (e.g., maximum likelihood estimates, which are consistent estimates of an examinee's θ, and Bayes estimates, which provide minimum mean-squared errors with respect to a reference population): *Point estimates that are optimal for individual examinees have distributions that can produce decidedly nonoptimal (specifically, inconsistent) estimates of population characteristics* (Little & Rubin, 1983). Plausible values, on the other hand, are constructed explicitly to provide consistent estimates of population effects.

In both IRT and ARM analyses in NAEP, plausible values are included for the small numbers of students who responded to background questions, but, although they were presented items in a subject area, did not respond to any of them. The conditional distribution employed for such a nonrespondent is based solely on his or her background values y. This special class of nonrespondents was included in this manner, even though they provided no information about their proficiencies, in order to maintain the representativeness of the sample. This convention provides estimates of population characteristics that have the *same expected value and precision* as would be obtained under the more familiar nonresponse adjustment of deleting the nonrespondents and boosting the sampling weights of respondents with the same y values, since their plausible values are drawn from the estimated θ distributions of the appropriately matched respondents.

9.2.3 Computing Plausible Values in IRT-based Scales

Plausible values for each respondent i are drawn from the conditional distribution $p(\theta|x_i,y_i)$. This subsection describes how, in IRT-based scales, these conditional distributions are characterized, and how the draws are taken. Using first Bayes' theorem, then the IRT assumption of conditional independence,

$$p(\theta|x_i,y_i) \propto P(x_i|\theta,y_i)\; p(\theta|y_i)$$

$$= P(x_i|\theta)\; p(\theta|y_i), \qquad (9.5)$$

where $P(x_i|\theta)$ is the likelihood function for θ induced by observing x_i (treating item parameter estimates as known true values) and $p(\theta|y_i)$ is the distribution of θ given the observed value y_i of background responses.

Equations (9.4) and (9.5) can also be employed with vector-valued θ, as in the 1986 NAEP mathematics subscales. In such cases, $P(x_i|\theta)$ is the product over subscales of the *independent likelihoods* induced by responses to items within each subscale, and $p(\theta|y_i)$ is the multivariate—and generally non-independent—*joint density* of proficiencies for the subscales, conditional on background variables y.

In the analyses of 1988 NAEP data, a normal (Gaussian) form was assumed for $p(\theta|y_i)$, with a common dispersion and with a mean given by a linear model for selected main-effects and two-way interactions of the complete vector of background variables. The included background variables will be referred to as the *conditioning variables*, and will be denoted y^c. (The conditioning variables used in 1988 NAEP analyses are listed in Appendix C.) The following model was fit in each subject area:

$$\theta = \Gamma' \, y^c + \varepsilon \quad , \tag{9.6}$$

where ε is normally distributed with mean 0 and dispersion Σ. Γ and Σ are the parameters to be estimated. In subject areas with only one scale, such as reading, Γ is a vector and Σ is a scalar. In subject areas comprising subscales, Γ is a matrix and Σ is a covariance matrix. As in regression analysis, Γ is a vector or matrix of *effects* and Σ is the matrix or scalar *variance of residuals*. Also as in regression, the interpretation of the effects depends on how the design vectors in y^c are coded—as contrasts, or for linear effects, as examples. Appendix C gives the codings and estimates of effects of the present assessment. Like item parameter estimates, the estimates of the parameters of conditional distributions were treated as known true values in subsequent steps of the analyses.

Maximum likelihood estimates of Γ and Σ were obtained with Sheehan's (1985) M-GROUP computer program, using a variant of the EM solution described in Mislevy (1985). The difference from the published algorithm lies in the numerical approximation that was employed. Note from (9.5) that $p(\theta|x_i,y_i)$ is proportional to the product of two terms, the likelihood $P(x_i|\theta)$ and the conditional distribution $p(\theta|y_i)$. The conditional distribution has been assumed multivariate normal, with mean $\mu_i^c = \Gamma' \, y_i^c$ and covariance matrix Σ, if the likelihood is approximated by another normal distribution, with mean μ_i^L and covariance matrix Σ_i^L, then the posterior $p(\theta|x_i,y_i)$ is also multivariate normal with covariance matrix

$$\Sigma_i^p = (\Sigma^{-1} + (\Sigma_i^L)^{-1})^{-1} \tag{9.7}$$

and mean vector

$$\tilde{\theta}_i = (\theta_i^c \, \Sigma^{-1} + \theta_i^L \, \Sigma_i^{L\;-1})(\Sigma_i^p)^{-1} \quad . \tag{9.8}$$

In the 1988 analyses, a normal approximation for $P(x_i|\theta)$ is accomplished in a given scale by the steps described below. (Recall that by the assumed conditional independence across scales, the joint conditional likelihood for multiple scales is the product of independent likelihoods for each of the scales.) These computations are carried out in the scale determined by BILOG (Mislevy & Bock, 1982) item parameter estimates, where the mean and standard deviation of the composite population formed by combining the three NAEP grade/ages has mean zero and standard deviation one. The steps were as follows.

1) Lay out a grid of Q equally spaced points from -5 to +5, a range that covers the region in each scale where all examinees from all NAEP grade/age groups are virtually certain to occur. The value of Q varies from 20 to 40, depending on the scale being used; smaller values suffice for scales with few items given to each respondent, while larger values are required for scales with many items.

2) At each point X_q, compute the likelihood $L(x_i|\theta=X_q)$.

3) To improve the normal approximation in those cases in which likelihoods are not roughly symmetric in the range of interest—as when all of a respondent's answers are correct—multiply the values from Step 2 by the mild smoothing function

$$S(X_q) = \frac{\exp(X_q+5)}{[1+\exp(X_q+5)][1+\exp(X_q-5)]} .$$

This is equivalent to augmenting each examinee's response vector with responses to two fictitious items, one extraordinarily easy item that everyone gets right and one extraordinarily difficult item that everyone gets wrong. This expedient improves the normal approximation for examinees with flat or degenerate likelihoods in the range where their conditional distributions lie, but has negligible effects for examinees with even modestly well-determined symmetric likelihoods.

4) Compute the mean and standard deviation of θ using the weights $S(X_q)L(x_i|\theta=X_q)$ obtained in Step 3.

At this stage, then, the likelihood induced by a respondent's answers to the items in a given scale is approximated by a normal distribution. In subject areas where there is only one scale, a single normal distribution thus summarizes information from item responses. In an area such as mathematics or science where there are several scales, independent normal distributions, one per subscale, summarize information from responses to items from the several scales.

This normalized-likelihood/normal posterior approximation was then employed in both the estimation of Γ and Σ and in the generation of plausible values. From the final estimates of Γ and Σ, a respondent's posterior distribution was obtained from the normal approximation using the four-step procedure outlined above. A plausible value was drawn from this normal distribution—univariate normal, in subject areas like those in the 1988 survey with only a single scale; multivariate normal in areas like 1986 mathematics and science, with multiple subscales. For those subject areas with multiple subscales, weighted-average composites over subscales were also calculated after appropriate rescaling (see Beaton, 1988, for details and definitions of composites).

9.2.3.2 Computing Plausible Values in ARM Scales

In 1988 NAEP, the average response method (ARM) was used to carry out analyses of writing data. The ARM writing composite variable is defined to be an average rating (on the 0-to-4 rating scale for responses to essay prompts) over the set of essay exercises detailed in Chapter 11. Under the NAEP BIB-spiraled sampling design, no single student is administered more than four of these prompts. It is possible nevertheless to estimate consistently the covariance among each pair of exercises, setting the stage for the construction of plausible values.

The key step in the scale-score analyses of writing was the creation of sets of student-level plausible values. If a respondent had answered all questions going into the composite, then that respondent's ARM score would be directly calculable, without error, by

$$\theta = a'x$$

where x is the vector of the subject's responses to the n questions in the composite and a is a vector of n constants, each equal to 1/n. However, since no respondent was presented all writing exercises, composite values must be estimated by an application of the ARM technology. Briefly, the ARM technology, which is a kind of multiple regression, produces for each student a set of plausible values, each of which predicts what that student's composite score might plausibly be, based on the student's responses to the questions in the composite that were presented to the student and based on the student's status on the conditioning variables listed in Appendix C. ("Conditioning variables" for the ARM, just as for IRT-based plausible values, are the background variables with which cognitive responses are combined to yield respondents' predictive distributions for θ).

Let x_i represent the responses of the i^{th} student to the questions in the composite that were presented to that student and let y_i be the values of that student's conditioning variables. Then the k^{th} plausible value of the

ARM composite θ, based on the student's observed responses and conditioning variables is

$$\tilde{\theta}_{ik} = y_i' \hat{\Gamma} + x_i' \hat{\beta} + y_i' \gamma_k + x_i' \alpha_k + \varepsilon_{ik}$$

where

$\tilde{\theta}_{ik}$ is the k^{th} plausible value of the ARM composite,

$\hat{\Gamma}$ is the vector of estimated effects for the conditioning variables,

$\hat{\beta}$ is the vector estimated as giving the change in the composite variable for a unit change in the scores on each of the questions in x_i^o, with the linear effect of the conditioning variables held fixed,

$[\gamma_k, \alpha_k]$ is a random draw from a $N(0,\Sigma)$ distribution, where Σ is the estimated variance-covariance matrix of the estimates of $\hat{\Gamma}$ and $\hat{\beta}$ and reflects the uncertainty due to using sample estimates of the regression equation; and

ε_{ik} is an estimated residual drawn from a $N(0,\sigma_\varepsilon^2)$ distribution where σ_ε^2 is the variance of the predictive distribution of the ARM value given the observed values of y_i and x_i.

The parameters relating the responses on a given set of background questions ($\hat{\beta}$) and values of the conditioning variables ($\hat{\Gamma}$) with the means of the responses each of the questions in the ARM composite were estimated by least-squares technology. To accomplish this it is sufficient to obtain estimates of the means, variances, and interitem covariances, by conditioning subgroup, for the complete set of writing questions going into the composite. Because the ARM composite is the mean of the individual questions, this in turn produces estimates, by conditioning subgroup, of the ARM value mean and variance, as well as of the covariances between the ARM composite and each of the individual writing questions. These provide a complete set of sufficient statistics (the normal equations) for the standard least-squares prediction of an ARM composite value given conditioning variable characteristics and responses to any subset of the writing questions. (See Chapter 11 for details of the generation of the normal equations.)

Solving these normal equations produces the standard least-squares point estimate of a student's score on the composite, which is, in the above notation,

$$\hat{\theta}_i = y_i' \hat{\Gamma} + x_i' \hat{\beta} .$$

This standard estimate does not take into account the distribution of potential scores for any individual. In fact, $\hat{\theta}_i$ is an estimate of the mean of the predictive distribution of potential θs for the individual and, as such, does not address the likelihood of other values from this distribution, any one of which might also have been the student's ARM composite score had the student answered all the questions. By including terms that account for the uncertainty in the estimation of a student's composite score, the plausible values ($\tilde{\theta}_{ik}$) provide a more complete representation of what we do and do not know about the student's "true" composite score. (Note: For convenience we are treating the ARM composite as a continuous variable, it is in fact discrete, but can take a large number of closely spaced values.)

A check on the impact of the approximations and simplifying assumptions employed in the ARM approach was carried out with the writing data from the 1984 NAEP writing assessment (Beaton & Johnson, 1987). As a comparison for subgroup average writing scores, the same statistics were calculated using a totally different approach—the model-free, unbiased estimate for average responses based on the methodology employed by the Education Commission of the States in previous NAEP analyses. The latter method is prohibitively expensive to be used for all NAEP statistics, but could be calculated for the 44 questions in the common background questionnaire. Beaton and Johnson found that statistics based on the ARM were nearly indistinguishable from the model-free averages for those subgroups distinguished as conditioning variables, and for subgroups whose memberships were well-predicted by conditioning variables. Estimated standard errors were also smaller for the ARM estimates. For those subgroups that were neither conditioned on nor well-predicted by conditioning variables, the ARM exhibited biases. The nature of such biases in plausible values methodology is discussed further in section 9.3.3 of this report. Their causes, properties, and remedies are discussed at length in Mislevy (1988a, in press).

9.3 ANALYSES

When survey variables are observed without error from every respondent, standard variance estimators quantify the uncertainty associated with sample statistics from the only source, namely the sampling of respondents. Item percents correct for NAEP cognitive exercises meet this requirement, but scale-score proficiency values do not. The IRT and ARM models used in their construction posit an unobservable proficiency variable θ to summarize performance on the items in the area. The fact that θ values are not observed even for the respondents in the sample requires additional statistical machinery to draw inferences about θ distributions and to quantify the uncertainty associated with those inferences. As described above, we have adapted Rubin's (1987) multiple imputations procedures to the context of latent variable models to produce the plausible values upon which many analyses of the 1988 NAEP data are based. This section describes how plausible values were en' 'yed in subsequent analyses to yield inferences about population and sul 'ılation distributions of proficiencies.

9.3.1 Computational Procedures

E en though we do not observe the θ value of respondent i, we do observe variables that are related to it: x_i, the respondent's answers to the cognitive items he or she was administered in the area of interest, and y_i, the respondent's answers to demographic and background variables. Suppose we wish to draw inferences about a number $T(\underline{\theta},\underline{Y})$ that could be calculated explicitly if the θ and y values of each member of the population were known. Suppose further that if θ values were observable, we would be able to estimate T from a sample of N pairs of θ and y values by the statistic $t(\underline{\theta},\underline{y})$ [where $(\underline{\theta},\underline{y}) = (\theta_1,y_1,\ldots,\theta_N,y_N)$], and that we could estimate the variance in t around T due to sampling respondents by the function $U(\underline{\theta},\underline{y})$. Given that observations consist of (x_i,y_i) rather than (θ_i,y_i), we can approximate t by its expected value conditional on $(\underline{x},\underline{y})$, or

$$t^* (\underline{x},\underline{y}) = E[t(\underline{\theta},\underline{y})|\underline{x},\underline{y}]$$

$$= \int t(\underline{\theta},\underline{y})\, p(\underline{\theta}|\underline{x},\underline{y})\, d\underline{\theta}\ .$$

It is possible to approximate t^* with random draws from the conditional distributions $p(\theta_i|x_i,y_i)$, which are obtained for all respondents by the method describ d above in section 9.1. Let $\hat{\underline{\theta}}_m$ be the m^{th} such vector of "plausible values," consisting of a (possibly multidimensional) value for the latent variable of each respondent. This vector is a plausible representation of what the true $\underline{\theta}$ vector might have been, had we been able to observe it. The following steps describe how an estimate of a scalar statistic $t(\underline{\theta},\underline{y})$ and its sampling variance can be obtained from M (>1) such sets of plausible values. (Note: Five sets of plausible values were used in NAEP analyses in each subject area, and are provided on the NAEP public-use data tapes for secondary analysis.)

1) Using each set of plausible values $\hat{\underline{\theta}}_m$ in turn, evaluate t as if the plausible values were true values of $\underline{\theta}$. Denote the results $\hat{t}_m$, for m=1,...,M.

2) Using the multiple weight jackknife approach (see Chapter 8), compute the estimated sampling variance of $\hat{t}_m$, denoting the result U_m.

3) The final estimate of t is

$$t^* = \sum_{m=1}^{M} \hat{t}_m / M \ .$$

4) Compute the average sampling variance over the M sets of plausible values, to approximate uncertainty due to sampling respondents.

$$U^* = \sum_{m=1}^{M} U_m / M \ .$$

5) Compute the variance among the M estimates $\hat{t}_m$, to approximate uncertainty due to not observing θ values from respondents:

$$B_M = \sum_{m=1}^{M} (\hat{t}_m - t^*)^2 / (M-1) \ .$$

6) The final estimate of the variance of t^* is the sum of two components:

$$V = U^* + (1+M^{-1})\ B_M \ .$$

Note: Due to the excessive computation that would be required, NAEP analyses did not compute and average jackknife variances over all five sets of plausible values, but only on the first set. Thus, in NAEP reports, U^* is approximated by U_1.

9.3.2 Statistical Tests

Suppose that if θ values were observed for sampled students, the statistic $(t - T)/U^{1/2}$ would follow a t-distribution with d degrees of freedom. Then the incomplete-data statistic $(t^* - T)/V^{1/2}$ is approximately t-distributed, with degrees of freedom given by

$$\nu = \frac{d}{d + r_M^{-2}(M-1)}\ (M-1)\ (1+r_M^{-1})^2 \approx d\ \frac{(1+r_M)^2}{1 + (d\ r_M^2 /(m-1))} ,$$

where r_M is the relative increase in variance due to not observing θ values.

$$r_M = (1+m^{-1})\ B_M / U^* \ .$$

When B_M is small relative to U^*, the reference distribution for incomplete-data statistics differs little from the reference distribution for the corresponding complete-data statistics. This is the case with main NAEP reporting variables. If in addition d is large, the normal approximation can be used to flag "significant" results.

For k-dimensional t, such as the k coefficients in a multiple regression analysis, each U_m and U^* is a covariance matrix, and B_M is an average of squares and cross-products rather than simply an average of squares. In this case, the quantity

$$(T-t^*)\ V^{-1}\ (T-t^*)'$$

is approximately F distributed, with degrees of freedom equal to k and ν, with ν defined as above but with a matrix generalization of r_M:

$$r_M = (1+M^{-1})\ \mathrm{Trace}\ (B_M U^{*-1})/k.$$

By the same reasoning as used for the normal approximation for scalar t, a chi-square distribution on k degrees of freedom often suffices.

9.3.3 Biases in Secondary Analyses

Statistics t^* that involve proficiencies in a scaled content area and variables included in the conditioning variables y^c, are consistent estimates of the corresponding population values T. Statistics involving background variables y that were *not* conditioned on, or relationships among proficiencies from *different* content areas, are subject to asymptotic biases whose magnitudes depend on the type of statistic and the strength of the relationships of the nonconditioned background variables to the variables that were conditioned on and to the proficiency of interest. That is, the large sample expectations of certain sample statistics need not equal the true population parameters.

The *direction* of the bias is typically to underestimate the effect of nonconditioned variables. For details and derivations, the interested reader is referred to Beaton and Johnson (1987, 1990), Mislevy (1988a, in press), and Mislevy and Sheehan (1987, section 10.3.5). For a given statistic t^* involving one content area and one or more nonconditioned background variables, the *magnitude* of the bias is related to (1) the extent to which observed responses x account for the latent variable θ, and (2) the degree to which the nonconditioned background variables are explained by conditioning background variables. The first factor—conceptually related to test reliability—acts consistently in that greater measurement precision reduces biases in *all* secondary analyses. The second factor acts to reduce biases in certain analyses but increase it in others. In particular,

- High shared variance between conditioned and nonconditioned background variables *mitigates* biases in analyses that involve only proficiency and nonconditioned variables, such as marginal means or regressions.

- High shared variance *exacerbates* biases in regression coefficients of conditional effects for nonconditioned variables, when nonconditioned and conditioned background variables are analyzed jointly as in multiple regression.

In the 1984 NAEP reading assessment, the magnitude of shrinkage for the subgroup ans of a background variable that was not conditioned on averaged about 15 percent. Biases in multiple regressions that included conditioning variables averaged about 35 percent. Since that time, two importar steps have been taken to greatly reduce potential biases of this type. First is the move to the "focused-BIB" matrix-sampling design, under which all the cognitive tasks a respondent is administered are drawn from the same subject area. On the average, respondents are presented about twice as y tasks in the subject area than would have been presented under the full spiraling design, which administered each examinee tasks from one, two, or three subject areas. This increases the extent to which x accounts for θ, and, as noted above, decreases potential biases in all secondary analyses. Second is the increase in the number of background variables that can be included in the conditioning vector. This increases the number of secondary analyses that can be carried out with little or no bias, and mitigates biases in analyses of the marginal distributions of θ in nonconditioned variables. Bruce Kaplan and Jennifer Nelson's analyses of 1988 reading data (some results of which are summarized in Mislevy, in press) indicate that these improvements have slashed the potential bias for nonconditioned variables in multiple regression analyses from the 1984 level of 35 percent to approximately 10 percent, and biases in simple regression of such variables from 15 percent to 5 percent.

Table 9-1 gives representative results from an analysis in which Nelson estimated a number of substantively important effects from the 1988 NAEP reading data for 13-year-olds in two ways: (1) with the operational conditioning process, which included the listed effects in the conditioning vector, so their estimates are consistent; and (2) with no conditioning at all, so that the biases would be at their maxima. Also shown are results of an experimental method of conditioning, which employed only the first 32 principal components of the matrix of all 64 original conditioning vectors. The encouraging result from this analysis is that bias in analyses involving the original effects is virtually eliminated in this approach through the use of only half as many conditioning variables.

Table 9-1
Estimated Effects Based on Full, No, and Partial Conditioning

	Conditioning				
Effect	Full*	None	Bias	32 Components	Bias
Male-Female	-15.7	-14.4	-8%	-15.9	1%
White-Black	26.6	23.8	-11%	26.3	-1%
High Metropolitan-Low Metropolitan	32.6	30.5	-6%	32.8	1%
Northeast-Southeast	10.4	9.4	-10%	10.1	-3%
13-year-old Eighth Graders-13-year-old Seventh Graders	32.7	29.2	-11%	32.7	0

* Imputations constructed with conditional distributions that included 64 contrasts, including those shown here.

9.3.4 A Numerical Example

To illustrate how plausible values are used in subsequent analyses, this subsection gives some of the steps in the calculation of 1988 grade-level reading means and their estimation-error variances.

The weighted mean of the first plausible values of the grade 4 students in the sample is 230.68, and the jackknife variance of these values is 1.17. Were these values true θ values, then 230.68 would be the estimate of the mean and 1.17 would be the estimation-error variance The weighted mean of the second plausible values of the same students, however, is 230.60; the third, fourth, and fifth plausible values give weighted means of 230.19, 230.32, and 230.06. Since all of these figures are based on precisely the same sample of students, the variation among them is due to uncertainty about the students' θs, having observed their item responses and background variables. Taking the jackknife variance estimate from the first plausible value, 1.17, as our estimate U^* of sampling variance, and the variance among the five weighted means, .09, as our estimate B of uncertainty due to not observing θ, we obtain as the final estimate V of total error variance $1.17 + (1+5^{-1})\ .09 = 1.28$.

With U^* and B defined as above, and with $M=5$, we may obtain a value for Rubin's (1987) index characterizing the relative increase in variance due to the latency of θ; $r = .09$.

Corresponding values were also calculated for grade 8 and grade 12. The results are shown in Table 9-2.

Table 9-2
Estimation Error Variances and Related Coefficients
for the 1988 Grade-level Reading Assessments

Grade	U^*	B	V	r
4	1.17	.09	1.28	.09
8	.96	.06	1.03	.07
12	.69	.02	.71	.03

9.4 OVERVIEW OF SCALES IN THE 1988 NAEP ASSESSMENT

Scale-score analyses were carried out in the following subject areas in the 1988 NAEP assessment.

- Reading: 1 IRT trend scale, linking 1988 results to the 1971 - 1984 assessments, and 1 IRT cross-sectional scale, newly created in 1988

- Writing: 1 ARM trend scale, linking 1984 and 1988, and 1 ARM cross-sectional scale, newly created in 1988

- Civics: 1 IRT trend scale, linking 1988 results to the 1976 and 1982 assessments (neither of which had been previously scaled), and 1 IRT cross-sectional scale, newly created in 1988.

- U.S. history: 1 IRT scale, newly created in 1988.

- Geography: 1 IRT scale, newly created in 1988.

One subset of conditioning variables was employed in the creation of plausible values for all of these areas; it included only variables from the common set of background questions administered to respondents in all subject areas. The variables involved are listed in Table C-1 in Appendix C. A second subset of subject-specific conditioning variables was additionally included in the creation of plausible values for the subject areas of reading, writing, civics, U.S. history, and geography; these subsets included variables based on subject-related questions administered only to respondents who were administered cognitive items in the corresponding area. Tables C-2 through C-11 in Appendix C give the relevant variables and details of exactly how the background effects were coded in order to produce the conditioning vector y^c Conditional effect parameters Γ and the associated residual variances Σ were estimated separately in each subject area and in each grade/age Estimated effects for each subject and sample are given in Tables C-12 through C-37. Additional information on these analyses is presented in Chapters 10 through 14.

In the course of administering bridge (or trend) sample test forms comparable to those of the 1986 survey to further study trends in reading, data were serendipitously obtained for mathematics and science. Supplementary analyses involving these data are described in Chapter 15. These analyses included scale-score analyses as follows:

- Mathematics: 1 IRT scale, linked with the mathematics trend scale created in 1986.

- Science: 1 IRT scale, linked with the mathematics trend scale created in 1986.

The conditioning variables and codings used in the creation of plausible values for these two subject areas are listed in Tables C-10 and C-11 in Appendix C; estimated effects for the mathematics and science conditioning variables are given in Tables C-38 through C-43.

Chapter 10

DATA ANALYSIS FOR THE READING ASSESSMENT[1]

Rebecca Zwick

Educational Testing Service

The analysis of the 1988 reading data had three main goals:

- to continue the long-term reading trend results begun in 1971 for ages 9, 13, and 17,

- to produce detailed analyses of reading proficiency for grades 4, 8, and 12, and

- to continue the investigation of the anomalous reading results in 1986.[2]

The details of the first two types of analyses are given in this chapter. The 1988-1989 investigation of the 1986 reading anomaly is described in *The Effect of Changes in the National Assessment: Disentangling the NAEP 1985-86 Reading Anomaly* (Beaton & Zwick, 1990).

This chapter has three main sections. First, the samples of students who received reading items in the 1988 NAEP assessment are described. Then, the procedures used for the long-term trend analysis are detailed. Finally, the procedures used for the cross-sectional analysis are discussed.

---

[1] David Freund provided statistical programming, with the assistance of Minhwei Wang and Kate Pashley. Robert Mislevy provided consultation on scaling. Jo-Ling Liang assisted with analyses.

[2] In 1986, reading trend results for ages 9 and 17 appeared implausibly low. Their release was therefore delayed and a three-year investigation ensued. (Despite the large change in reading proficiency estimates, the 1986 findings resembled past assessments and other reading measures with respect to the ordering of subgroup differences and the relation of reading proficiency to background variables and to performance in mathematics and science. It was therefore decided to release a cross-sectional report, *Who Reads Best?* [Applebee, Langer, & Mullis, 1988], using a scale metric that differed from the 1984 reading scale.) The anomalous 1986 reading trend results are not included in NAEP's reading trend report, but are documented separately in Beaton and Zwick, 1990, which describes the investigation of these results and their subsequent adjustment. In brief, the study showed that changes in assessment technology had affected the 1986 estimates of reading proficiency

10.1 SAMPLES OF STUDENTS

In 1988, reading items were administered to 14 samples of students in the 1988 assessment, as shown in Table 10-1:

- The samples that constituted the bridge to 1984 (9[Br84-RW], 13[Br84-RW], and 17[Br84-RW]) played a dual role in the 1988 reading analyses: They served as the basis of the long-term reading trend estimates, reported in *The Reading Report Card, 1971 to 1988: Trends from the Nation's Report Card* (Mullis & Jenkins, 1990) and were used in the 1988-1989 phase of the investigation of the 1986 reading anomaly.

- The "focused-BIB" samples for reading (i.e., the balanced incomplete block spiral samples 9[Main-Rdg], 13[Main-Rdg], and 17[Main-Rdg], which received three reading blocks) served as the basis of the grade-level cross-sectional analyses reported in *Learning to Read in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Langer, Applebee, Mullis, & Foertsch, 1990).

- The intercorrelation samples, 9[Main-Int], 13[Main-Int], and 17[Main-Int], which were also a part of the main assessment, received booklets that contained reading, civics, and U.S. history items, and, at grade 12/age 17, geography items. The scaling of the results from these samples allows researchers to examine the relation between these subject areas.

- For grade 8/age 13 and grade 12/age 17, the main assessment also included samples of students (9[Main-Doc] and 13[Main-Doc]) who received two booklets consisting of NAEP reading blocks along with document literacy items that had been administered as part of NAEP's 1985 study of adult literacy (Kirsch & Jungeblut, 1986). These booklets differed from the remaining main NAEP booklets for grade 8/age 13 and grade 12/age 17 in that the grid of response ovals for the items followed immediately after each item in the booklet; for the remainder of the main NAEP samples for the two upper age classes, a separate scannable answer sheet was used. (Table 10-2 shows the response modes for all portions of the 1988 reading assessment.) Preliminary item analyses showed that when reading items were administered in these document literacy booklets, they behaved in a substantially different manner than in the focused-BIB assessment. Therefore, the reading item responses from these booklets were not scaled, but are available for secondary analysis. Item results for the document literacy blocks (which, unfortunately, showed that large proportions of students failed to reach some items) are also available on NAEP user tapes.

Table 10-1
NAEP 1988 Reading Samples

Sample Code	Subject Areas	Booklet Numbers	Cohort	Time of Assessment	Age Defn.	Modal Grade	Sample Size
9[Br84-RW]	R, W	51-56	Grade 4/age 9	Winter	CY	4	5,188
13[Br84-RW]	R, W	51-56	Grade 8/age 13	Fall	CY	8	5,500
17[Br84-RW]	R, W	51-56	Grade 11/age 17	Spring	not CY	11	4,622
9[Main-Rdg]	R	8-14	Grade 4/age 9	Winter, spring	CY	4	6,177
13[Main-Rdg]	R	8-14	Grade 8/age 13	Winter, spring	CY	8	5,912
17[Main-Rdg]	R	8-14	Grade 12/age 17	Winter, spring	CY	12	5,768
9[Main-Int]	R, C, H	17-19	Grade 4/age 9	Winter, spring	CY	4	2,638
13[Main-Int]	R, C, H	29-31	Grade 8/age 13	Winter, spring	CY	8	2,590
17[Main-Int]	R, C, H, G	30-32	Grade 12/age 17	Winter, spring	CY	12	2,438
13[Main-Doc]	R, D	35-36	Grade 8/age 13	Winter, spring	CY	8	2,533
17[Main-Doc]	R, D	36-37	Grade 12/age 17	Winter, spring	CY	12	2,425
9[Br86-RMS]	R, M, S	91-93	Age 9	Winter	CY	4	3,711
13[Br86-RMS]	R, M, S	91-93	Age 13	Fall	CY	8	3,942
17[Br86-RMS]	R, M, S	61-66	Grade 11/age 17	Spring	not CY	11	4,703

R - Reading
W - Writing
C - Civics
H - U.S. history
G - Geography
D - Document Literacy

M - Mathematics
S - Science
CY - Calendar year: birth dates in 1978, 197 , and 1970 for ages 9, 13, and 17
not CY - (Age 17 only): birth dates between October 1, 1970 and September 30, 1971

Table 10-2

Response Modes for NAEP 1988 Reading Items

Sample Code	Sample Type	Booklet Numbers	Response Mode
9[Br84-RW]	Bridge to 1984	51-56	Circle response in booklet
13[Br84-RW]	Bridge to 1984	51-56	Circle response in booklet
17[Br84-RW]	Bridge to 1984	51-56	Circle response in booklet
9[Main-Rdg]	BIB Main	8-14	Fill in oval in booklet
13[Main-Rdg]	BIB Main	8-14	Fill in oval on separate sheet
17[Main-Rdg]	BIB Main	8-14	Fill in oval on separate sheet
9[Main-Int]	BIB Intercorrelation	17-19	Fill in oval in booklet
13[Main-Int]	BIB Intercorrelation	29-31	Fill in oval on separate sheet
17[Main-Int]	BIB Intercorrelation	30-32	Fill in oval on separate sheet
13[Main-Doc]	BIB Document Literacy	35-36	Fill in oval in booklet
17[Main-Doc]	BIB Document Literacy	36-37	Fill in oval in booklet
9[Br86-RMS]	Bridge to 1986	91-93	Fill in oval in booklet
13[Br86-RMS]	Bridge to 1986	91-93	Fill in oval in booklet
17[Br86-RMS]	Bridge to 1986	61-66	Fill in oval in booklet

- The reading blocks in the bridge-to-1986 samples, 9[Br86-RMS], 13[Br86-RMS], and 17[Br86-RMS], were used only in the analysis of the reading anomaly and are, therefore, discussed further in this report.

10.2 LONG-TERM TREND ANALYSIS—BRIDGE TO 1984 SAMPLES

The 1988 bridge to 1984 included, at each age level, six of the assessment booklets administered in 1984. These booklets contained both reading and writing blocks, as well as background items. Although these bridge booklets represented only about a tenth of the reading booklets administered in the complex 1984 BIB design,[3] they contained 10 of the 12 reading blocks that were scaled at each grade/age level in 1984. The number of students who were given each of the bridge items ranged from 768 to 927. The samples of students who received these bridge booklets are described in Table 10-1 and in Chapter 3. The purpose of the long-term reading trend analysis was to add to the reading trend results that extended from 1971 to 1984 for ages 9, 13, and 17.[4] Dimensionality analyses conducted following the 1984 assessment showed that the reading items were well summarized by a unidimensional scale (Zwick, 1987a). The analysis steps were as follows:

1. *Conduct preliminary item analyses and verify the fit of the 1984 item parameters to the 1988 bridge data.* For each item, calculations were made of the percent of students selecting each response, the percent who omitted the item, the percent who did not reach the item, and the correlation between the item score and the block score. Also, for each item block, the internal consistency reliability was computed. The block-level KR-20 reliabilities (for scaled multiple-choice items only) ranged from .66 to .83, with a median of .74, at age 9; from .46 to .74, with a median of .64, at age 13; and from .31 (for a block with only four scaled items) to .75, with a median of .67, at age 17. The item analyses revealed that at least 10 percent

---

[3]The bridge to 1984 included 1984 booklets 16, 17, 27, 34, 55, and 60 at age 9 and booklets 13, 16, 17, 21, 34, and 57 at ages 13 and 17 (see J. R. Johnson, 1987, pp. 120-121). The 1984 BIB assessment included 57 booklets that contained at least one scaled reading block at age 9 and 56 such booklets at ages 13 and 17.

[4]Note that the current estimates of reading proficiency results for 1971 through 1984 differ slightly from those that appear in *The Reading Report Card* (NAEP, 1985) for three reasons. The first is that an improvement in the capacity of the computer program used for the conditioning phase of proficiency estimation (see Chapter 9) allowed for an increase in the number of conditioning variables, leading to improved proficiency estimates for the years 1971 through 1984. Also, additional records for 1971 that had been omitted from tapes provided by the previous contractor were recovered. The third change resulted from a recomputation of the 1984 weights for grade 4/age 9 and grade 8/age 13. This change is detailed in Appendix E.

of the students failed to reach 22 items at age 9, seven items at age 13, and one item at age 17. Unlike the main NAEP assessment, however, the bridge blocks cannot be altered to reduce speededness. Evidence of speededness in the main assessment is presented below.

Because the purpose of the analysis of the bridge data was to provide a link to the 1984 reading results, it was desirable to use the parameters that had been obtained for the bridge items in the 1984 item calibration, provided that the fit of the 1984 parameters to the 1988 data was acceptable. (The 1984 analysis included all three grade/age cohorts in a single calibration.) In 1988, all multiple-choice items were included in the scaling process. There were 99 such items at ages 9 and 13 and 87 at age 17. The total number of scaled items was 182. (The decision to exclude open-ended items was influenced by consideration of the role of the bridge data in the reading anomaly investigation: Because a previous study [Zwick, 1988] had shown that scoring inconsistencies had affected these items, their inclusion could have complicated the study of the reading anomaly.) Out of the 182 items, 65 items were administered to grade 8/age 13 and grade 12/age 17, 37 items were given to grade 4/age 9 and grade 8/age 13, 25 items were given to grade 4/age 9 and grade 12/age 17, and 24 items were given to all three age classes.

For all 182 items to be included in the 1988 scaling, plots of the 1984 item response functions along with the 1988 data showed the fit of the 1984 items to the 1988 data to be acceptable. These item parameters are given in Table F-2 of Appendix F.

2. *Obtain the proficiency means and standard errors, and percents of students above each scale anchoring point for the NAEP reporting groups.* Although scale values were obtained for the grade/age samples, only the age samples were used for trend reporting, following the NAEP tradition. Sample sizes were 3,782, 4,005, and 3,652 for ages 9, 13, and 17. For each of the NAEP reporting categories, reading proficiency means and standard errors were computed using the technology described in Chapter 9. A list of the conditioning variables, the scheme for coding them, and their estimated effects appear in Tables C-19 to C-21 of Appendix C. In addition, for each reporting category, a determination was made of the percent of students exceeding each of the scale anchoring levels determined in 1984: Rudimentary (150), Basic (200), Intermediate (250), Adept (300), and Advanced (350). The methods for deriving the anchor points is outlined in Chapter 7 and described in detail in Beaton (1987a).

10.3 CROSS-SECTIONAL ANALYSIS—MAIN SAMPLES

In 1988, reading items were administered in seven focused BIB booklets, each of which contained three reading blocks, in addition to background items. The samples of students who received these items are listed in Table 10-1 and described in Chapter 3. In addition, at each grade/age level, one reading block was included in each of the three booklets administered to the intercorrelation samples. The numbers of reading items administered to the main samples are given in Table 10-3, along with the numbers of items that were new in 1988. Reading objectives for the new items are documented in *Reading Objectives, 1986 and 1988 Assessments* (NAEP, 1987a; see Chapter 2).

Table 10-3
NAEP 1988 Reading Items—BIB Samples

Sample Code	Number of Reading Scale Items	Number of Reading Scale Items New to 1988
9[Main-Rdg]	83	43
13[Main-Rdg]	100	53
17[Main-Rdg]	110	66
Total	215	126

Of the total of 215 items, 52 items (including three intact blocks) were administered to both grade 8/age 13 and grade 12/age 17, 26 items were given to both grade 4/age 9 and grade 8/age 13, and 5 items were given to all three age classes. The number of students in the focused-BIB and intercorrelation samples who were given each item ranged from 2,459 to 3,536.

Four of the 215 items were open-ended; the remainder were multiple-choice. Two of the four open-ended items were administered at grade 12/age 1 only, one at grade 4/age 9 only and one at both grade 4/age 9 and grade 8/ag 13. These items were rated on ordinal scales by professional judges. Scoring procedures and reliability results are described in Chapter 6.2. For purposes of item response theory (IRT) scaling, the range of possible scores was then dichotomized into "correct" and "incorrect" categories using rules provided by reading experts. Table 10-4 shows the number of points in the ordinal scale for each item, along with the dichotomization rule.

Table 10-4
Dichotomization Rules for Open-ended Reading Items Used in Scaling

NAEP Item	Grade/Age	Score Range for Valid Responses	Scores Considered Correct
N015905 High Tech Pizza	12/17	1 - 4	2 - 4
R002406 Small Fruits	12/17	1 - 2	2
R000106 Ant and Dove	4/9	1 - 5	4 - 5
R000806 Grandpa and Wind	4/9, 8/13	1 - 5	3 - 5

The primary purpose of the analysis of the main NAEP data was to provide reading results for grades 4, 8, and 12, and to investigate the relation of reading proficiency to student background and attitudes and to teacher attributes and instructional methods. The analysis steps were as follows.

1. *Conduct item analyses.* Item analysis procedures like those described in connection with the trend analysis were applied to the main NAEP blocks. Block-level KR-20 reliabilities (for multiple-choice items only) ranged from .62 to .84, with a median of .81, at grade 4; from .66 to .87, with a median of .84, at grade 8; and from .74 to .85, with a median of .81, at grade 12. An important finding of the item analyses was that, although NAEP assessments are not intended to be speeded, some of the main NAEP reading blocks, like the bridge blocks, tended to be too long for substantial numbers of examinees. At least 10 percent of the students failed to reach 13 items at grade 4, 10 items at grade 8, and three items at age 17. The percent of students who failed to reach (or chose not to respond to) the open-ended items, which are placed at the end of reading blocks in NAEP, was always at least 20. The information gained from the 1988 assessment about the numbers of items that students can complete was used in determining the number of items for the 1990 item blocks. It is hoped that, by reducing the length of blocks, speededness problems can be avoided in the future.

It was also discovered in examining the reading results that the performance on each item block was affected by the position of the block within the booklet. Table 10-5 shows the average percent of students who failed to reach the items in each block for each of the three possible block positions. In nearly every case, the average percent not reached is greater for position 3 than for positions 1 and 2. Particularly notable are the large average percents not reached for the grade 4 blocks R3 and R7 in position 3. The average percent correct for each block in each of the three possible positions is shown in Table 10-6. In nearly every case, the average percent correct declines as the position moves from 1 to 3. The effect of block position is most severe for grade 4, followed in order by grade 8 and grade 12. The effect is most striking for block R3 at grade 4, for which the average percent correct for position 1 is nearly 11 percent points higher than for position 3. In NAEP's computation of the percent correct for an item, students who did not reach the item are not included[5]; therefore, these differences in average percents correct are not simply the result of students' failure to reach items that occur in later blocks. Perhaps fatigue leads to more errors in these blocks. Further investigation of this issue is planned.

---

[5]Percent correct is defined as R/(R + W + O + DK), where R, W, O, and DK represent the sum of the student weights for those who got the item right, those who got the item wrong, those who reached the item but omitted it, and those who indicated that they did not know the answer, respectively. The DK option is included only in certain trend items. Students who did not reach the item are not included in the computation. An item is considered "not reached" if the student did not respond to the item and did not give a valid response to any of the succeeding items within the item block. Each block of items is separately timed and therefore, the determination of which items are to be considered "not reached" is made separately by block.

Table 10-5

Average Percent of Students Failing to Reach Items
for Each Block Position

Block*	Number of Items	Position 1	Position 2	Position 3
Grade 4				
R2	14	5.8	5.0	7.0
R3	7	4.7	4.8	9.8
R4	7	2.4	2.2	6.0
R5	15	5.5	4.1	6.8
R6	14	5.0	6.5	5.2
R7	15	3.7	3.1	10.0
R8	12	4.6	3.6	8.4
Grade 8				
R2	14	3.6	3.0	5.5
R3	9	0.9	0.3	2.8
R4	10	0.6	1.7	2.3
R5	14	2.3	1.8	2.3
R6	18	5.1	4.8	7.0
R7	19	3.5	4.[illegible]	3.9
R8	1[illegible]	2.5	2.3	3.9
Grade 12				
R2	14	3.9	3.0	6.1
R3	11	4.2	4.7	6.3
R4	19	0.9	1.6	3.1
R5	14	1.4	1.4	7.0
R6	15	3.4	4.1	5.7
R7	19	2.4	3.0	3.4
R8	18	1.5	3.7	5.9

---

* Blocks R2, R5, and R7 are identical for grades 8 and 12. Except in these cases, identity of block numbers across grades does not imply that blocks are identical.

Table 10-6

Average Item Percent Correct
for Each Block Position

Block*	Number of Items	Position 1	Position 2	Position 3
Grade 4				
R2	14	74.5	72.9	69.9
R3	7	72.0	69.2	61.4
R4	7	56.1	57.7	52.7
R5	15	75.3	74.7	73.5
R6	14	69.2	65.9	65.4
R7	15	71.0	69.6	67.5
R8	12	72.7	69.2	65.7
Grade 8				
R2	14	58.8	57.2	53.0
R3	9	85.7	84.9	81.4
R4	10	73.5	75.7	68.7
R5	14	45.9	47.0	47.6
R6	18	67.3	65.3	64.9
R7	19	71.3	67.1	65.8
R8	16	84.7	83.6	79.4
Grade 12				
R2	14	70.9	71.5	69.5
R3	11	70.8	69.7	67.5
R4	19	73.6	73.2	72.1
R5	14	62.2	62.8	61.2
R6	15	66.0	65.9	62.0
R7	19	80.2	79.3	76.3
R8	18	76.3	75.9	72.4

* Blocks R2, R5, and R7 are identical for grades 8 and 12. Except in these cases, identity of block numbers across grades does not imply that blocks are identical.

2. *Calibrate the items and investigate differential item functioning across grade levels and racial/ethni and gender groups.* Strictly speaking, the findings on reading block position effects imply the need for an item response model that allows item parameters to depend on block position; that is, a model in which each item in the main assessment would have three sets f item parameters, one corresponding to each possible position. Such a model would, of course, be unwieldy and its parameters could not be estimated accurately with current sample sizes. Fortunately, however, the focused-BIB portion of the assessment has the property that each block appears in each position exactly once. Therefore, item parameter estimates based on the focused-BIB assessment can be viewed as appropriately weighted averages of these three sets of parameter estimates. This property of the item parameter estimates would not hold if the three intercorrelation booklets were included in the item calibration (see Tables 4-2, 4-3, and 4-4); hence, the calibration was based only on the focused-BIB portion of the assessment.

For all three grade/ages combined, the BILOG program (Mislevy & Bock, 1982) was used to obtain item parameter estimates on a provisional scale, based on the three-parameter logistic model. Sampling weights were not used in this phase of the analysis. Parameters were estimated even for previously administered items; that is, parameter values for old items were not assumed equal to their previous values. To reduce costs, a systematic sample consisting of half of the students in the focused-BIB samples 9[Main-Rdg], 13[Main-Rdg], and 17[Main-Rdg] was used in the item calibration. The three samples were treated as distinct subpopulations in the BILOG run; that is, in estimating the item parameters, the densities for the three grade/age groups were not assumed to be the same.

Using the method described in Chapter 7, a graphical analysis was conducted to determine whether it was reasonable to assume common item response functions for all three age classes. For each item and each group, expected proportions correct (see Mislevy & Sheehan, p. 302) for each of approximately eight proficiency levels were obtained. The departures of these proportions from the common estimated item response function were examined. In the case of four items, the assumption of a common item response function for all age classes appeared to be seriously violated. Therefore, a second calibration was conducted in which these items were allowed to have different item response functions across age classes. Table 10-7 lists these items and indicates how they were treated in the second calibration. The first two items refer to a single passage about getting summer jobs. It is not surprising that these items functioned differently for students of different ages: Given an eighth grader and a twelfth grader with equal reading proficiency, the older student is more likely to answer correctly because of greater familiarity with methods for getting summer jobs. The next two items in Table 10-7 required treatment as a separate item for grade 4. As noted earlier, the response mode for grade 4 differed from that for grades 8 and 12 (Table 10-2). It is unclear, though, why these items in particular should be affected. Perhaps certain items are more susceptible to response mode differences. Using the same method, residuals were also examined for Black, Hispanic, and White students and for males and females. No major or systematic departures were found

Table 10-7
Items for Which Multiple Item Response Functions Were Estimated

NAEP Item	Grade/Age	Calibration Procedure
N003202 Summer job-best time to find	8/13, 12/17	Treat separately
N003204 Summer job-references	8/13, 12/17	Treat separately
N002401 Mosquito-size exaggerated	4/9, 8/13	Treat separately
R000904 Naomi James-year she set record	4/9, 8/13, 12/17	Treat 4/9 separately; treat 8/13 and 12/17 together

Item parameter estimates for the 215 items administered to the focused-BIB and intercorrelation samples are given in Table F-1 of Appendix F.

3. *Obtain the proficiency means and standard errors on a provisional scale for the NAEP reporting groups for the focused-BIB and intercorrelation samples.* As in 1984, responses to reading items were summarized in a single reading scale. Although scale values were obtained for the grade/age samples, only the grade subsamples of the focused-BIB reading sample were used as the basis of the reading cross-sectional report; sample sizes were 4,534, 4,404, and 4,250 for grades 4, 8, and 12. For each of the NAEP reporting categories, means and standard errors were computed using the technology described in Chapter 9. Note that reading background items were used as conditioning variables for the focused-BIB samples. However, with the exception of a single item, these items were not administered to the students in the intercorrelation samples, who received civics and history background items (with one reading background item) instead. Because different conditioning models were required, the conditioning process was conducted separately for the focused-BIB and intercorrelation samples. For each of these samples, a list of the conditioning variables, the scheme for coding them, and their estimated effects, appear in Tables C-12 to C-17 of Appendix C.

4. *Determine the appropriate metric for reporting, investigate linkage of the 1988 main NAEP results with the 1984 reading scale.* Another component of the analysis was the exploration of the feasibility of linking the 1988 main NAEP results with the reading trend scale established in 1984. Such a linkage appeared unlikely to succeed because the 1988 assessment differed in major ways from the 1984 assessment. The time of year at which assessments were administered, the age definition for 17-year-olds, and the response mode

(see Table 10-2) all differed. Furthermore, although 72 of the items administered in 1988 were also given in 1984, these items did not appear within intact 1984 item blocks. Through our investigation of the 1986 reading anomaly, we learned that reading items may behave in a substantially different manner when they appear in altered contexts and that other seemingly minor changes can have major effects on equating. Nevertheless, three equating approaches were applied in an attempt to achieve an approximate linking. If the three equating efforts were to yield similar equating functions, this would support the validity of the link.

Two of the equating efforts took advantage of the existence, at ages 9 and 17, of subsamples of the bridge to 1984 and the 1988 main NAEP samples that had common age definitions and times of assessment. At each of these age levels, the RESOLVE program (which implements the procedures described in Mislevy, 1984) was used to estimate the reading proficiency distribution of the bridge and main NAEP students, based on item responses and item parameters. (The 1984 item parameters were used for the bridge to 1984, item parameters from step 2, above, were used for the main NAEP sample.) The linear function required to match the means and standard deviations of the estimated bridge and main NAEP distributions was then obtained. In addition to these two common population equating methods, a common item equating procedure was applied to items common to 1984 and 1988 despite the fact that items did not appear in intact blocks. The Stocking-Lord (1983) procedure, implemented in the TBLT program (Stocking, 1986; see also Sheehan, 1988) was used to find the best-fitting linear transformation for mapping the provisional parameters for the 68 common items, obtained in the 1988 calibration, to the existing 1984 parameters for these items. (The three items common to 1984 that were treated separately across age classes in 1988 were not included in the set of items common between 1988 and 1984. In addition one 1988 open-ended item that had been administered in 1984 was not treated as a common item because a study of score drift [Zwick, 1988] showed that rating criteria were not applied consistently in 1984 and 1988. The resulting number of effective common items was 68.)

The estimated equating functions from these three methods were quite disparate, reinforcing the preliminary conclusion that the two assessments were too different to be linked in a satisfactory way. Therefore, as in 1984, the mean and standard deviation for the three grade/ages combined were set to 250.5 and 50.[6] We plan to link the 1990 main NAEP results to this new scale.

After determining the metric for reporting, another decision that had to be made was whether to combine the results from the intercorrelation samples with those from the focused-BIB samples. As noted earlier, only the focused-BIB samples were used in item calibration, but scale values were obtained for both samples. However, the results from the intercorrelation samples differed

[6]The metric in which the item parameters in Appendix F are given differs by a linear transformation from the reading scale used for reporting. Letting θ represent the proficiency metric that corresponds to the item parameters and letting RS represent the metric of the reporting scale, the required transformation is $RS = 50\theta + 250.5$.

in two ways from the focused-BIB results. First, they were not balanced with respect to block position and second, they were based on a less complete conditioning model. Examination of subgroup means and standard deviations for the two samples showed some small but systematic differences between the two. In particular, the subgroup means for sample 9[Main-Int] tended to be slightly higher than those for sample 9[Main-Rdg] and the standard deviations for sample 13[Main-Int] tended to be lower than those for 13[Main-Rdg]. Therefore, only the focused-BIB results were used as the basis for *Learning to Read in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Langer, Applebee, Mullis, & Foertsch, 1990). The scale values for the intercorrelation samples are available on the 1988 public-use data tapes.

5. *Conduct a study of the relation between the reading proficiency of fourth graders and teacher background and instructional variables.* Teacher background data and information on instructional methods used with specific students were collected from the reading teachers of 3,901 fourth graders (about 86 percent of the fourth-grade students in sample 9[Main-Rdg]). There are several possible reasons for the absence of teacher data for 14 percent of the students: Some teachers may not have been identified, others were not sampled, and others may have failed to complete questionnaires. Also, teachers were not required to complete questionnaires for more than ten students. Further detail about the process for sampling teachers is given in Chapters 3 and 8.

For reasons explained here, the analysis of the relation of these teacher data to reading proficiency requires the use of separate sets of reading scale values and sample weights. The goal of these analyses was to report reading proficiency results for each possible response to selected teacher items. For example, student reading proficiencies were computed for each of several levels of teacher education and for each of several levels of frequency of individualized instruction. In order to avoid biases in these analyses (see Chapter 9), it was necessary to include as conditioning variables the teacher items that had been selected for reporting. These conditioning variables were, of course, unavailable for the nonrandom 14 percent of the focused-BIB students whose teachers did not complete questionnaires. Because the missing data problem affected a large, nonrandom portion of the sample, the use of a common conditioning model in which the teacher items were simply treated as missing for certain students could produce a substantial violation of the assumption of equal residual variances in each conditioning cell. Therefore, a separate set of plausible values was obtained for students with teacher data and these were compared to the original set. Because there were some slight differences, the two sets of plausible values were retained—one to be used for analyzing the relation of reading proficiency to the teacher data and one for other analyses of reading data. For the teacher-based plausible values, a list of the conditioning variables, the scheme for coding them, and their estimated effects, appear in Table C-18 of Appendix C.

In any case, analyses involving the teacher data must be conducted separately from other analyses because of the need to use a separate set of sampling weights. For the teacher analyses, the teacher-based student weights must be used. The teacher-student weight is the nonresponse-adjusted student weight further adjusted for the teacher's probability of selection and for teacher nonresponse. (Some further adjustments were also applied; see Chapter 8 and *The 1988 National Assessment of Educational Progress—Sampling and Weighting Procedures, Final Report* [Rust, Bethel, Burke, & Hansen, 1990].) The teacher-based plausible values and the teacher-student weights are available on the 1988 public-use data tapes.

Analyses involving the teacher background and instructional variables are included in *Learning to Read in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Langer, Applebee, Mullis, & Foertsch, 1990).

Chapter 11

DATA ANALYSIS FOR THE WRITING ASSESSMENT[1]

Eugene G. Johnson

Educational Testing Service

This chapter describes the analyses carried out on the responses to the writing tasks and the background items in the 1984 and 1988 assessments of writing that led to the results reported in *The Writing Report Card, 1984-88: Findings from the Nation's Report Card* (Applebee, Langer, Mullis, & Jenkins, 1990) and *Learning to Write in Our Nation's Schools: Instruction and Achievement in 1988 at Grades 4, 8, and 12* (Applebee, Langer, Mullis, Jenkins, & Foertsch, 1990). The emphasis is on the methods and results of the procedures used to develop the average response method scale scores that formed the basis of those reports.

The analysis of the 1988 writing data consisted of two components. The first component was designed to measure trends in writing achievement since 1984. Trends in writing achievement were measured by comparing the responses to a set of writing tasks by students assessed in 1984 with the responses to the same set of writing tasks for students assessed in 1988. The major analyses were made for trends in average task accomplishment (primary trait), based on a writing scale developed using the average response method, although trends in writing mechanics and trends in overall writing fluency (based on holistic scoring) were also measured. The data forming the bas for these analyses were from the 1988 bridge-to-1984 (trend) samples and 1984 main samples (these samples are d ed below). The techniques used to measure trends in writing achievemer discussed in section 11.1.

The second component in the analyses of the writing data was designed to analyze the responses to the writing tasks in the main assessment, to develop a cross-sectional writing scale based on these data, to examine the association between writing ability and writing process and instruction, and to determine the effect that time allocated has on writing ability. These analyses were based on the data from the 1988 main assessment samples. Section 11.2 provides a description of the analyses of the cross-sectional data.

---

[1]The statistical programming for the average response method was ably performed by Bruce Kaplan Data analysis and additional statistical programming were performed by Bruce Kaplan, Jo-Ling Liang, Mike Narcowich, and Inge Novatkoski. The author is indebted to Mary Varone for typing the manuscript.

The specific samples used for the analysis of writing achievement in 1984 and 1988 are (by age cohort):

Sample Code	Sample Type	Time of Assessment	Age Definition	Modal Grade
9:84	1984 Main	Winter	Calendar year	4
9[Main-Wrt]	1988 Main	Winter, spring	Calendar year	4
9[Main-LWr]	1988 Long Writing	Winter, spring	Calendar year	4
9[Br84-RW]	1988 Bridge to 1984	Winter	Calendar year	4
13:84	1984 Main	Fall	Calendar year	8
13[Main-Wrt]	1988 Main	Winter, spring	Calendar year	8
13[Main-LWr]	1988 Long Writing	Winter, spring	Calendar year	8
13[Br84-RW]	1988 Bridge to 1984	Fall	Calendar year	8
17:84	1984 Main	Spring	Oct. - Sept.	11
17[Main-Wrt]	1988 Main	Winter, spring	Calendar year	12
17[Main-LWr]	1988 Long Writing	Winter, spring	Calendar year	12
17[Br84-RW]	1988 Bridge to 19	Spring	Oct. - Sept.	11

11.1 MEASUREMENT OF TRENDS IN WRITING ACHIEVEMENT

The data for the 1988 point in the analysis of trends in writing comes from the 1988 bridge-to-1984 samples, which provide a bridge to the 1984 assessment and match that assessment in terms of the time of administration and age definitions. The data for the 1984 point in writing trend comes from the responses of students within the 1984 main (i.e., BIB) assessment. All analyses of trends in writing performance were based on grade-eligible students only. For reasons given below, both the 1988 and the 1984 points were determined by scores provided by raters scoring the papers in 1988.

The items on which the trends in writing achievement are based are shown in Table 11-1. The table shows the block that contained the item in 1984 and the trend booklets containing the item in 1988. A total of twelve writing tasks were used to measure trends in writing achievement, with six tasks presented at each grade in 1984 and 1988. To allow comparisons in writing ability across grades, three of the six tasks presented to the fourth-grade students in 1988 were also presented to the eighth-grade students; three of the eighth-grade tasks were also presented to the eleventh-grade students; and one task was presented at all three grades.

11.1.1 Primary Trait Scoring of the Writing Tasks and Measures of Scorer Effect

All writing exercises from the 1988 assessment were scored for task accomplishment (primary trait). For the purposes of analysis, the student responses were coded as 0=not rated, 1=unsatisfactory, 2=minimal, 3=adequate, and 4=elaborated. A 20 percent random subsample of all the papers scored were

Table 11-1

Assignment of 1984-1988 Writing Trend Items in 1984 and 1988

Writing Task	1984 BIB-spiral Blocks Grades 4	8	11	1988 Bridge to 1984 Booklets Grades 4	8	11
N0003 Recreation Opp.	-	C	C	-	52,54	52,54
N0004 Food On Frontier	D	D	D	-	51,54	51,54
N0005 Dissecting Frogs	-	E	-	-	53,55	-
N0006 XYZ Company	E	E	-	52,54	53,55	-
N0009 Radio Station	G	G	-	54,55	55,56	-
N0010 Appleby House	G	G	G	54,55	55,56	55,56
N0076 Flashlight	V	V	V *	56	-	-
N0147 Plants	C	-	-	51,53	-	-
N0148 Spaceship	E	-	-	52,54	-	-
N0180 Space Program	-	-	E	-	-	53,55
N0190 Job Application	-	-	E	-	-	53,55
N0210 Bike Lane	-	-	G	-	-	55,56

* Block V never appeared with any other writing block in 1984 (all other blocks appeared with every other block at the same grade in 1984).

rescored by a second rater to provide an estimate of interrater reliability. Table 11-2 shows scorer reliability for each essay as measured by the intraclass correlation. Additionally, Table 11-2 shows the percentage of exact score agreement between the first and second raters. The reliabilities and percents of exact agreement are generally high and are consistent with equivalent values from prior assessments.

Although the measures of scorer agreement in NAEP have been consistently high, we recognized the possibility that there might be variation between the ratings provided by the group of scorers assembled in 1988 and the scorers assembled in 1984. If present, this variation would add a confounding effect in the measurement of trend. The most direct way of controlling the effect of across-year variation in scoring would be to eliminate it entirely by rescoring all of the 1984 data, using the same set of scorers who scored the 1988 data. Unfortunately, resources did not allow for the rescoring of the full set of 1984 writing papers but did allow for a rescoring of approximately 11,000 of the papers given in 1984. The rescored papers for a given item constituted approximately a 25 percent sample of all 1984 papers and consisted of all grade-eligible respondents to two or three of the 1984 booklets containing that item.

The hope was that the between-year variability in scoring would be low enough to permit the use of the full set of the 1984 data. Tables 11-3 and 11-4 show the results of the comparison of the rescore of the 1984 data with the scores assigned to the papers in 1984. Table 11-3 shows, by grade and item, the average difference between the 1988 rescore and the 1984 score and the standard deviation of the difference. Table 11-4 shows the distribution of the difference between the rescore and the original score, again by age and item. The average difference between the rescore and the original score is -.019 for grade 4, -.046 for grade 8 and -.056 for grade 11, differences of the same general magnitude as the between-group differences that were reported in the previous *Writing Report Card*. Although the 1988 scorers appear to have been more stringent on average than the 1984 scorers, the tables show that the difference between the two groups of scorers is not simply a consistent drift.

In light of the differences between the 1984 and the 1988 scoring, direct comparisons between the 1988 results and the original 1984 results were not considered acceptable. Consequently, the 1984 trend point was based on the rescored data only. The resultant sample sizes for the trend report analyses are given in Table 11-5.

(It should be noted that another option that was considered was to adjust the original 1984 scores to reflect the observed relationship between the rescores and the original scores on the set of rescored data. Such an adjustment of the 1984 scores would use a multiple imputation procedure in which adjusted 1984 scores for a given item are drawn from the multinomial distribution of possible original scores for that item conditional on a given rescore value. While such an approach would use the full set of 1984 data, the results would include a component of uncertainty inherent in the imputation of scores. This component can be large: Limited analyses indicate

Table 11-2

Percentages of Exact Score Agreement and Interrater Reliability for the Primary Trait Scoring of the 1988 Writing Trend Items

	Grade 4		Grade 8		Grade 11	
NAEP Item	Percent Exact Agreement	Reliability	Percent Exact Agreement	Reliability	Percent Exact Agreement	Reliability
N000602 XYZ Company	97.1	.985	93.5	.920		
N000902 Radio Station	93.5	.950	87.0	.886		
N001002 Appleby House	90.3	.916	75.3	.688	89.3	.888
N007602 Flashlight	87.5	.874				
N014702 Plants	94.3	.946				
N014802 Spaceship	91.8	.952				
N000302 Recreation Opp.			85.4	.820	90.8	.931
N000402 Food on Frontier			79.9	.683	93.1	.863
N000502 Dissecting Frogs			76.1	.642		
N018002 Space Program					89.9	.928
N019002 Job Application					92.3	.922
N021002 Bike Lane					84.9	.866
AVERAGE	92.4	.937	83.0	.773	90.0	.900

Table 11-3

Mean and Standard Deviation of (Rescore - Original) for the 20% Rescore of 1984 Writing Responses

NAEP Item	Grade 4 Mean	Grade 4 S. D.	Grade 8 Mean	Grade 8 S. D.	Grade 11 Mean	Grade 11 S. D.
N000602 XYZ Company	-.010	.311	.057	.559		
N000902 Radio Station	-.039	.342	-.034	.450		
N001002 Appleby House	.003	.339	-.055	.538	-.064	.375
N007602 Flashlight	-.043	.367				
N014702 Plants	.000	.356				
N014802 Spaceship	-.021	.399				
N000302 Recreation Opp.			-.046	.588	-.118	.547
N000402 Food on Frontier			-.117	.533	-.092	.454
N000502 Dissecting Frogs			-.063	.563		
N018002 Space Program					-.086	.517
N019002 Job Application					.031	.504
N021002 Bike Lane					-.006	.527
OVERALL	-.019		-.046		-.056	

Table 11-4

Distribution of (Rescore-Original),
for the 20% Rescore of 1984 Writing Responses
Percent of Responses Where (Rescore-Original) Equals -1,0 or 1

NAEP Item	Grade 4 Rescore-Original			Grade 8 Rescore-Original			Grade 11 Rescore-Original		
	-1	0	1	-1	0	1	-1	0	1
N000602 XYZ Company	2.6	93.7	2.6	2.5	83.0	9.7			
N000902 Radio Station	5.9	90.8	3.[illegible]	10.5	81.4	7.5			
N001002 Appleby House	5.3	89.5	4.9	15.8	73.1	10.3	9.5	86.4	3.8
N007602 Flashlight	8.2	87.5	3.9						
N014702 Plants	4.8	90.2	4.1						
N014802 Spaceship	6.0	87.7	5.3						
N000302 Recreation Opp.				11.9	75.0	10.1	15.5	75.4	6.9
N000402 Food on Frontier				16.2	75.0	7.2	13.3	81.2	4.6
N000502 Dissecting Frogs							17.0	70.1	12.2
N018002 Space Program							14.7	77.4	6.7
N019002 Job Application							6.4	83.7	7.6
N021002 Bike Lane							11.6	75.7	11.6
OVERALL	5.4	89.9	4.0	12.3	76.2	9.5	11.8	80.0	6.9

Table 11-5

Sample Sizes for Primary Trait Analyses
of Trends in Writing Performance

NAEP Item	Grade 4		Grade 8		Grade 11	
	1988 Bridge	1984 Rescore	1988 Bridge	1984 Rescore	1988 Bridge	1984 Rescore
N000602 XYZ Company	1152	544	1334	616		
N000902 Radio Station	1234	585	1364	612		
N001002 Appleby House	925	530	1256	588	1041	599
N007602 Flashlight	614	609				
N014702 Plants	1285	656				
N014802 Spaceship	1258	611				
N000302 Recreation Opp.			1372	494	1242	521
N000402 Food on Frontier			1339	603	1212	629
N000502 Dissecting Frogs			1356	641		
N018002 Space Program					1195	632
N019002 Job Application					1169	603
N021002 Bike Lane					11783	636

that the standard errors based on the rescored data only are of the same general size as the standard errors of the full, imputed-up-to-1988 data.)

11.1.2 The Writing Trend Scale Based on the Average Response Method

The major analysis of trends in writing achievement was based on the average response method (ARM) of scaling nonbinary data (Beaton and Johnson, 1987, 1990), although analyses of individual items were conducted also. Since the measurement of trend was based on the rescored data, the ARM writing scale established in 1984 could not be used. Instead, a new ARM writing scale was constructed based on the trend items given in 1988. Because certain of the between-item correlations are not estimable, the ARM technology has been generalized to allow for the imputation of missing correlations.

11.1.2.1 Overview of the Average Response Method

The average response method begins with a defined composite of the (primary trait) scores for a set of p exercises and provides, for each assessed student, draws from the distribution of potential values for that composite. If a student had responded to all the exercises going into the composite, then that student's ARM score would be directly calculable, without error, by

$$\theta = a'x$$

where x is the vector of the subject's scores on the p questions in the composite and a is a vector of p arbitrary constants, which for the ARM writing scale are each equal to 1/p (since the ARM writing scale is defined as the predicted average performance across the set of p writing questions).

Because each respondent is presented only a subset of the questions, the respondent's composite value is unknown and so must be estimated. Such an estimate is provided by the ARM technology. Briefly, the ARM technology is a kind of multiple regression that produces for each student a set of plausible values, each of which predicts what that student's composite score might plausibly be, based on the student's scores on the exercises in the composite that were presented to the student and based on the student's status on a selected set of background variables, called the conditioning variables. (The purpose of the conditioning variables is to improve the prediction of subgroup effects given the sparse quantity of information available for any individual. Their exclusion would lead to biased estimates of subgroup effects. See Chapter 9 for a general definition and description of plausible values and conditioning variables).

Let x_i represent the (row) vector of responses of the ith student to the questions in the ARM composite that were presented to that student and let y_i be the (row) vector of values of that student's conditioning variables. Then a plausible value from the conditional distribution of θ given the observed data x_i and y_i for student i is

$$\tilde{\theta}_{ik} = x_i\hat{\beta} + y_i\hat{\Gamma} + x_i\alpha_k + y_i\gamma_k + \epsilon_{ik} \tag{11.1}$$

where

$\tilde{\theta}_{ik}$ is the kth plausible value of the ARM composite

$\hat{\beta}$ is the (column) vector giving the change in the composite for unit change in the scores on each of the questions in x_i

$\hat{\Gamma}$ is the (column) vector of effects for the conditioning variables

α_k and γ_k are random draws from a multivariate normal distribution with mean vector 0 and variance-covariance matrix Σ where Σ is the variance-covariance matrix of the parameter estimates $\hat{\beta}$ and $\hat{\Gamma}$. (α_k and γ_k reflect the uncertainty due to using sample estimates $\hat{\beta}$ and $\hat{\Gamma}$ in the regression equation.)

ϵ_{ik} is an estimated residual drawn from a normal distribution with mean 0 and variance σ_ϵ^2 where σ_ϵ^2 is the variance of the predictive distribution of θ given the observed values of x_i and y_i.

All parameters in equation (11.1) were estimated by least-squares technology. To accomplish this, it is sufficient to obtain estimates of the elements of the population sum of squares and cross products matrix of the conditioning variables and the writing questions:

$$C = \text{an estimate of } V'V = \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix} \tag{11.2}$$

In the above, Y is a $N \times q$ matrix containing the values of the q conditioning variables for each of the N students in the population; X is a $N \times p$ matrix containing the scores of the N students in the population on the p exercises; and $V = [Y\ X]$. If Y and X were known for all students in the population, C would be trivially equal to V'V. However, since only a sample of the students in the population were assessed for writing and since each sampled student was only presented a few writing questions, many of the elements of Y and X are unknown. Accordingly, V'V must be estimated. The procedures used to determine an estimate C of V'V will be deferred to the next subsection.

Since the ARM composite is the mean of the individual questions, the estimate C generates a complete set of sufficient statistics (the normal equations) for the standard least-squares prediction of an ARM composite value given conditioning variable characteristics and responses to any subset of writing questions. Define the N element column vector T by

$$T = Xa$$

where the elements of T, θ_i, are the values of the composite for each student in the population. The exact value of θ_i will not be known unless the student i was administered all p of the exercises. The plausible values, $\tilde{\theta}_{ik}$, of equation (11.1) are determined by operations on the matrix C_θ where C_θ is the estimated population sum of squares and cross product matrix of the conditioning variables, the writing exercises and the composite. C_θ is generated by the matrix C and the transformation matrix

$$H = \begin{bmatrix} I_q & 0 & 0 \\ 0 & I_p & a \end{bmatrix}$$

by

$$C_\theta = H'CH = \begin{bmatrix} Y'Y & Y'X & Y'T \\ X'T & X'X & X'T \\ T'Y & T'X & T'T \end{bmatrix}$$

The matrix C_θ can be used to estimate a value of θ_i for student i as follows:

Let X_1 consist of the columns of X corresponding to the writing exercises presented to student i and let $V_1 = [Y\ X_1]$. The least-squares estimates of $\hat{\beta}$ and $\hat{\Gamma}$ in equation (11.1) are

$$\begin{bmatrix} \hat{\Gamma} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} Y'Y & Y'X_1 \\ X_1'Y & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y'T \\ X_1'T \end{bmatrix} \qquad (11.3)$$

and the standard least-squares point estimate of the composite score for student i is

$$\hat{\theta}_i = x_i\hat{\beta} + y_i\hat{\Gamma} .$$

This value is the mean of the predictive distribution of potential θs for the individual and, thus, does not take into account the fact that any other value from this predictive distribution might also have been the student's score. By including the terms accounting for the uncertainty in the estimation of a student's composite score, the plausible values $\tilde{\theta}_{ik}$ allow the more complete representation of what is known and what is not known about the student's composite scores. The terms accounting for uncertainty are of two types:

1) ϵ_{ik}, accounting for variability of potential scores of an individual about the conditional mean (of the distribution given y_i and x_i) and

2) α_k and γ_k, accounting for uncertainty due to using sample estimates of $\hat{\beta}$ and $\hat{\Gamma}$ in the regression equation.

ϵ_{ik} is a random draw from a $N(0,\sigma_\epsilon^2)$ distribution, where σ_ϵ^2 is the residual mean-squared-error for the regression defined by (11.3). The vector

$$\begin{bmatrix} \gamma_k \\ \alpha_k \end{bmatrix}$$

is a draw from a multivariate normal distribution with mean 0 and variance-covariance matrix

$$\Sigma = (V_1'V_1)^{-1} \sigma_\epsilon^2.$$

The values of α_k and γ_k are held fixed for all student. with the same pattern of missing data.

A further discussion of the generation of ARM plausible values, given an estimate C of the sum of squares and cross product matrix V'V, appears in Beaton and Johnson (1990). The next section considers the estimation of V'V for the 1988 and 1984 writing trend scale.

11.1.2.2 Estimation of V'V

As noted in the previous subsection, the basis for the estimation of a predicted value for any student is an estimate C of the full sums-of-squares-and-cross products matrix

$$V'V = \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix} \qquad (11.4)$$

from which all other necessary matrices and estimates are derived. For the construction of the NAEP writing trend scale, six separate estimates of the cross-product matrix were created: one for each of the three grades for each of the years 1984 and 1988. The elements of the estimate C of V'V for a particular grade and year fall into three general types:

Type 1: Elements that are directly estimable from the available data for that grade and year; these are sums of squares and cross-products involving the conditioning variables and the items presented to that grade in that year.

Type 2: Elements that must be estimated based on relationships observed for another grade in the same year or for the same grade in the other year; these are sums of squares and cross-products involving items and pairs of items not administered to the target grade and year but administered to another grade or year.

Type 3: Elements requiring the imputation of between item correlations; these are cross-product terms involving pairs of items that have never been presented together so that the between item correlation is not estimable.

The complexity of the estimation of these three types of elements of the matrix C increases as one moves down the list, with elements of type 1 being the most straightforward to estimate and elements of type 3 being the most complex. We will consider the estimation of each of the three types in turn.

11.1.2.3 Estimation of Type 1 Elements of C

The most directly estimable elements of the matrix C for a given grade and year are those involving the conditioning variables and the items presented to students of that grade who were assessed in the specified year.

In the matrix C, the conditioning matrix Y contains the values of 39 conditioning variables measuring each student's status on a set of demographic, background, and attitude questions as well as specific questions related to the student's attitudes to and experiences in writing. A list of the conditioning variables and the scheme for coding them appears as Table C-5 in Appendix C.

The values of the conditioning variables are known for all students and so Y'Y in the matrix C is directly obtained by taking the sum of squares and cross-products of the conditioning variables for each student, weighting these by the student's sampling weight and then summing across all students of the given grade and assessment year.

Also directly estimable are sums of squares and certain cross-products involving the writing tasks presented to students in the given grade in the given assessment year. Specifically, if the item X_1 was presented to a sample of students of the given grade in the given year, then the following terms are estimable:

$Y'X_1$ the cross-product vector between the conditioning variables and the item, and

$X_1'X_1$ the sum-of-squares for the item. Furthermore, if another item, X_2 was also presented to the grade in the given year and if a sample of students responded to both X_1 and X_2, then the between-item cross-product terms $X_1'X_2$ and $X_2'X_1$ are also directly estimable.

Because of the spiral design, each item is presented to a subsample of the full sample of grade-eligible students assessed for writing in a given year. Consequently, the simple procedure used to estimate the elements of Y'Y is inappropriate for estimating the elements $Y'X_1$, $X_1'X_1$ and $X_1'X_2$. However, these elements can be consistently estimated in the following manner.

Because of the spiral design, we can assume that the set of students who responded to the item X_1 is a representative sample of the population of all grade-eligible students in the given year. Consequently, the appropriately weighted sample mean $\overline{X}_1$, and the weighted sample variance, S_1^2, based on the total sample of students of the grade in the assessment year responding to the item, are consistent and unbiased estimates of the population mean and variance for the item. A consistent estimator of the sum-of-squared scores in the population for the item is

$$X_1 \hat{'} X_1 = W_{TOT}(S_1^2 + \overline{X}_1^2)$$

where W_{TOT} is the sum of weights for all grade-eligible students assessed for writing in the given year.

A consistent estimate of the cross-product element $X_1'X_2$, when a sample of grade-eligible students were presented both items in the given year, is

$$X_1 \hat{'} X_2 = W_{TOT}(S_1 S_2 r_{12} + \overline{X}_1 \overline{X}_2)$$

where S_1 and $\overline{X}_1$ are the weighted sample standard deviation and mean based on the full sample of grade-eligible students responding to item 1, S_2 and $\overline{X}_2$ are the analogous statistics based on the full sample of grade-eligible students responding to item 2, and r_{12} is the appropriately weighted sample correlation coefficient based on the students responding to both items.

The estimation of cross-products between conditioning variables and items, $Y'X_1$, is accomplished in an analogous manner.

Table 11-6 shows the items used in the writing trend scale (by supplying the last three digits of the ID number from Table 11-1) and indicates the elements of the cross-product matrix V'V that can be directly estimated at each grade and year. The entries in the table are codes giving the grade and year for which the corresponding element of item by item cross-product matrix is estimable. The presence of a 4, 8, or E indicates that the element can be estimated from the 1984 sample of students in grade 4, 8 or 11, respectively. The presence of an asterisk indicates that the element can also be estimated from the 1988 data. For example, the first element in the table, in the row labeled 003 and the column labeled 003, indicates the grades and years for which the sum-of-squares of the item N003 can be directly estimated from the

Table 11-6

Estimable Between-item Correlations
with Grade and Year Where Estimable

Item	1984 ARM Items						Grade 4 Trend Items		Grade 11 Trend Items		
	003	004	005	006	009	010	147	148	180	190	210
003	8E*	8E*	8	8	8	8E			E	E	E
004		48E#	8	48	48	48E	4	4	E	E	E
005			8*	8*	8*	8*					
006				48*	48*	48*	4	4*			
009					48*	48*	4	4*			
010						48E*	4	4*	E*	E*	E*
147							4*	4			
148								4*			
180									E*	E*	E*
190										E*	E*
210											E*

4: correlation estimable at grade 4 in 1984
8: correlation estimable at grade 8 in 1984
E: correlation estimable at grade 11 in 1984

* indicates correlations estimable from 1988 data
# indicates correlations estimable from 1988 data for grades 8 and 11

data. The "8E" indicates that $X_1'X_1$ for this item can be estimated for grade 8 and grade 11 students in 1984, the asterisk indicates the $X_1'X_1$ can also be estimated for grade 8 and grade 11 students in 1988. The "8" in the third column of the first row (under the 005) indicates that the cross-product term $X_1'X_2$ between items N003 and N005 can only be estimated for grade 8 students in 1984. The table also indicates for which grades and years the cross-products between conditioning variables and item responses can be estimated. any time the item sum-of-squares $X_1'X_1$ can be estimated, the conditioning variable by item cross-product vector $Y'X_1$ can also be estimated.

Besides indicating the terms of the matrix V'V that are directly estimable for a given grade and year, Table 11-6 also indicates the terms that are not: the Type 2 and Type 3 terms. For example, the sum-of-squares of the item N003 is a Type 2 term for grade 4 students in 1984 and in 1988 since it cannot be directly estimated from data from grade 4 students but can be estimated from data from other grades. An example of a Type 3 element is the cross-product term between items N003 and N0147. The blank at that position in Table 11-6 indicates that this term cannot be directly estimated for any grade or year. This is because items N003 and N0147 have never been presented together to any student. The procedures for estimating Type 2 terms are given in the next subsection (11.1.2.4). Subsection 11.1.2.5 gives the procedure for estimating Type 3 terms.

11.1.2.4 Estimation of Type 2 Elements of C

As described above, a Type 2 element of the cross-product matrix for a given grade and year corresponds to a term that is not directly estimable based on data for that grade and year (the *target* grade and year), but that can be estimated based on data from another grade or another year (the *donor* grade and year). The estimation procedure is as follows.

Let X_1 be the set of items held in common between the target grade and year and the donor grade and year and let X_2 be the set of items presented to the donor but not to the target. For notational convenience, we will operate as if the entire donor and target population had been measured and that complete information by student is available for all items presented to the students grade and year. There is no loss of generality because only estimates of the terms of the cross-product matrix are required.

The known information for the target population is the matrix

$$V_t = [Y_t \; X_{1t}]$$

consisting of the conditioning variables and the items held in common with the donor population. The known information for the donor population is the corresponding matrix

$$V_d = [Y_d \; X_{1d}]$$

plus the set of items X_{2d}.

We seek estimates of the Type 2 terms $V_t'X_{2t}$ and $X_{2t}' X_{2t}$ in the cross-product matrix

$$C_t = \begin{bmatrix} V_t'V_t & V_t'X_{2t} \\ X_2'{}_tV_t & X_{2t}'X_{2t} \end{bmatrix}$$

Beaton and Johnson (1987) show that the estimator

$$V_t\hat{'}X_{2t} = (V_t'V_t)\ (V_d'V_d)^{-1}V_d'X_{2d}$$

is unbiased under the assumption that $[X_{1d}\ X_{2d}]$ and $[X_{1t}\ X_{2t}]$ are each distributed with common variance matrix Σ and mean matrices Y_dB and Y_tB. Under the same assumptions they show that $X_{2t}'X_{2t}$ can be unbiasedly estimated by

$$X_{2t}\hat{'}X_{2t} = X_{2d}'V_d\ (V_d'V_d)^{-1}\ V_t'V_t(V_d'V_d)^{-1}\ V_d'X_{2d} + (N - G)\ \hat{\Sigma}_{2\cdot1}$$

where

$G =$ trace $[(V_d'V_d)^{-1}\ V_t'V_t]$,

N is the number of students taking any pair of items

$\Sigma_{2\cdot1} = (X_{2d}'X_{2d} - X_{2d}'V_d(V_d'V_d)^{-1}\ V_d'X_{2d})/(N-m)$

and

m is the number of columns of V_d.

All estimates of Type 2 elements were accomplished in the above manner. Frequently, there were a number of candidates for the donor population. The following hierarchy was followed in selecting the population to be the donor for the given target population:

1) same grade as the target and the other year
2) same year as the target and the closest grade
3) any grade or year where data was available

11.1.2.5 Estimation of Type 3 Elements of C

The final step in the construction of an estimate C of the cross-product matrix V'V for a given grade and year is the estimation of the Type 3 elements. These are cross-product terms, $X_1'X_2$, between pairs of items that were never jointly presented to any student in any grade or year. Such item pairs are identified by a blank in Table 11-6.

For a concrete example, consider the cross-product term between the items N0003 and N0147. The item N0003 was presented to grade 8 and grade 11 students in 1984 and 1988 while the item N0147 was presented to grade 4 students in 1984 and 1988. Consequently, the cross-product between this pair of items is not directly estimable at every grade and year since the items have never been jointly presented to a sample of students.

Let X_1 and X_2 represent such a pair of items and consider estimating $X_1'X_2$ for the target grade and year, t. At most one of the two items can have been administered to the target population, although it is possible that neither item was administered to the target population. Let a index the donor grade and age for item X_1 and let b index the donor grade and age for item X_2. (Either a or b, but not both, might correspond to the target population t.)

The known data from the donor population a are the conditioning variables, Y_a, and the scores X_{1a} on the item 1. The known data from the donor population b are the conditioning variables, Y_b, and the scores X_{2b} on the item 2. Assume that, to a reasonable degree of approximation,

$$X_{1a} = Y_a\beta_1 + E_{1a}$$

and $$X_{2b} = Y_b\beta_2 + E_{2b}$$

where β_1 and β_2 are vectors of unknown constants and E_{1a} and E_{2b} are vectors of independent and identically distributed errors where each element of E_{1a} has mean 0 and variance σ_1^2 and each element of E_{2b} has mean 0 and variance σ_2^2. Further assume that, had the items been presented to the target population, then the scores X_{1t} and X_{2t} on the items for the target population could be modeled as

$$X_{1t} = Y_t\beta_1 + E_{1t}$$

$$X_{2t} = Y_t\beta_2 + E_{2t}$$

where β_1 and β_2 are as above, where the means of E_{1t} and E_{2t} are zero, and where the variances of each element of E_{1t} and E_{2t} are, respectively, σ_1^2 and σ_2^2.

Then, if X_{1t} and X_{2t} had been jointly presented to the target population, the expectation of $X_{1t}'X_{2t}$ would be

$$\beta_1'Y_t'Y_t\beta_2 + N\,\sigma_{12} \qquad (11.5)$$

where σ_{12} is the covariance between the two items after adjusting for the conditioning variables. An unbiased estimate of the first term of (11.5) is

$$\hat{\beta}_1'Y_t'Y_t\hat{\beta}_2 \qquad (11.6)$$

where $\hat{\beta}_1 = (Y_a'Y_a)^{-1}\, Y_a'X_{1a}$

and $\hat{\beta}_2 = (Y_b'Y_b)^{-1}\, Y_b'X_{2b}$.

An estimator of σ_{12} is

$$\hat{\sigma}_{12} = \hat{\sigma}_1\ \hat{\sigma}_2\ r_{12} \qquad (11.7)$$

where $\hat{\sigma}_1^2 = [X_{1a}'X_{1a} - X_{1a}'Y_a\ (Y_a'Y_a)^{-1}\ Y_a'X_{1a}]/(N-p)$

and $\hat{\sigma}_2^2 = [X_{2b}'X_{2b} - X_{2b}'Y_b\ (Y_b'Y_b)^{-1}\ Y_b'X_{2b}]/(N-p)$

and where p is the rank of Y_a and Y_b. It remains to get an estimate of r_{12}, the correlation between the two items. The estimate of r_{12} was based on the observed distributions of the between item correlations for the target grade and age.

Let R_t be the vector of between item correlations for all pairs of items jointly presented to the target population t. A typical element of R_t, r_{ij}, is the observed correlation between item i and item j, after adjusting for the conditioning variables Y_t. Let

$$z_{ij} = \operatorname{arctanh}(r_{ij}) = \tfrac{1}{2}\log_e\{(1 + r_{ij})/(1 - r_{ij})\}$$

be the result of applying Fisher's variance stabilizing transformation to the correlation r_{ij} and let Z_t be the vector of the z_{ij}'s. The distribution of z_{ij} is modeled as

$$z_{ij} = a + b_i + b_j + \epsilon_{ij} \qquad (11.8)$$

where

a is a constant common to all elements of Z_t;

b is a constant common to all elements of Z_t involving the item i;

b_j is a constant common to all elements of Z_t involving the item j; and

ϵ_{ij} is a normal random variable with mean 0 and variance σ_Z^2.

Let $\hat{a}$, $\hat{b}_i$, and $\hat{b}_j$ be the least-squares estimates of a, b_i and b_j in (11.8) and let $\hat{\epsilon}_{ij}$ be the least-squares residual. The set of all least-squares residuals from the additive fit to Z_t was assumed to be a set of normally distributed random variables with common mean 0 and common variance σ_Z^2.

Let $E_t = \{\hat{\epsilon}_{ij}$: i> j, items i and j jointly presented to population t} be the set of unique residuals. (That is, the elements of Z_t come in pairs, there being a value z_{ij} and a value z_{ji}, which are equal, for any pair of items presented to the target population. Consequently, the least-square residuals $\hat{\epsilon}_{ij}$ and $\hat{\epsilon}_{ji}$ are identical; only one of these is included in E_t.)

Let E_t have K elements and let

$$e_{(1)} \le e_{(2)} \le \ldots \le e_{(K)}$$

be the ordered values of E_t. Under the assumptions, the $e_{(i)}$ are order statistics of a sample of size K from a $N(0, \sigma_Z^2)$ distribution. Form the standard normal working values

$$w_1 \le w_2 \le \ldots \le w_K$$

where w_i is the standard normal deviate for p bility (i - 1/3)/(K + 1/3). (w_i is a close approximation to the median o the distribution of the i^{th} order statistic from a sample of K standard normal random variables.)

A regression of the $e_{(i)}$'s on the w_i's produces estimates of the mean and standard deviation of the normal distribution best fitting the distribution of the residuals in E_t: The intercept estimates the mean and the slope estimates the standard deviation, σ_Z. The above process was conducted separately for each of the six grade-by-year populations. It was found in each case that the empirical distribution of the residuals was quite closely approximated by a normal distribution with mean 0 and standard deviation σ_Z and that the values of the standard deviations for each of the six populations were nearly the same.

Accordingly, the elements of E_t for all six grade-by-year populations were pooled to form a combined set of least-squares residuals with M elements (where M is the sum of the number of elements in each of the E_t), sorted to form the order statistics $e_{(1)} \le e_{(2)} \le \ldots \le e_{(M)}$ and regressed against the standard normal working values for a sample of size M. The result of this regression was an intercept of 0 and a slope of .09. The fit of this regression was excellent, with an $R^2 = .996$, implying that the prediction of the least squares residuals by the regression on the normal working values accounted for 99.6 percent of the variance of the actual values. Consequently, the distribution of the least-squares residuals from the model (11.8) is closely approximated by a normal distribution with a mean of 0 and a standard deviation of .09, taken to estimate σ_z for all six grade-by-year populations.

The imputation of a value for population t of $\tilde{r}_{12}$ and, hence, of $X_1'X_2$, for a pair of items that never appeared together was accomplished via the following steps:

1) draw a random variable $\tilde{e}_{12}$ from a normal distribution with mean 0 and standard deviation $\sigma_z = .09$.

2) form the imputed value $\tilde{z}_{12} = \hat{a}_t + \hat{b}_{1a} + \hat{b}_{2b} + \tilde{e}_{12}$ where

   $\hat{a}_t$ is the least-squares estimate of the constant common to all Fisher-transformed correlations in the target population;

   $\hat{b}_{1a}$ is the least-squares estimate of the constant common to all Fisher-transformed correlations involving the item 1 in the donor population a for the item; and

   $\hat{b}_{2b}$ is the least-squares estimate of the constant common to all Fisher-transformed correlations involving the item 2 in the donor populations b for the item.

3) back-transform $\tilde{z}_{12}$ to form the estimate of the correlation r_{12}:

$$\tilde{r}_{12} = [\exp(2\,\tilde{z}_{12}) - 1]/[\exp(2\,\tilde{z}_{12}) + 1]$$

4) form the imputed value of $X_1'X_2$ for the target population by

$$X_1'\tilde{X}_2 = \hat{\beta}_1'Y_t'Y_t\hat{\beta}_2 + N\,\hat{\sigma}_1\,\hat{\sigma}_2\,\tilde{r}_{12}$$

   where $\hat{\beta}_1$ and $\hat{\beta}_2$ are defined by (11.6) and $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are defined by (11.7).

11.1.2.6 Application to the Writing Trend Scale

For each of the three grades and each of the two years, separate estimates of the matrix C were made. The order of estimation was as follows. First, all Type 1 elements in all six cross-product matrices were estimated. The second step was the estimation of the Type 2 elements for each grade and year based on the Type 1 estimates from the donor grade and year. The order of preference for selecting the donor population was: (1) same grade, (2) another adjacent grade in the same year, and (3) any available grade and year Finally, the Type 3 elements were estimated. To allow for the measurement of the error due to imputation of the between-item correlations, five cross-product matrices were generated for each grade and year, each matrix C being based on a different set of draws from the distribution of correlations.

The resultant matrices were then used as the basis for constructing the C_θ matrices as was detailed in 11.1.2.1. Five C_θ matrices were created for each grade and year, each based on one of the five C matrices. To approximately account for the effects of the sample design and the amount of information available, each of the C_θ matrices was scaled to be consistent with a sample of 1,000. For each of the five C_θ's for a given grade and age, a single plausible value was computed for each student according to the formulas in 11.1.2.1. This resulted in five plausible values for each student.

11.1.3 Other Analyses of Trends in Writing Performance

In addition to trends in primary trait scores, trends were also measured for mechanics of writing and for overall writing fluency. Trends in components of mechanics of writing at each age were based on a selected writing prompt given to the age group in both 1984 and 1988. The writing items used for the assessment of the mechanics of writing were "Spaceship" (N014840) for grade 4 and "Recreation Opportunity" (N000310) for grades 8 and 11. All analyses were based on representative subsamples of around 500 responses to each item at each grade and year. In the sample selection, Black students were sampled at a higher rate to provide sufficient sample size to allow for comparisons in performance between Black and White students. The student weights were adjusted to reflect this oversampling of Black students by the following poststratification process: 1) for each grade, the students selected for the writing mechanics analysis were categorized by gender and by race/ethnicity (White, Black, Hispanic, other), producing eight cells; 2) the sampling weights of the students within each cell were then multiplied by a poststratification factor computed as a ratio whose denominator is the sum of weights of all students in the cell selected for the mechanics analysis and whose numerator is the sum of the weights of all students in the writing assessment of the specified grade, gender, and race/ethnicity. All papers used in this analysis were scored in 1988; the actual sample sizes are shown in Table 11-7.

Table 11-7
Sample Sizes for Mechanics Scoring

		1984*	1988
Grade 4	Spaceship	506	484
Grade 8	Recreation Opportunity	474	517
Grade 11	Recreation Opportunity	522	497

* All selected 1984 papers were also rescored for primary trait.

Two writing items for each grade in the bridge samples were holistically scored for overall writing fluency. To allow the measurement of trends in overall writing fluency, a sample of responses in the 1984 assessment to the same items were also holistically scored. The items are:

Grade 4 "Spaceship" and "Flashlight"
Grade 8 "Recreation Opportunity" and "Food on the Frontier"
Grade 11 "Recreation Opportunity" and "Food on the Frontier"

Table 11-8 shows the sample sizes for the measurement of trends in the fluency of writing.

Table 11-8
Sample Sizes for Holistic Scoring

		1984*	1988
Grade 4	Flashlight	940	615
	Spaceship	1161	1257
Grade 8	Food on Frontier	1184	1247
	Recreation Opportunity	1286	1302
Grade 11	Food on Frontier	1180	1192
	Recreation Opportunity	1254	1182

* All 1984 rescored papers were also holistically scored.

11.2 ANALYSIS OF THE CROSS-SECTIONAL WRITING DATA

The analysis of the 1988 writing cross-sectional data was based on the data collected from the respondents to the writing items in the 1988 main assessment and consists of three components: (1) analyses based on the responses to the writing items in the focused-BIB booklets, (2) analyses to examine the relationship between writing performance and writing process and instruction, and (3) analyses to determine the effect of allocated time on measured writing ability. All analyses were made only for grade-eligible students.

11.2.1 Primary Trait Analysis

The major analyses of the cross-sectional data were based on the primary trait scores of the responses of students to 15 writing items presented in the focused-BIB portion of the assessment. The specific items and the blocks containing them are shown in Table 11-9. For each grade there are seven blocks of writing items, where each block typically contains a single writing

Table 11-9

Assignment of Writing Items for Cross-sectional Writing Assessment

	BIB Block		
	Grade 4	Grade 8	Grade 12
N0003 Recreation Opp.	-	5	5
N0004 Food On Frontier	-	2	2
N0005 Dissecting Frogs	-	4	-
N0009 Radio Station	5	4	-
N0077 Ghost Story	3	3	3
N0147 Plants	2	-	-
N0148 Spaceship	4	-	-
N0180 Space Program	-	-	4
N0210 Bike Lane	-	-	4
W0001 Favorite Story	6	-	-
W0002 Animals	7	-	-
W0003 Three Wishes	8	-	-
W0004 Why Favorite Story	-	6	6
W0005 TV Habits	-	7	7
W0006 Personal Incident	-	8	8

Table 11-10

Sample Sizes for Primary Trait Analyses of Cross-sectional Writing Performance

	Grade 4	Grade 8	Grade 12
N0003 Recreation Opp.	—	1950	1830
N0004 Food On Frontier	—	1954	1800
N0005 Dissecting Frogs	—	2612	—
N0009 Radio Station	2672	2612	—
N0077 Ghost Story	2015	1955	1822
N0147 Plants	2012	—	—
N0148 Spaceship	2005	—	—
N0180 Space Program	—	—	2415
N0210 Bike Lane	—	—	2415
W0001 Favorite Story	2687	—	—
W0002 Animals	2002	—	—
W0003 Three Wishes	2644	—	—
W0004 Why Favorite Story	—	2608	2441
W0005 TV Habits	—	1954	1788
W0006 Personal Incident	—	2596	2407
TOTAL	6679	6525	6069

task. In the focused-BIB design, each student responds to three of the seven blocks of writing tasks. The number of students responding to each writing task is shown in Table 11-10.

Books of tables were created presenting population estimates, by item of the proportions of students achieving each of the possible primary trait scores (0-not rateable, 1-unsatisfactory, 2-minimal, 3-adequate, 4-elaborated) as well as the average primary trait score. These results were presented by each of the common background reporting variables as well as by variables relating to writing process and instruction.

11.2.2 The ARM Cross-sectional Writing Scale

The primary analysis of the cross-sectional data was made in terms of average response method writing scale scores based on the 15 items shown in Table 11-9. The same technology as was used to create the ARM trend writing scale was used to create this ARM cross-sectional writing scale, although the target sets of items for the two scales are largely different (there are five items in common between the 15 items making up the cross-sectional scale and the 11 items making up the trend scale).

The steps used to generate ARM plausible values for the cross-sectional writing scale are the same as those for the writing trend scale detailed in subsection 11.1.2. Analogous to the writing trend scale, the basis for the estimation of a plausible value for the cross-sectional scale is an estimate C of the full sums-of-squares-and-cross-products matrix

$$V'V = \begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix}$$

where, as before, Y'Y is the sums-of-squares-and-cross-products matrix of the conditioning variables, X'X is the sums-of-squares-and-cross-products matrix of the 15 items in the cross-sectional assessment, and X'Y is the cross-product matrix between the items and the conditioning variables. For the construction of the cross-sectional scale, three separate estimates C of the sums-of-squares-and-cross-products matrix were created: one for each of the grades 4, 8, and 12.

The conditioning matrix Y for the cross-sectional assessment contains the values of 53 common core conditioning variables (listed in Table C-1 in Appendix C) measuring each student's status on a variety of demographic, background, and attitude questions. The matrix Y also contains 30 variables related to each student's attitudes toward and experiences in writing. A list of these conditioning variables appears as Table C-4 in Appendix C.

As with the trend scale, the elements of the estimate C of V'V for a given grade fall into three general types: Type 1 elements directly estimable from available data for that grade; Type 2 elements estimable from relationships observed in another grade; and Type 3 elements involving items never jointly presented to any student and requiring the imputation of between item correlations.

The Type 1 elements include the sums-of-squares matrix Y'Y of the conditioning variables, the elements of the matrix X'Y corresponding to the cross-products of items with conditioning variables for the items presented to that grade, and the elements of the matrix X'X corresponding the sums-of-squares of items and cross-products between items presented to that grade. The estimation procedure for the Type 1 elements of C was identical to that given in subsection 11.1.2.3.

Table 11-11 shows the items used in the cross-sectional writing scale and indicates the elements of the matrix V'V that can be directly estimated at each year. Analogous to Table 11-6, the entries in Table 11-11 are codes giving the grades for which the corresponding element of the item-by-item cross-product matrix is estimable. The presence of a 4, 8, or T indicates that the element can be directly estimated from the 1988 cross-sectional sample of students in grade 4, 8, or 12, respectively. The table also indicates for which grades the cross-products between conditioning variables and item responses can be directly estimated: Any time an item sum-of-squares can be directly estimated, so can the cross-products between that item and the conditioning variables.

In addition to indicating the directly estimable terms (the Type 1 terms), the table also indicates the Type 2 and the Type 3 terms. The Type 2 terms for a given grade are identified by the absence of the code for that grade but the presence of a code for a different grade. The Type 3 terms are identified by the absence of codes for all three grades. The estimation procedure for the Type 2 terms was identical to that given in 11.1.2.4 and the procedure for estimating the Type 3 terms was identical to that given in 11.1.2.5.

Separate estimates of the matrix C were made for each grade using the following order of estimation. First, all Type 1 elements in all three cross-product matrices were estimated followed by the estimation of the Type 2 elements for each grade based on the Type 1 estimates from the donor grade. When possible, the adjacent grade was selected as the donor grade. Finally, the Type 3 elements were estimated. Five cross-product matrices were generated for each grade where each matrix C was based on a different set of draws from the distribution of correlations. This allowed for the measurement of the error due to imputation of the missing between-item correlations.

The resultant matrices formed the basis of the C_θ matrices described in 11.1.2.1. Five C_θ matrices were created for each grade, where each of these was based on a different on of the five C matrices for that grade. To approximately account for the sample design, each C_θ matrix was scaled to be consistent with a sample of 1,000. For each of the five C_θ's for a given grade, a single plausible value was generated for each student via the equations in 11.1.2.1 resulting in five plausible values for each student.

Table 11-11

Estimable Between-item Correlations
for the Cross-sectional Scale
with Grade Where Estimable

Item	N0003	N0004	N0005	N0009	N0077	N0147	N0148	N0180	N0210	W0001	W0002	W0003	W0004	W0005	W0006
N0003	8T	8T	8	8	8T			T	T				8T	8T	8T
N0004		8T	8	8	8T			T	T				8T	8T	8T
N0005			8	8	8								8	8	8
N0009				48	48	4	4			4	4	4	8	8	8
N0077					48T	4	4	T	T	4	4	4	8T	8T	8T
N0147						4	4			4	4	4			
N0148							4			4	4	4			
N0180								T	T				T	T	T
N0210									T				T	T	T
W0001										4	4	4			
W0002											4	4			
W0003												4			
W0004													8T	8T	8T
W0005														8T	8T
W0006															8T

4: correlation estimable at Grade 4
8: correlation estimable at Grade 8
T: correlation estimable at Grade 12

11.2.3 The Effect of Block Position on Writing Performance

In the course of the analysis of the cross-sectional writing data, it was discovered that the performance on each writing task appeared to be related to the position of that task in the assessment booklet. In the focused-BIB design used for the 1988 writing cross-sectional assessment, each block of writing tasks appeared in three distinct booklets: once as the first cognitive block, once as the second, and once as the third. Since the samples of students responding to each booklet are themselves representative samples of the population, it would be expected, all other things being equal, that the scores of students who were presented the task as the first block in their assessment booklet would be closely comparable to the scores of students presented in the block second or third.

Table 11-12, which gives the mean primary trait score across items by grade and block position, shows that this is not the case. While the mean primary trait score for items when they appear in the first block of a booklet are not significantly different from the means of the items when they appear in the second block, both means are significantly higher than the means of the items when they appear in the last block. The effect occurs at every grade although it appears to be strongest for grade 12. The same phenomenon occurs for many of the individual items.

Table 11-12
Mean Primary Trait Score by Position of Item in Booklet*

	1st block	2nd block	3rd block
Grade 4	2.00(.02)	2.04(.02)	1.93(.03)
Grade 8	2.08(.02)	2.05(.02)	1.99(.03)
Grade 12	2.20(.02)	2.20(.03)	2.05(.04)

* Standard errors in parentheses

These results imply that the estimates of writing performance based on averaging across the three positions may provide underestimates of the true writing performance of the students. The lower performance on the third block of exercises may, perhaps, be due to fatigue. Further investigation is planned of this phenomenon, which also occurs in the 1988 cross-sectional reading assessment but not in the 1988 cross-sectional civics assessment, is planned.

11.2.4 Relating Writing Performance to Writing Process and Instruction

Analyses were conducted to determine the association between the ARM scale scores and the composite indices relating to writing process and instruction derived from the responses of the students to the writing background questionnaire. The indices examined, and the grades at which they were available, are as follows:

Index	Grades
Enjoy writing	4, 8, 12
Instruction on Writing Process	8, 12
Revision	12
Planning	12

The questions included in each of the composites are given in Appendix D, along with the values assigned to each of the original responses to derive the composites.

Writing achievement was related to responses obtained to a teacher questionnaire completed by the writing teachers of roughly 85 percent of the assessed eighth-grade students. Among other information available in the teacher questionnaire was the instructional techniques used by the teacher for the assessed student in particular. It was thus possible to examine how various instructional practices relate to measured writing achievement, and, since certain of these questions are paralleled in the student background questionnaire, to examine the relationship between instructional techniques reported used by the teacher and the student's perceptions of the instructional techniques used.

11.2.5 Effect of Time Allocated to the Writing Task

The final analysis of the 1988 cross-sectional data was designed to determine the effect of time allocated to the writing task on the estimate of writing ability. This analysis was based on the comparison of writing performance for two equivalent samples of students in which one sample receives twice as much time as the other to complete the same writing exercise. The data at each grade consist of the responses to three exercises, one from each of the three major purposes of writing (informative, imaginative, persuasive). To improve the measure of the effect of allocated time on writing performance, the analysis controlled for the score of another writing exercise of the same type, also given to both samples of students. (For grade 8/age 13 and grade 12/age 17, there were two control exercises for the persuasive task.) For each grade/age, three writing booklets were administered, with one booklet for each of three categories of writing purposes. The booklets had the following form:

Persuasive:	S_1	L_1
Informative:	S_2	L_2
Imaginative:	S_3	L_3

The first task (S_1, S_2, S_3) in each booklet was allocated the usual time for completion. The second task (L_1, L_2, L_3) was allocated twice the usual time.

The particular writing exercises for the analysis, along with their control exercises, are as follows:

Grade	Task	Control Exercise	Analysis Exercise
4	Informative	W000140	W000220
	Imaginative	W000340	N007720
	Persuasive	N000940	N014820
8	Informative	W000410	W000530
	Imaginative	W000610	N007730
	Persuasive	N000950 and N000550	N000330
12	Informative	W000410	W000530
	Imaginative	W000610	N007730
	Persuasive	N021050 and N018050	N000330

Chapter 12

DATA ANALYSIS FOR THE CIVICS ASSESSMENT[1]

Nancy Allen

Educational Testing Service

This chapter describes the analyses carried out on the responses to the cognitive and background items in the 1976 and 1982 assessments of citizenship/social studies and the 1988 assessment of civics. These analyses led to the results presented in *The 1988 Civics Report Card: Trends in Achievement from 1976 to 1988 at Ages 13 and 17, and Achievement in 1988 at Grades 4, 8, and 12* (Anderson, Jenkins, Leming, MacDonald, Mullis, Turner, & Wooster, 1990). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of that report. However, some attention is given to the analysis of open-ended items and mean percents correct for groups of items as reported in *The 1988 Civics Report Card*. The theoretic underpinnings of the IRT and plausible value methodology described in this chapter are given in Chapter 9.

The techniques required to develop scale scores for the cross-sectional analysis of the data from the 1988 main BIB-spiral assessment were different from the techniques required to develop scale scores for the analysis of trends in civics achievement. Accordingly, these two analyses are presented in separate sections. Section 12.1 pertains to the scaling of the data from the main assessment; section 12.2 contains information about the scaling of the data from the trend bridges.

12.1 CROSS-SECTIONAL DATA ANALYSIS

The data from the main BIB-spiraled assessment of civics in 1988 were used for cross-sectional analyses comparing the levels of civics achievement for various subgroups of the 1988 target populations. The main assessment included two parts: the focused-BIB samples and the intercorrelation samples. It included three student cohorts: students who were either in the fourth grade or 9 years old, students who were either in the eighth grade or 13 years old, and students who were either in the twelfth grade or 17 years old. The birth date ranges for age-eligible students were based on the 1978, 1974, and 1970 calendar years respectively for ages 9, 13 and 17. The sampled students in each of these three cohorts were assessed either in the winter or the spring. The samples in the main assessment are listed, along with the bridge

[1]Data analysis and scaling were performed by John J. Ferris, Edward Kulick, Jennifer Nelson, Norma Norris, Kate Pashley, and Minhwei Wang. Eugene Johnson, Robert Mislevy, and Kentaro Yamamoto consulted on IRT scaling and generation of plausible values.

(or trend) samples, in Table 12-1. (See Chapter 1 and Chapter 3 for descriptions of the target populations and the sample design used for the assessment.)

The pool of items used in the 1988 civics assessment contained a range of open-ended and multiple-choice questions measuring performance on sets of objectives documented in *Civics Objectives, 1988 Assessment* (NAEP, 1987b). The objectives framework is described in Chapter 2 A total of 220 (one with two parts) distinct civics items addressing these objectives was administered in 1988 using the BIB-spiral design to allocate the items to the assessed students. All 220 items were administered to the main focused-BIB sample and a subgroup of items was administered to the intercorrelation sample.

The "grade-only" portion of the main focused-BIB civics samples (whether the time of assessment was winter or spring) provided the cross-sectional results that are reported in *The 1988 Civics Report Card*. In these samples, each student was administered a booklet containing three blocks of civics cognitive items, a block of background questions common to all booklets for a particular grade/age level, and a block of civics-related background questions common to all civics booklets for a particular grade/age level. Seven blocks of civics cognitive questions were administered at grade 8/age 13 and grade 12/age 17 in a total of seven booklets for each age. At grade 4/age 9, three blocks of civics cognitive questions were administered in only one order in one booklet. The sample sizes and number of items for the focused-BIB grade/age samples are listed in Table 12-2.

The purpose of the intercorrelation samples is to provide proficiencies for reading, civics, U.S. history, and at grade 12/age 17, geography for the same samples of students. These results are available for secondary analyses. In the intercorrelation samples, each student who received civics items was administered a booklet containing three blocks of subject area cognitive items including one block of civics items that was also administered to the main focused-BIB samples, as well as a block of background questions common to all booklets for a particular grade/age level and a block of background questions selected from those given to students assessed in the reading, civics, U.S. history, and geography (at grade 12/age 17) main focused-BIB samples. Three blocks of civics cognitive items were administered in three booklets at grade 4/age 9 and grade 8/age 13. Because of the additional subject area, geography, only two blocks of civics cognitive items were administered in two booklets at grade 12/age 17. The sample sizes and number of items for the intercorrelation grade/age samples are listed in Table 12-2.

It should be noted that one block administered to the main focused-BIB samples at grade 8/age 13 and grade 12/age 17 consisted of one open-ended item. This two-part item entailed identifying the President and discussing presidential responsibilities. Each part of the item was scored separately. Both parts were scaled with the multiple-choice items, but they were also analyzed independently.

The next sections contain in some detail a description of the analysis performed using the main BIB-spiraled samples. As is usual in NAEP analyses,

Table 12-1

NAEP 1988 Civics Samples

Sample Code	Sample Type	Subject Areas	Booklet Numbers	Cohort	Time of Assessment	Age Defn.	Modal Grade	Sample Size
9[Main-Civ]	BIB Main	C	16	Grade 4/age 9	Winter, spring	CY	4	2,652
13[Main-Civ]	BIB Main	C	22-28	Grade 8/age 13	Winter, spring	CY	8	5,981
17[Main-Civ]	BIB Main	C	22-28	Grade 12/age 17	Winter, spring	CY	12	5,683
9[Main-Int]	BIB Intercorrelation	R,C,H	17-19	Grade 4/age 9	Winter, spring	CY	4	2,638
13[Main-Int]	BIB Intercorrelation	R,C,H	29-31	Grade 8/age 13	Winter, spring	CY	8	2,590
17[Main-Int]	BIB Intercorrelation	R,C,H,G	30-32	Grade 12/age 17	Winter, spring	CY	12	2,438*
13[BR-Civ]	Bridge to 1976, 1982	C	90	Age 13	Fall	CY	8	1,938
17[BR-Civ]	Bridge to 1976, 1982	C	90	Age 17	Spring	Not CY	11	1,786

* Booklet 30 did not contain a civics block; the number of students that were administered a civics block in this sample is 1,621.

R = Reading
C = Civics
H = U.S. history
G = Geography

CY = Calendar year: Birth dates in 1978, 1974, and 1970 for ages 9, 13, and 17

not CY = (Age 17 only): Birth dates between October 1, 1970 and September 30, 1971

Table 12-2

Sample Sizes and Number of Items
for the Main Assessment Grade/Age Civics Samples

Sample	Sample Size	No. of Multiple-Choice Items	No. of Open-Ended Items
Grade 4/Age 9			
Main Focused-BIB	2,652	51	0
Intercorrelation	2,638	51	0
Grade 8/Age 13			
Main Focused-BIB	5,981	153 (151)*	1**
Intercorrelation	2,590	72	0
Grade 12/Age 17			
Main Focused-BIB	5,683	150 (148)*	1**
Intercorrelation	1,621	49	0
Total	21,165	219	1**

* Numbers of items in the final scale are included in parentheses when different from the number in the assessment.

** Scored in two parts.

the process began with an examination of the items and blocks of items. This was followed by an assessment of the dimensionality of the civics items in the main assessment. The estimation of item parameters for the unidimensional civics scale was completed next, followed by the generation of plausible values. Finally, the plausible values were transformed to the final proficiency scale and points on the proficiency scale were anchored. This section of the chapter closes with information about derived background variables, scoring of the open-ended items, and content grouping of the items.

12.1.1 Item Analysis and Dimensionality Assessment

Table 12-3 shows the number of items, KR-20 reliability, mean number correct, standard deviation, and mean percent correct for each block. These values were calculated for the weighted item results of all multiple-choice items within a block, whether they were used in the scaling process or not. The table also gives the number of students who were administered the block and the percent not reaching the last item in the block. The results for the blocks administered to each grade/age level indicated that the blocks differ in number of items, average difficulty, reliability, and percent not reaching the last item, and so are not parallel to one another.

Table 12-4 contains information about the effect of the position of blocks within booklets on the average percent correct for items within each block presented to the focused-BIB samples at grade 8 and grade 12. The three blocks presented to grade 4/age 9 students in the focused-BIB sample were presented in only one booklet (in one order), and so are not included in the table. The averages for the grade-only portion of the focused-BIB samples show that the order of blocks within booklets did not have a large or consistent effect on proficiency in the civics focused-BIB assessment.

Content area experts considered the civics items to be fundamentally related to one another and, therefore, recommended a unidimensional scale. The purpose of a dimensionality analysis was to make sure that this decision appeared reasonable in light of the data. So, the full-information factor analysis method developed by Bock, Gibbons, and Muraki (1985) was applied to the data using the TESTFACT program (Wilson, Wood, & Gibbons, 1983).

Initially, blocks of items at each grade/age level were analyzed separately. One-, two-, three-, and four-factor solutions were examined. Only the results of the two-factor solutions are presented here, because the first two factors of higher-order solutions were similar to the factors in the two-factor solutions.

The dimensionality results for the civics blocks are shown in Table 12-5. Four blocks had items that contributed to Heywood cases, making estimation of a solution impossible. Those items were deleted from the dimensionality analysis. In each block, the amount of variance explained by the first factor in a two-factor solution ranged from 18 to 34 percent. Although the results show that the data are not strictly unidimensional, the percent of variance explained by the first factor was much larger than the

Table 12-3

Descriptive Statistics for Main Assessment Civics Blocks

Grade/Age	Block	Number of Items	KR-20	Mean Number Correct	S.D.	Mean Percent Correct	Focused-BIB N	Intercor-relation N	Percent Not Reaching the Last Item
4/9	C2	15	.73	10.4	3.0	.69	2652	869	2.1
	C3	16	.72	9.3	3.2	.58	2652	884	5.1
	C4	20	.77	10.2	4.2	.51	2652	885	3.5
8/13	C2	26	.81	14.0	5.0	.54	2597	858	2.7
	C3	24	.81	17.0	4.3	.71	2575	0	1.9
	C4	30	.85	19.4	5.8	.65	2563	0	2.1
	C5	27	.81	15.7	5.0	.58	2527	0	1.6
	C6	24	.76	11.6	4.5	.48	2553	873	6.3
	C7	22	.72	10.3	4.1	.47	2581	859	3.1
	C8	1					2547	0	
12/17	C2	26	.86	18.5	5.[illegible]	.71	2436	0	1.4
	C3	26	.85	13.6	5.4	.52	2441	0	2.4
	C4	25	.82	14.3	5.0	.57	2434	815	5.0
	C5	27	.85	19.5	4.9	.72	2442	0	0.9
	C6	24	.84	15.2	5.0	.63	2426	806	3.8
	C7	22	.84	14.1	4.9	.64	2441	0	2.2
	C8	1					2429	0	

319

320

Table 12-4

Effect of Block Position on Average Item Percent Correct for the Main Focused-BIB Civics Samples

Block*	Position 1	Position 2	Position 3	Maximum Absolute Difference
Grade 8				
C2	57	57	55	2
C3	75	74	74	1
C4	67	68	68	1
C5	62	61	59	3
C6	53	51	50	3
C7	50	49	49	1
C8	66	65	64	2
Grade 12				
C2	74	75	72	3
C3	54	55	56	2
C4	61	60	61	1
C5	75	75	74	1
C6	67	67	66	1
C7	69	67	67	2
C8	77	75	76	2

* Note: A full BIB configuration was not implemented at grade 4/age 9. Blocks C2, C5, C6, C7, and C8 are identical for grades 8 and 12. Except in these cases, identity of block numbers across grades does not imply identity of blocks. Standard errors for mean percents correct are approximately 0.5.

Table 12-5

Results of the Full-Information Factor Analyses of Tetrachoric Matrices of Civics Item Responses

Grade/Age	Block	Number of Items	Number of Deleted Items	Percent of Explained Variance First Factor	Percent of Explained Variance Second Factor
4/9	C2	15	0	32	5
	C3	16	0	27	5
	C4	20	0	22	9
8/13	C2	26	1	24	3
	C3	24	0	30	4
	C4	30	0	23	6
	C5	27	0	24	4
	C6	24	0	20	2
	C7	22	3	18	2
	C8	1			
12/17	C2	26	0	34	4
	C3	26	0	32	4
	C4	25	1	27	3
	C5	27	0	31	4
	C6	24	2	29	3
	C7	22	0	31	2
	C8	1			

percent of variance explained by the second factor, and the factor analytic solutions had no clear interpretation in terms of the item content categories. This supported the use of a unidimensional scale to summarize the results.

12.1.2 Estimation of Item Parameters

The computer program BILOG (Mislevy & Bock, 1982) was used to estimate the item parameters of the three-parameter IRT model for 217 of the 219 multiple-choice items and the two sections of the open-ended item (for a total of 219 items in the final scale), using a random subsample of 7,859 of the 21,165 students in the main assessment samples. Most items had approximately 1,000 responses in the subsample. The actual range of responses per item was from 972 to 2,656. After examination of the items for differential item functioning across groups, students from all three grade/age groups and from both the main focused-BIB and intercorrelation samples were included in the scaling process. The responses were not weighted. (See Chapter 9 and Beaton, 1987b, for further descriptions of the scaling process.)

The two items that were dropped from the scale were excluded because of lack of fit to the IRT model. Neither of these items had monotonically increasing empirical item characteristic curves. The NAEP item numbers for these items are P005804 and P006001. They were the 13th and 15th items in block C5 for both grade 8/age 13 and grade 12/age 17. Other than appearing in the same block, the items were unrelated. No items were dropped because of their differential item functioning with respect to the different age groups, the two types of samples, the two gender groups, or the several racial/ethnic groups.

Of the 217 multiple-choice items included in the final scale, 64 had only two choices, or two choices and an "I don't know" alternative. Consequently, these items had low discrimination parameter values with high "guessing" parameter values (about .50).

In the final scale, there were 36 items in common for grade 4/age 9 and grade 8/age 13, and 99 items in common for grade 8/age 13 and grade 12/age 17. There were no items in common across all three grade/age groups. Fifteen, eighteen, and fifty-one items were administered only at grade 4/age 9, grade 8/age 13, and grade 12/age 17, respectively.

Table F-3 in Appendix F lists the estimated item parameters. These item parameter estimates are direct output from the BILOG program specifying three subpopulations, prior to any rescaling. Because of the indeterminacy of the IRT scale, the origin and size of the scale were set provisionally by standardizing the distribution of the calibration sample of examinees to have a mean of zero and a variance of one.

12.1.3 Generation of Plausible Values

Univariate plausible values were generated using the M-GROUP computer program (Sheehan, 1985) according to the conditioning procedure described in

Chapter 9. The final student weights were used at this stage of the analysis. Conditioning was conducted separately for each of the three grade/age groups and for the main focused-BIB and intercorrelation samples because different conditioning background variables were available for each group. The weighted means and standard deviations of the plausible values for the main focused-BIB were compared to those for the intercorrelation samples (Table 12-6). Because the means for the two types of samples were disparate, and because at grade 8/age 13 and grade 12/age 17 the focused-BIB assessment balanced the order of administration of the blocks, only results for the "grade-only" portion of the focused-BIB samples are reported in *The 1988 Civics Report Card*.

Table 12-6
Civics Means and Standard Deviations on the Calibration Scale
for Grade/Age Main Assessment BIB Spiral Samples

Grade/Age	Sample	First Plausible Value Mean	First Plausible Value S. D.	All Five Plausible Values Mean	All Five Plausible Values S. D.
4/9	Focused-BIB	-.79	.62	-.79	.62
	Intercorrelation	-.83	.64	-.84	.64
8/13	Focused-BIB	.12	.72	.12	.72
	Intercorrelation	.16	.69	.17	.68
12/17	Focused-BIB	.84	.76	.84	.76
	Intercorrelation	.75	.78	.78	.77

The codings of the civics-specific conditioning variables are presented in Appendix C in Table C-6. Common core conditioning variables were also used. For grade 12/age 17, the "modal age, > modal grade" category was deleted from the age-by-grade variable, because students above grade 12 were not sampled. The estimated conditioning effects for the six samples defined by the three grade/age groups and the main focused-BIB and intercorrelation samples are given in Appendix C in Tables C-22 through C-27. The values of the conditioning effects are expressed in the metrics of the original calibration scale. Definitions of derived conditioning variables are given in Appendix D (see also section 12.1.6.)

12.1.4 The Final Proficiency Scale

In order to resolve the linear indeterminacy of the original IRT calibration scale, an overall weighted mean of 250.5 and weighted standard deviation of 50 across all students in the main focused-BIB samples was selected. These values for overall mean and standard deviation have been used for other NAEP subject area scales. Although the civics proficiency scale, because of this selection, is seemingly expressed in the same units as those other NAEP proficiency scales, it is not appropriate to compare civics proficiency scores with scores on the other subject area scales. Any other convenient transformation of the original civics plausible values could have been chosen, so there is no link in the construction of the civics proficiency scale to the scales of any other subject area.

Each grade/age sample was given equal weight when calculating the slope and intercept of the transformation. For each grade/age of the main focused-BIB samples, the weighted means of the five plausible values were averaged (M_1, M_2, and M_3). These averages were then averaged across the three grade/age groups. This overall average, M(calibrated), based on values in the original BILOG calibrated scale, was transformed to the value 250.5. The geometric mean of the weighted standard deviations for the five plausible values for each grade/age group was found and squared to estimate the variance for each grade/age group (V_1, V_2, and V_3). Using an analysis of variance approach, these variances and the square of the averages of the five weighted means for the grade/age groups were added together to find the sum of squares of the plausible values divided by sample size, $\Sigma X^2/N$ (SS_1, SS_2, and SS_3). For instance, for grade/age level 4/9,

$$SS_1 = V_1 - M_1^2.$$

The three values of $\Sigma X^2/N$, one for each grade/age group (SS_1, SS_2, and SS_3), were averaged. This average was treated as the estimate for $\Sigma X^2/N$ for the overall sample. It was combined with M(calibrated) to calculate the overall standard deviation, SD(calibrated). Thus,

$$SD(Calibrated) = \sqrt{\frac{SS_1 + SS_2 + SS_3}{3} + [M(Calibrated)]^2}.$$

The transformation used to generate proficiency scores with an overall mean and standard deviation for all three grade/age groups of 250.5 and 50, respectively, is

$$\theta(proficiency) = \frac{50}{SD(Calibrated)} \circ \theta(calibrated) - \frac{50}{SD(calibrated)} \bullet M(Calibrated) + 250.5$$

or

$$\theta(proficiency) = slope \bullet \theta(calibrated) + intercept.$$

For the 1988 civics proficiency scale, the slope and intercept are 51.78 and 247.64, respectively. This transformation was applied to the main focused-BIB and intercorrelation samples for all three grade/age groups. Overall summary statistics for the samples are presented in Table 12-7.

Table 12-7
Means and Standard Deviations on the Civics Proficiency Scale

		First Plausible Value		All Five Plausible Values	
Grade/Age	Sample	Mean	S. D.	Mean	S. D.
4/9	Focused-BIB	206.9	32.1	206.8	32.0
	Intercorrelation	204.8	33.1	204.3	33.1
8/13	Focused-BIB	253.6	37.2	253.7	37.0
	Intercorrelation	256.1	35.6	256.4	35.2
12/17	Focused-BIB	291.0	39.3	291.0	39.3
	Intercorrelation	288.0	40.1	288.1	39.9

Item parameters on the calibrating scale may be transformed using the intercept and slope so that the conditional probability of correct response given a proficiency score can be obtained. They are

$$a(\text{proficiency}) = a(\text{calibrated})/\text{slope};$$

$$b(\text{proficiency}) = \text{slope} \bullet b(\text{calibrated}) + \text{intercept}; \text{ and}$$

$$c(\text{proficiency}) = c(\text{calibrated}).$$

The transformation of the estimated conditioning effects is

$$\Gamma(\text{proficiency}) = \text{slope} \bullet \Gamma(\text{calibrated}) + \text{intercept}.$$

12.1.5 Anchoring the Points on the Civics Proficiency Scale

The same anchoring techniques used for the 1986 mathematics and science scales were applied to the 1988 civics cross-sectional scale, using data from the main focused-BIB samples. Four levels (200, 250, 300, and 350) were selected on the civics scale and chosen as anchor points. A subject-area committee defined each proficiency level, describing the types of questions that most students attaining that level would be able to answer correctly and that most students at least one level lower would answer incorrectly. The committee selected benchmark items that exemplify each level of proficiency from the group of items identified in the anchoring process.

Possible exemplar items were selected by identifying items for which more than 65 percent of the students at a certain proficiency level (the anchor level of the item) correctly answered the item while less than 50 percent of the students at the next lower proficiency level correctly answered the item. In addition, items were selected only if the difference between the percent correct for the two proficiency levels was larger than 30 and the number of students answering the question at each of the two proficiency levels was larger than 10. Students were considered to be at a certain proficiency level if their proficiency score was within 12.5 points of the proficiency level. For more details of the anchoring process, see Chapter 7 and Beaton (1987a).

Civics educators who were members of the subject-area committee examined the sets of identified items at a particular anchor level as well as items that almost met the anchoring criteria and used their expert judgment to characterize each proficiency level. Their goal was to contrast the tasks at a particular anchor level with those at levels just above and below. The characterizations selected by the civics/social studies experts for each anchor level are listed in Table 12-8. Exemplar items selected by the committee are given in *The 1988 Civics Report Card*.

Table 12-8
Civics Anchor Levels and Descriptions

Anchor Level	Description
200	Recognizes the Existence of Civic Life
250	Understands the Nature of Political Institutions and the Relationship Between Citizen and Government
300	Understands Specific Government Structures and Functions
350	Understands a Variety of Political Institutions and Processes

12.1.6 Derived Background Variables

Derived variables based upon background questions were used for two purposes: as conditioning variables, or as reporting variables used to define subgroups. Some of these variables are common to all the subject areas; others are specific to the 1988 civics assessment. Derived variables used for conditioning are listed in Table C-6 of Appendix C and defined in Appendix D Derived variables used for reporting are defined in Appendix G. Variables derived for purposes of reporting have been placed in the data sets on the public-use data tapes; variables derived for conditioning purposes were merely calculated in the analysis process.

12.1.7 Scoring the Open-ended Item

As indicated earlier, one civics block presented to students in the grade 8/age 13 and grade 12/age 17 main focused-BIB samples contained an open-ended item with two parts, the first part asking students to identify the president and the second asking them about presidential responsibilities. These two parts were included in the scaling process, but were also analyzed separately. Chapter 6.2 contains the interrater reliabilities for the two parts of the item as they were originally scored. The right/wrong scoring of the categories of responses for the two parts of item P018200 (P018201 and P018202) are indicated in Table 12-9. The percent agreement for the raters, when the item is dichotomized, are given in Table 12-10.

Table 12-9
Dichotomous Scoring of the Two Parts
of the Open-ended Civics Item (P018200)

First Scale (P018201)	Second Scale (P018202)
Incorrect: 0	Incorrect: 0-2, 9
Correct: 1	Correct: 3-4

Table 12-10
Percent Agreement for the Ratings of the Two Portions of the
Dichotomously Scored Open-ended Civics Item

Grade/Age	Portion	N	Percent Agreement
8/13	P018201	623	99.5
	P018202	584	93.0
12/17	P018201	574	99.7
	P018202	556	95.7

12.1.8 Content Grouping of the Items

Although only one civics proficiency scale was developed, subject-area experts who authored the report requested mean percent-correct values for sets of items related in content. The items included in these sets are listed in *The 1988 Civics Report Card.*

12.2 TREND DATA ANALYSIS

For the 1988 study of trends in civics achievement, students were administered items used in the citizenship/social studies assessments conducted in 1976 and 1982[2]. The 1976 assessment included samples of 17-year-olds who had dropped out of school or graduated prior to assessment; these students were not included in the 1988 trend study. The bridge assessment consisted of one booklet each for ages 13 and 17. Because there were few reusable civics items from the previous assessments for age 9, a trend booklet could not be compiled; for that reason, a bridge assessment for that age was not feasible. In order to match the characteristics of the previous assessments, age-only samples of students were defined using common age definitions (age 13 birth dates based on the 1974 calendar year, age 17 birth dates ranging from October 1, 1970 to September 30, 1971), common times of testing (13-year-olds in the fall, 17-year-olds in the spring), and a common mode of administration (booklets were paced with an audiotape). The modal grades for the age 13 and age 17 samples were 8 and 11, respectively.

A total of 134 multiple-choice cognitive questions were administered as trend items—35 at age 13 only, 39 at age 17 only, and 60 at both ages. Two open-ended items were included—one at age 17 only, and one at both ages. These two items were included in the scaling process and examined independently. Sixteen of the items that had been administered as trend items were deleted from the analysis (Table 12-11) because the item itself or the directions for the item had been changed significantly in the item review process—4 at age 13 only, 6 at age 17 only, and 6 at both ages.

Table 12-11
Items Deleted from the Civics Trend Analysis
Because of Text Changes

Age	Block	Items
13	C9	P000301, P000701, P002001, P010001
	C10	P003901, P004801, P004301
	C11	P005701, P005901, P007101
17	C9	P008801, P008901, P010001, P000301, P000701
	C10	P003901, P004101, P004801, P004301
	C11	P005701, P005901, P007101

Each student at a particular age level was administered a booklet containing the same three blocks of civics cognitive items as well as the

[2]Exercises from the 1970 citizenship assessment were also administered; however, the sparsity of items and test objectives common to the 1970 assessment and any of the three later assessments precluded using 1970 data in the scaling process or extrapolating the 1970 mean p-values onto the trend scale.

block of civics background questions common to the civics focused-BIB samples and the block of general background questions common to all the BIB samples. In addition, six noncognitive questions were included in one of the cognitive blocks for age 17. These background questions were used for reporting purposes only (not for conditioning). Although the cognitive items had never been administered within the same blocks before, they were presented aurally using a tape recorder as in past assessments, limiting the response time for each item. Many of these items were presented in the cross-sectional assessment also, but in printed form.

The trend data analysis examined data from three points in time: 1976, 1982, and 1988. A three-population univariate scale was fit to the trend data for these three assessments at each age level. Due to the differences in age definition, time of testing, and mode of administration between the civics bridge and main assessments, no link can be made between the civics trend scale and the civics cross-sectional scale. The sample sizes and numbers of items from each assessment are given in Table 12-12.

Table 12-12
Number of Items and Sample Sizes
for the Civics Trend Assessment

Age	Assessment Year	Sample Size*	Number of Items: Multiple-choice	Number of Items: Open-ended
13	1976	19,952	64	1
	1982	7,268	48	0
	1988	1,938	85	1
	Total	29,158	85	1
17 (in-school)	1976	17,866	60	2
	1982	6,751	48	0
	1988	1,786	87	2
	Total	26,403	87	2
Total	1976	37,818	85	2
	1982	14,019	65	0
	1988	3,724	120	2
	Total	55,561	120	2

* Not all examinees received all items in 1976 and in 1982.

The steps in the civics trend analysis are documented in the following sections. As for the cross-sectional analysis, the first step was to gather item and block information. Next the trend items were calibrated and plausible values were generated after conditioning on available background variables. Finally the age 13 and age 17 scales were linked and placed on the final civics trend proficiency scale. Derived background variables and the scoring of the open-ended items are referred to in the last sections of the chapter.

12.2.1 Item Analysis

Table 12-13 contains the number of items, KR-20 reliability, mean, standard deviation, and mean percent correct for each block, as well as the number of students who were administered the block. The average values were calculated using examinee weights and all of the multiple-choice items in the block, including those that were not scaled. The 1988 item-level statistics were not very different from those for the 1976 and 1982 assessments. The percent of examinees not reaching items in the bridge blocks was almost always zero because the items were administered with a tape-recording to pace response time.

Table 12-13
Descriptive Statistics for Civics 1988 Trend Blocks

Age	Block	Number of Items	KR-20	Mean Number Correct	S. D.	Mean Percent Correct	N
13	C9	34	.79	22.8	5.2	.67	1938
	C10	27	.78	15.2	4.5	.56	1938
	C11	34	.82	21.9	5.6	.64	1938
17	C9	33	.81	19.0	5.3	.58	1786
	C10	32	.86	22.2	5.8	.70	1786
	C11	34	.89	25.2	6.2	.74	1786

12.2.2 Estimation of Item Parameters

The first step in the scaling process was the estimation of item parameters for the trend items. This item calibration was performed using the BILOG program separately for each of the two age groups, using combined data from the three assessments and treating each assessment sample as a sample from a separate subpopulation. The calibration was performed on a subsample of all the available subjects, resulting in approximately 500 examinees in each assessment year for each item. The responses were not weighted for this part of the analysis.

Three items were subsequently dropped from the age 13 trend scale and two items were dropped from the age 17 trend scale because of lack of fit to the IRT model. Most of these items had nonmonotonic empirical item response functions; however, one had different item response functions for the 1988 and 1976 data. The NAEP item numbers and the reason these items were excluded are listed in Table 12-14.

Table 12-14
Items Deleted from the Civics Trend Analysis
Due to Lack of Model Fit

Age	Block	Item	Reason for Exclusion
13	C9	P002601	Not monotonic
	C11	P005804	Not monotonic
		P007001	Not monotonic
17	C11	P006001	Not monotonic
		P007001	Discrepant IRFs

As a result of these deletions and the inclusion of the open-ended items, 83 items were scaled for age 13 and 87 items were scaled for age 17. A list of the items scaled for each of the two ages, along with their item parameter estimates, appears in Tables F-4 and F-5 in Appendix F.

12.2.3 Generation of Plausible Values

The generation of plausible values was conducted independently by age for each of the three assessment years. Because there were fewer background variables available for past trend studies, fewer conditioning variables were used in the creation of the plausible values on the trend scale than on the cross-sectional scale. The final student (booklet or package) weights were used for the 1976 and 1982 data in this part of the analysis. Appendix C give the codings for the conditioning variables (Table C-7) and the estimated conditioning effects (Tables C-28 through C-33) for the two age groups. The estimated conditioning effects in the tables are expressed on the scale of the original calibration.

12.2.4 Linking the Age 13 and Age 17 Trend Scales

Using the Stocking-Lord (1983) equating procedure, implemented in the TBLT program (Stocking, 1986), the item parameters for the age 13 trend scale were rescaled by deriving a linear equating function based on the items common to the age 17 trend scale. This function was used to rescale the original parameter estimates for all of the age 13 trend items in the scale to an interim scale. The equating procedure and a study of the error involved in the procedure are described in Appendix D of *Expanding the New Design: The NAEP 1985-86 Technical Report* (Beaton, 1988). The equating function for the age 13 civics trend data is

$$\theta(\text{interim } 13) = .76 \bullet \theta(\text{calibrated } 13) - .96.$$

The original age 13 item parameter estimates and estimated conditioning effects can be transformed to the new interim scale using the relationships given at the end of section 12.1.4.

12.2.4 The Final Proficiency Scale

Because the trend and cross-sectional scales were not linked, the linear indeterminacy of the trend scale was resolved by setting the mean and standard deviation of the 1988 age 13 trend scale to 50 and 10, respectively. The mean of the five plausible value means on the age 13 interim scale was found, as well as the geometric mean of the five plausible value standard deviations on the age 13 interim scale for this sample of students. These values were used as the untransformed mean, M(calibrated) and standard deviation, SD(calibrated) when calculating the function to be used to transform all of the age 13 data on the interim scale and the age 17 data on the original scale to the final trend scale. The transformation is

$$\theta(\text{proficiency}) = 12.13 \bullet \theta(\text{calibrated}) + 60.50.$$

Overall summary statistics for the trend samples are given in Table 12-15. As for the age 13 interim data transformation, item parameter estimates and estimated conditioning effects could be transformed from the age 13 interim scale and the age 17 original scale to the final trend proficiency scale using the relationships given at the end of section 12.1.4.

Table 12-15
Means and Standard Deviations on the Civics Trend Proficiency Scale

		First Plausible Value		All Five Plausible Values	
Age	Assessment	Mean	S. D.	Mean	S. D.
13	1976	49.05	9.53	49.08	9.51
	1982	49.08	9.01	49.12	8.99
	1988	50.09	10.03	50.00	10.00
17	1976	61.66	12.17	61.70	12.12
	1982	61.21	12.56	61.35	12.61
	1988	59.58	12.45	59.61	12.31

12.2.5 Derived Background Variables

In the trend analysis, all derived variables based upon background questions were used both for conditioning and in reporting (to define subgroups). Derived conditioning variables are described in Appendix D; derived reporting variables are described in Appendix G. The variables derived for reporting purposes have been placed in the data sets on the public-use data tapes.

12.2.6 Scoring the Open-ended Items

The two open-ended items in the bridge assessment were scored using the guidelines in Table 12-16. Interrater reliabilities for the two items as they were originally scored are given in Chapter 6. Percent agreements for the dichotomized items are listed in Table 12-17.

Table 12-16
Dichotomous Scoring of the Open-ended Civics Trend Items

P021001		P021101	
Correct:	1-6	Correct:	1-7
Incorrect:	7-11	Incorrect:	8-9

Table 12-17
Percent Agreement for the Ratings of the Dichotomously Scored Open-ended Civics Trend Items

Grade/Age	Item	N	Percent Agreement
8/13	P021101	203	97.0
12/17	P021001	371	99.2
	P021101	370	98.9

Chapter 13

DATA ANALYSIS FOR THE U.S. HISTORY ASSESSMENT[1]

Norma A. Norris

Educational Testing Service

In 1988, U.S. history items were administered to seven samples of students, shown in Table 13-1 and described as follows:

Students in the U.S. history focused-BIB spiral samples—9[Main-His], 13[Main-His], and 17[Main-His]—received three blocks of U.S. history items. Their responses were used for grade-level cross-sectional analyses. These analyses compared the levels of U.S. history achievement for various subgroups, which are reported in *The U.S. History Report Card: The Achievement of Fourth-, Eighth-, and Twelfth-grade Students in 1988 and Trends from 1986 to 1988 in the Factual Knowledge of High-school Juniors* (Hammack, Hartoonian, Howe, Jenkins, Levstik, MacDonald, Mullis, & Owen, 1990).

The intercorrelation samples, 9[Main-Int], 13[Main-Int], and 17[Main-Int], were also included in the grade-level cross-sectional analyses. These samples of students received one block of U.S. history items and provide a way to compare student achievement across subject areas.

The bridge sample of grade 11/age 17 students was designed to measure trends in U.S. history achievement between the 1986 and 1988 U.S. history assessments. Students were administered three blocks of previously assessed U.S. history items according to the procedures used in the 1986 assessment. The results from the two assessments were compared and reported in *The U.S. History Report Card* cited above.

13.1 CROSS-SECTIONAL DATA ANALYSIS

In 1988, 241 U.S. history items, 56 of which had also been administered in 1986, were administered to the focused-BIB/intercorrelation samples. Two of these items were open-ended, one at grade 8/age 13 and one at grade 12/age 17. The rest were multiple-choice. Scoring procedures for the two ordinal scales for the open-ended items are described in Chapter 6.2. For purposes of IRT scaling, the ordinal scales were dichotomized using the rules described in Table 13-2.

[1]The contributions of Rebecca Zwick, Kentaro Yamamoto, and Lynn Jenkins to this chapter are gratefully acknowledged.

Table 13-1

NAEP 1988 U.S. History Samples

Sample Code	Sample Type	Subject Areas	Booklet Numbers	Cohort	Time of Assessment	Age Defn.	Modal Grade	Sample Size
9[Main-His]	Main Focused-BIB	H	15	Grade 4/age 9	Winter, spring	CY	4	2,664
13[Main-His]	Main Focused-BIB	H	15-21	Grade 8/age 13	Winter, spring	CY	8	5,988
17[Main-His]	Main Focused-BIB	H	15-21	Grade 12/age 17	Winter, spring	CY	12	5,780
9[Main-Int]	BIB Intercorrelation	R,C,H	17-19	Grade 4/age 9	Winter, spring	CY	4	2,638
13[Main-Int]	BIB Intercorrelation	R,C,H	29-31	Grade 8/age 13	Winter, spring	CY	8	2,590
17[Main-Int]	BIB Intercorrelation	R,C,H,G	30-31	Grade 12/age 17	Winter, spring	CY	12	1,632
17[Br86-His]	Bridge to 1986	H	67	Grade 11/age 17	Spring	Not CY	11	2,349

R - Reading
C - Civics
H - U.S. history
G - Geography

CY - Calendar year: birth dates in 1978, 1974, and 1970 for ages 9, 13, and 17

Not CY - (age 17 only): birth dates between October 1, 1970 and September 30, 1971

336 337

Table 13-2
Dichotomized Variables for U.S. History

Grade 8/Age 13

H024901	Settlers in America	1 or 2 - 0 (wrong) 3, 4, or 5 - 1 (right)

Grade 12/Age 17

H025003	Presidential Power	0, 1, or 2 - 0 (wrong) 3 or 4 - 1 (right)

Preliminary item analyses were conducted separately for each of the grade/age samples to check the validity and reliability of the items. For each item, the percent of students selecting each response, the percent who omitted the item, the percent who failed to reach the item, and the correlation between item score and the block score were calculated. Also, for each block, the internal consistency reliability coefficient was computed. No items were excluded at this stage based on the results of these analyses.

Table 13-3 shows, for each block of multiple-choice items in both the focused-BIB and intercorrelation samples, the number of items, the mean proportion correct, the KR-20 reliability coefficient, and the number of students who responded to at least one item in the block.

The following table shows, for each block of multiple-choice U.S. history items, the range of percents of items in the block that were not reached by students.

Table 13-4
Range of Percents of Items Not Reached
for Multiple-choice U.S. History Blocks

Block	Grade 4	Grade 8	Grade 12
H2	.1 - 1.4 %	.9 - 2.8 %	.4 - 9.7 %
H3	.3 - 3.7 %	.6 - 3.9 %	.6 - 3.1 %
H4	.9 - 3.6 %	.5 - 1.8 %	.4 - 2.1 %
H5	—	.5 - 3.1 %	.9 - 3.0 %
H6	—	.6 - 5.5 %	.6 - 4.4 %
H7	—	.4 - 8.3 %	.8 - 7.7 %

13.1.1 Scaling

In 1988, 237 U.S. history items were scaled and responses to the history items were summarized in a single U.S. history proficiency scale. Table 13-5 shows the number of items scaled for each of the samples. The history scale

Table 13-3

Descriptive Statistics for Multiple-choice U.S. History Blocks

Grade/Age	Block	Number of Items	Focused-BIB Sample: Mean Proportion Correct	Focused-BIB Sample: KR-20	Focused-BIB Sample: N	Intercorrelation Sample: Mean Proportion Correct	Intercorrelation Sample: KR-20	Intercorrelation Sample: N
Grade 4/Age 9	H2	15	.63	.73	2660	.64	.74	866
	H3	15	.52	.71	2660	.53	.70	880
	H4	15	.56	.72	2660	.57	.69	861
Grade 8/Age 13	H2	26	.75	.86	2534	.75	.87	858
	H3	26	.65	.82	2560	—	—	—
	H4	26	.64	.81	2538	—	—	—
	H5	26	.44	.69	2543	.43	.67	854
	H6	28	.54	.78	2574	.57	.76	858
	H7	28	.44	.75	2552	—	—	—
Grade 12/Age 17	H2	28	.54	.82	2458	.57	.81	815
	H3	25	.56	.80	2458	—	—	—
	H4	25	.54	.81	2461	—	—	—
	H5	26	.56	.79	2459	—	—	—
	H6	28	.6[illegible]	.85	2463	.72	.85	810
	H7	28	.61	.85	2459	—	—	—

339

340

was derived using the item response theory (IRT) methodology extensively documented in Chapter 9 of this report. Only a brief outline of the scaling procedures is given here.

Table 13-5
U.S. History Proficiency Scale Item Information, by Sample

Sample Code	Cohort	No. of History Scale Items
9[Main-His] and 9[Main-Int]	Grade 4/age 9	44
13[Main-His] and 13[Main-Int]	Grade 8/age 13	158
17[Main-His] and 17[Main-Int]	Grade 12/age 17	159
17[Br86-His]	Grade 11/age 17	0 (Only item percents correct were reported)

Three U.S. history items were excluded from the U.S. history proficiency scale because of lack of fit to the IRT model. On these items, two grade/age samples had different empirical response curves. In one case (item H003401), one grade/age sample had an item response curve that was not monotonically increasing. Table 13-6 identifies the U.S. history items excluded from the scale. Only the first rater's (dichotomized) scores on the two open-ended items (H024901 and H025003) were used in scaling. The second rater's scores were used only for estimation of interrater reliability.

Table 13-6
U.S. History Item Scores Excluded from IRT Scale

NAEP ID	Description	Grade/Age
H003401	Emancipation Proclamation	8/13, 12/17
H012601	American Revolution	4/9, 8/13
H022401	Association with F. D. Roosevelt	8/13, 12/17

Item Calibration

For the combined focused BIB/intercorrelation samples across all three grade/age levels, the BILOG program (Mislevy & Bock, 1982) was used to obtain item parameter estimates on a provisional scale based on the three-parameter logistic model. Parameters were estimated even for previously administered items; parameter values were not assumed to be equal to their 1986 values.

Conditioning

Before proficiency estimation was initiated, conditioning variables were derived for each respondent. A common set of conditioning variables was used in all subject areas in the 1988 assessment, along with background items specific to the subject area being scaled. The common set of conditioning variables is listed in Table C-1 of Appendix C. The specific history background items that were used as conditioning variables in the U.S. history scaling and reporting are listed in Table C-8 of Appendix C. Each of the U.S. history grade/age samples for which proficiency values were estimated were conditioned using the common set of conditioning variables (common across all subjects areas within a given grade/age sample) and the grade/age background items specific to the respective U.S. history sample. The one exception was the sample at the grade 12/age 17 level, where the category "modal age > modal grade" was not used, inasmuch as no one in this sample can be enrolled in a grade above the modal grade (12). Tables C-34, C-35, and C-36 of Appendix C list the estimated effects for the common conditioning variables and the U.S. history attitude items that were obtained from the M-GROUP program for each of the grade/age samples.

Proficiency Estimation

Using the iterative method described by Mislevy (1985) and implemented in the M-GROUP program (Sheehan, 1985), a U.S. history proficiency distribution was estimated for each respondent. The M-GROUP program was applied to each grade/age separately, and from each respondent's distribution five random values were drawn and saved. These "plausible values" were transformed by standardizing the combined distribution for the three grade/ages to a mean of 250.5 and a standard deviation of 50.0. The five plausible values for each respondent were then used for estimating demographic group statistics.

13.1.2 Anchoring

The U.S. history proficiency scale was "anchored" using the same procedures described in Chapter 7. Based on the range of student performance in the assessment, the following levels of U.S. history proficiency were defined:

- Level 200: Knows Simple Historical Facts
- Level 250: Knows Beginning Historical Information and has Rudimentary Interpretive Skills
- Level 300: Understands Basic Historical Terms and Relationships
- Level 350: Interprets Historical Information and Ideas

To provide the basis for the anchoring process, NAEP used empirical procedures to delineate sets of items that discriminated between adjacent performance levels on the scale—that is, items likely to be answered correctly by students performing at a particular level on the scale and much less likely to be answered correctly by students performing at the next lower level.

The sets of items represented at each of the four levels were studied by a panel of distinguished history educators, who carefully considered and articulated the types of knowledge, skills, and reasoning abilities demonstrated by correct responses to the items in each set. This information was then placed in the context of the assessment framework and used to characterize students' understanding of the significant events, people, and policies that have shaped our nation's history, as well as their ability to interpret and analyze historical information.

Table 13-7 shows the percentages of students in each grade who performed at or above each level of U.S. history proficiency.

Table 13-7
U.S. History Scale Anchoring:
Percentages At or Above Each Level (with Standard Errors), by Grade

Level	Grade 4	Grade 8	Grade 12
200	76.0 (1.0)	96.0 (0.3)	99.4 (0.0)
250	15.9 (0.9)	67.7 (0.9)	88.9 (0.6)
300	0.2 (0.1)	12.7 (0.5)	45.9 (1.3)
350	0.0 (0.0)	0.1 (0.0)	4.6 (0.5)

13.1.3 Classification

The primary objectives of the 1988 U.S. history assessment were to measure students' knowledge of the significant events, persons, and documents that have shaped our nation's history, and to assess their understanding of how these fit into the major chronological time periods of our past. To address these objectives, the U.S. history items were classified as follows:

Context

- Political Life
- Economic Life
- Cultural, Social, and Family Life

Chronology of Events

- Chronology
- Documents
- Persons

Historical Periods

- Exploration and Colonization, up to 1763
- The Revolutionary Era, the Constitution, and the New Republic, 1763-1815
- Economic and Social Development of the Antebellum Republic, 1790-1861
- Crisis of the Union: Origins of the War, the War, and Reconstruction, 1850-1877
- The Rise of Modern America and World War I, 1877-1920
- The United States, 1920-1941
- World War II and the Postwar Era, 1931-1968
- Modern Post-industrial Era. 1968 to the present

Based on these classifications, statistical analyses were performed for different subgroups (e.g., male, female) in each of the grade/age samples. The means and standard errors were generated at the item level, and across the group of items within each of the classification categories. Differences between the subgroups within each of the categories also were compared.

13.2 TREND ANALYSES

To provide information about trends in students' knowledge of U.S. history, 105 U.S. history items were re-administered in 1988 to a bridge sample of grade 11/age 17 students. The 1988 sample of grade 11 students was compared to the equivalent sample from the 1986 assessment.

The U.S. history trend items were not scaled. Given time and cost constraints, it did not seem beneficial to do so, because the only previous U.S. history assessment was from the 1986 sample of grade 11/age 17 students and the modal grade of 11 is no longer being assessed.

Based on preliminary item analyses in which the 1988 results were compared with the results for the 1986 items, the item statistics for the 1988 and the 1986 samples were found to be comparable.

Using the item classifications previously mentioned for the cross-sectional analyses, the means and standard errors were generated for both the 1986 and 1988 samples. The results of the U.S. history trend analyses are reported in *The U.S. History Report Card: The Achievement of Fourth-, Eighth-, and Twelfth-grade Students in 1988 and Trends from 1986 to 1988 in the Factual Knowledge of High-school Juniors* (Hammack, Hartoonian, Howe, Jenkins, Levstik, MacDonald, Mullis, & Owen, 1990).

Chapter 14

DATA ANALYSIS FOR THE GEOGRAPHY ASSESSMENT[1]

Norma A. Norris

Educational Testing Service

In 1988, geography items were administered to two samples of students, as shown in Table 14-1.

Students in the focused-BIB sample for geography, 17[Main-Geo] received three blocks of geography items. Students in the intercorrelation sample for geography, 17[Main-Int], received a block of geography items, a block of reading items, and a block of either U.S. history or civics items.

These two samples were analyzed as a single population in the cross-sectional analyses, in which the levels of geography achievement for various subgroups were compared. The results of these analyses are reported in *The Geography Learning of High-School Seniors* (Allen, Bettis, Kurfman, MacDonald, Mullis, & Salter, 1990).

14.1 CROSS-SECTIONAL DATA ANALYSES

In total, 78 geography items were administered in 1988 to the focused-BIB/intercorrelation sample. All of these items were multiple-choice.

Preliminary item analyses were conducted to check the validity and reliability of the items. For each item, the percent of students selecting each response, the percent who omitted the item, the percent who did not reach the item, and the correlation between the item score and the block score were calculated. Also, for each block, the internal consistency reliability was computed. No items were excluded based on the results of these analyses.

Table 14-2 shows, for each block in both the focused-BIB and intercorrelation samples, the number of items, the mean proportion correct, the KR-20 reliability coefficient, and the number of 12/age 17 students who responded to at least one item in the block.

Percents of students not reaching items ranged from .2 to 8.2 percent for items in block G2, from .5 to 9.9 percent for items in block G3, and from 1.1 to 9.0 percent for items in block G4.

[1]The contributions of Kentaro Yamamoto, Rebecca Zwick, and Lynn Jenkins to this chapter are gratefully acknowledged.

Table 14-1

NAEP 1988 Geography Samples

Sample Code	Sample Type	Subject Areas	Booklet Numbers	Cohort	Time of Assessment	Age Defn.	Modal Grade	Sample Size
17[Main-Geo]	Focused-BIB Main	G	29	Grade 12/age 17	Winter, spring	CY	12	2,446
17[Main-Int]	BIB Intercorr.	R,C,H,G	30, 32	Grade 12/age 17	Winter, spring	CY	12	1,623

R = Reading
C = Civics
H = U.S. history

G = Geography
CY = Calendar year: birth dates in 1978, 1974, and 1970 respectively for ages 9, 13, and 17

Table 14-2

Descriptive Statistics for Geography Blocks

		Focused-BIB Sample			Intercorrelation Sample		
Block	Number of Items	Mean Proportion Correct	KR-20	N	Mean Proportion Correct	KR-20	N
G2	26	.57	.79	2438	.56	.79	803
G3	26	.54	.79	2438	—	—	—
G4	26	.56	.84	2438	.57	.83	805

14.1.1 Scaling

In 1988, the responses to the 78 geography items were summarized on a single geography profici n y scale, derived using the item response theory (IRT) methodology extensively documented in Chapter 9 of this report. Only a brief outline of the scaling procedures is given here.

Item Calibration

For the grade/age sample, the BILOG program (Mislevy & Bock, 1982) was used to obtain item parameter estimates on a provisional scale, based on the three-parameter logistic model.

Conditioning

Before proficiency estimation was initiated, conditioning variables were derived for each respondent. A common set of conditioning variables was used in all subject areas in the 1988 assessment, along with background items specific to the subject area being scaled. The common set of conditioning variables is listed in Table C-1 of Appendix C. The geography-specific background items that were used as conditioning variables in the geography scaling and reporting are listed in Table C-9 of Appendix C. The one exception was the sample at the grade 12/age 17 level, where the category "modal age, > modal grade" was not used, inasmuch as no one in that sample can be enrolled in a grade above the modal grade (12). Table C-37 of Appendix C lists the estimated effects for the common conditioning variables and the geography background items that were obtained from the M-GROUP program.

Proficiency Estimation

Using the iterative method described by Mislevy (1985) and implemented in the M-GROUP program (Sheehan, 1985), a geography proficiency distribution was estimated for each respondent, and from each respondent's distribution five random values were drawn and saved. These "plausible values" were transformed by standardizing to a mean of 285.0 and a standard deviation of 40.0. (This mean and standard deviation were chosen to be similar to those obtained for grade 11 reading proficiency in 1984. The same standardization was used in the U.S. history and literature proficiency scales in 1986.) The five plausible values for each respondent were then used for estimating demographic group statistics.

14.1.2 Classification

The objectives for NAEP's 1988 geography assessment reflected a broad-based consensus of university professors, classroom teachers, social science researchers, school administrators, and curriculum specialists. In brief, the assessment framework emphasized that students should be able to use e skills and tools of geography, including maps, charts, and globes; that they should know and understand the concepts underlying cultural and physical geography, including the locations of places, resources, and cultural areas; and that they should be able to apply geography principles. To address these objectives, the geography items were classified as follows:

- Knowing locations
- Using the skills and tools of geography
- Understanding cultural geography
- Understanding physical geography

Based upon these classifications, statistical analyses were performed for different subgroups (e.g., male, female). The means and standard errors were generated at the item level and across the group of items within each of the classifications. Differences between the subgroups within each of the classifications also were compared.

Chapter 15

DATA ANALYSIS FOR MATHEMATICS AND SCIENCE[1]

Kentaro Yamamoto

Educational Testing Service

The mathematics and science items were part of the samples used in 1988 to bridge back to the 1986 trend samples. This chapter describes the technical details of the item-parameter estimation and scaling that were performed for trend analyses of responses to mathematics and science cognitive items in the 1988 assessment. The results of the analyses are presented in greater detail in *The Effect of Changes in the National Assessment: Disentangling the NAEP 1985-86 Reading Anomaly* (Beaton & Zwick, 1990)

To maintain the comparability of measurement instruments, booklets for the 1988 reading bridge to 1986 were identical to those used in 1986 and therefore included science and mathematics blocks. The 1988 mathematics and science trend analyses are limited to data from blocks that appeared in the same booklets as the reading blocks in the 1986 assessment. For age 17, the number of mathematics and science blocks available for trend analysis was smaller in 1988 than in 1986. However, since every 1986 trend booklet for ages 9 and 13 contained a block from each of the three subject areas, the complete sets of trend blocks that were available in 1986 for both ages were also available for analysis in 1988.

The combination of blocks within booklets, the composition of item blocks, the mode of administration, the sample definition, and the time of testing were identical for the age 9 and age 13 samples in the 1986 assessment and the 1988 bridge to 1986. Consequently, trend analyses for these two ages were straightforward; trend analyses for age 17, however, were not.

In 1986, the reading trend for age 17 was assessed as part of the BIB spiral portion of the assessment, while the science and mathematics trends were assessed apart from reading under a paced-tape mode of administration Since the overarching aim of the 1988 bridge study was to replicate the booklets and administration procedures for the 1986 assessment of trends in reading, booklets from the BIB spiral portion of the 1986 assessment were again administered in 1988 under the same administration conditions as in 1986. In particular, the administration of mathematics and science items in the spiral portion made use of paper and pencil, rather than paced tape. An implication of this mode of administration was that the data from the 1988 age

[1]Maxine Kingston, Edward Kulick, Michael Narcowich, and Minhwei Wang performed the data analyses for this chapter; Edward Kulick produced the figure. Robert Mislevy provided consultation on scaling and Rebecca Zwick provided valuable editorial assistance.

17 trend assessments of mathematics and science are comparable to the 1986 BIB assessment, but not directly to the 1986 trend assessment. As a result, the design to align the 1988 trend point for age 17 student to the past trend was more complicated than before. For age 17 in 1988, two types of equating were necessary—one based on common populations across different modes of administration for the 1986 BIB and trend, and one based on common items (similarly placed) for the 1986 BIB and the 1988 trend.

The main objective of the 1988 trend assessments of mathematics and science was to evaluate the differences between the 1986 and 1988 assessments. The 1988 trend point was to be added to the existing trend line. Since these analyses closely follow those conducted in 1986, readers desiring more detailed descriptions are referred to relevant chapters in *Expanding the New Design: The NAEP 1985-86 Technical Report* (Beaton, 1988). This chapter will consider details specific to the 1988 analysis.

15.1 SAMPLING OF STUDENTS AND ITEMS FOR MATHEMATICS AND SCIENCE

For ages 9 and 13, the combination of blocks, composition of item blocks, mode of administration, age definition, and time of testing in 1986 were identical to those in the 1988 bridge to 1986. Three booklets, identical to those used in 1986 including background questions, were used to measure trend for these ages. Each booklet contained one reading, one mathematics, and one science block. Each student in the sample was administered one of these booklets. The mathematics and science portions were presented aurally using a tape recorder as in past assessments. The tape recorder was turned off for the reading block.

For age 17, the mathematics and science booklets of the 1986 trend assessment were not used in 1988, since the 1986 mathematics and science trend booklets for age 17 did not include reading blocks. Instead, the booklets used in 1988 were identical to a subset of booklets used for the 1986 BIB assessment and consisted of six booklets, five of which contained at least one reading block and either a mathematics or a science trend block from the 1986 assessment. The sixth booklet, which did not contain mathematics or science blocks, was included only for the reading assessment in 1988. Three trend blocks that appeared in the 1986 age 17 trend assessments were used for the 1988 age 17 trend assessment, even though some booklets included additional mathematics and science blocks used only for the cross-sectional assessment in 1986. Only one of the first two trend blocks of either mathematics or science was included in four of the booklets; the fifth booklet contained both a mathematics and a science block (details of block arrangements are documented in Chapter 4). The 1988 age 17 sample was defined using the same age definition as the 1986 BIB assessment and received a print-administered assessment instead of the paced administration of the pre-1988 trend assessments. Unlike the samples at ages 9 and 13, in which every student received both a mathematics and a science block, about one-fifth of the age 17 sample received both; the rest received a block of either mathematics or science items.

The proficiencies of the three ages cannot be placed on a single scale without a cross-sectional study or a vertical equating across ages, neither of which were possible in the 1988 mathematics and science trend assessment. The mathematics and science scales were derived from the 1986 cross-sectional assessment (see E. G. Johnson, 1988, and Yamamoto, 1988). The 1988 trend analysis added a new trend point to the existing trend line up to 1986.

The specific mathematics and science samples for 1988 and 1986 are shown in Table 15-1.

The items used for the analysis of the 1988 data set are the same as those used for the 1986 trend analyses; that is, the same items were excluded as in 1986 for reasons of lack of fit of the estimated item response function to the empirical regression curve in either the bridge or the cross-sectional data. Three mathematics items, one from each age group, were excluded from the scaling. The calculator items were excluded from the analysis. Seven science items were dropped from the scaling for age 9, seven were dropped for age 13, and four were dropped for age 17.

Using current methods, it is possible to assess the change over time in either item characteristics or proficiencies of populations, but not both at the same time. This is true for any analysis, whether based on classical test theory, item response theory, or proportions correct. To assess change in item characteristics, we are forced to assume that the ability distribution of the population remains stable; to assess change in the ability distribution of the population, we must assume that the item characteristics are stable (see the discussion of common-item equating in Chapter 10). However, we know that these assumptions are not strictly justified. Societal and instructional changes may produce gradual alterations in item functioning over time. If there is evidence that this is occurring, it may be desirable to allow for changes in the parameters of these common items. Permitting item characteristics to vary in this way is feasible only if common-population equating methods are available to link the newly obtained results to past trend lines. This is the approach that was used in analyzing the 1988 mathematics data at age 17 and science data at all three ages.

15.2 SCALING OF THE MATHEMATICS TREND DATA

Mathematics Trend Scaling for Ages 9 and 13

From the item analysis, it was found that the 1988 response distributions of all response choices, including "omits," were quite similar to the 1986 data. The mean weighted proportion correct at the block level was computed; these values were compared with the 1986 results, as shown in Table 15-2. At each block level for all age groups, the 1988 sample showed higher weighted proportion correct values than the 1986 sample.

In estimating item parameters in 1986, combined data from the three most recent trend assessments (1977, 1982, and 1986) were used. Thus, the 1986 trend analysis assumed the characteristics of all items were stable across the three assessments. Item parameters estimated in 1986 were kept unchanged for

Table 15-1

Mathematics and Science Samples, 1988 Assessment

Sample Code	Sample Type	Subject Areas	Booklet Numbers	Mode of Administration	Cohort	Time of Testing	Age Defn.	Modal Grade	Sample Size
Mathematics and Science									
86:9a	1986 bridge	RMS	91-93	Tape	Age 9	Winter	CY	4	6,932
9[Br86-RMS]	1988 bridge	RMS	91-93	Tape	Age 9	Winter	CY	4	3,711
86:13a	1986 bridge	RMS	91-93	Tape	Age 13	Fall	CY	8	6,200
13[Br86-RMS]	1988 bridge	RMS	91-93	Tape	Age 13	Fall	CY	8	3,942
Mathematics									
86:17	1986 main	RMS	14,36,68	Print	Grade 11/age 17	Spring	Not CY	11	6,151*
86:17b	1986 bridge	RMS	94-95	Tape	Age 17	Spring	Not CY	11	3,868
17[Br86-RMS]	1988 bridge	RMS	61,62,65	Print	Grade 11/age 17	Spring	Not CY	11	1,852*
Science									
86:17	1986 main	RMS	47,62,68	Print	Grade 11/age 17	Spring	Not CY	11	5,611*
86:17b	1986 bridge	RMS	94-95	Tape	Age 17	Spring	Not CY	11	3,868
17[Br86-RMS]	1986 bridge	RMS	63,64,65	Print	Grade 11/age 17	Spring	Not CY	11	1,862*

* Number of age-only students who answered any one of the trend blocks.

Note: 1) For all three ages, mathematics 1988 trend blocks are identical to those administered in 1986; 2) Only the subset of the 86:17 and 17[Br86-RMS] samples that were age-eligible and received trend blocks were used, and numbers on the table reflect such samples.

Table 15-2

Mathematics Weighted Mean Proportion Correct

	Block	1986	(N)	1988	(N)	No. of Items[c]	
Age 17 (paper)	1	59.1	(2211)[a]	61.3	(619)	35	
	2	63.4	(2233)[a]	65.7	(624)	35	
	3	65.3	(2263)[a]	67.0	(609)	24	(19)
	Total	62.3	(6157)[a]	64.4	(1852)	94	
	Noncalculator	61.0		62.7		75	
Age 17 (taped)	1	60.3	(1934)[b]			35	
	2	62.1	(1934)[b]			35	
	3	64.5	(1934)[b]			24	(19)
	Total	62.0	(3868)[b]			94	
	Noncalculator	60.8				75	
Age 13 (taped)	1	63.9	(2075)	65.3	(1405)	37	
	2	58.5	(2054)	60.5	(1281)	37	
	3	57.4	(2071)	60.0	(1256)	24	(16)
	Total	60.3	(6200)	62.2	(3942)	98	
	Noncalculator	61.4		63.2		82	
Age 9 (taped)	1	55.2	(2315)	58.2	(1274)	26	
	2	57.3	(2361)	62.4	(1240)	26	
	3	73.0	(2256)	76.7	(1197)	16	(11)
	Total	60.2	(6932)	64.2	(3711)	68	
	Noncalculator	57.1		62.1		57	

[a] Age-only BIB sample with at least one mathematics trend block.

[b] 1986 age 17 trend sample blocks 1 and 2 were paired.

[c] Includes some items that were excluded from IRT scaling; parentheses in this column indicate the number of calculator items excluded from IRT scaling.

the 1988 assessment for ages 9 and 13 after the fit of the 1988 data to the item parameters was examined visually, as well as by means of the chi-square test, for every item. Consequently, the same linear function as in 1986 was used to transform provisional imputed values to the mathematics proficiency scale. Although the use of previously estimated item parameters was justified at ages 9 and 13, this was not the case at age 17. Hence, the item parameters applicable to age 9 and age 13 were kept unchanged for the mathematics trend analysis; they are presented in Tables F-8 and F-9 in Appendix F.

The coexistence of item parameters that fit in various degrees to the data from a particular year comes from the need to place several samples from different years on a scale based upon common-item equating.[2] When common-item parameters are estimated on multiple data sets, the fit of the estimated item response functions to the weighted means of expected proportions correct, given an ability level, is maximized. Because of this averaging over multiple data sets, it is possible that the estimated item parameters will fit very well to the combined data sets as a whole, but less well to each data set separately. For each item, fit of the estimated item parameters was examined also for the some of the major subpopulations, such as male, female, White, Black, and Hispanic. Although a few items indicated that the estimated item parameters fit less well to a particular subpopulation, the examination of these items did not reveal any explanation for the misfit.

For ages 9 and 13, the same common-item equating procedure that was employed in the 1986 trend analysis was used to align the 1988 point to the trend up to 1986. A brief description of the procedure follows. From the item parameters estimated in 1986 and background variables of 1988, the proficiency scores were imputed for the 1988 bridge data for each age using the M-GROUP computer program based on the plausible values methodology (Sheehan, 1985; see Mislevy, 1988b, for a detailed discussion). Appendix C gives the conditioning variables (Table C-10) and the estimated conditioning effects (Tables C-38 and C-39) for ages 9 and 13. The same linear constants as in 1986 were used to transform provisional imputed scores to the final proficiency scores for mathematics trend. The transformation constants for all three ages are listed in Table 15-3.

Table 15-3
Coefficients of the Linear Transformation of the Trend Scale from Original Units to the Mathematics Proficiency Scale

Age	Intercept	Slope
9	218.42	35.84
13	266.58	34.57
17	303.25	31.84

[2]The IRT equating of two assessments can be based either on an assumption that the proficiency distributions for the two assessments are the same (common-population equating) or an assumption that a set of common items are functioning in the same manner in both assessments (common-item equating). The 1988 mathematics and science trend assessments used common-item equating methods. In the future, we may consider an alternate equating method based on the distributions of item parameters.

Mathematics Trend Scaling for Age 17

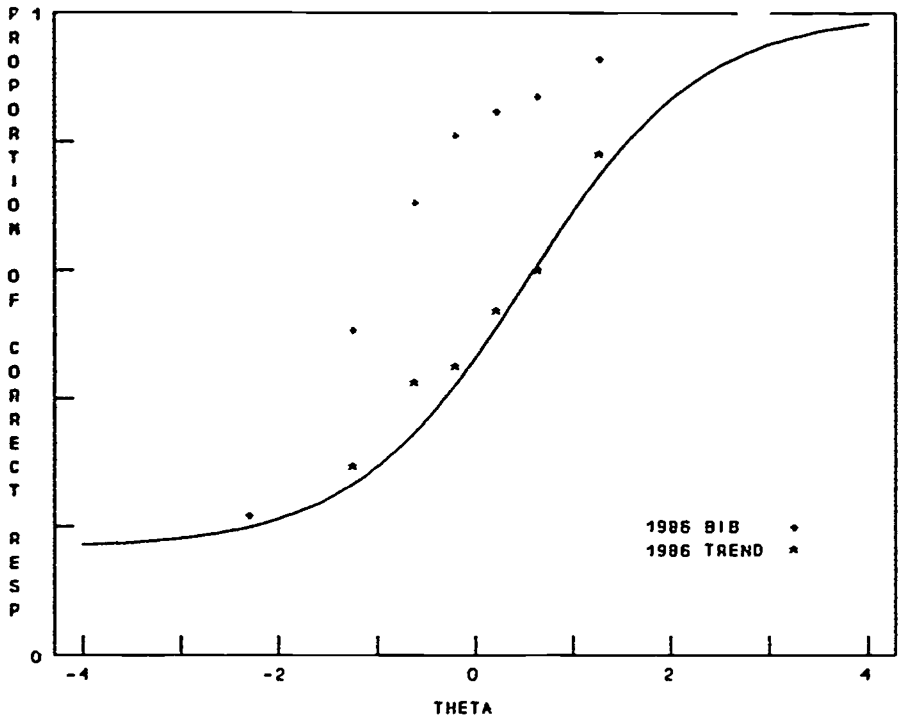
To scale the age 17 mathematics trend data, new item parameters were estimated using the subsample from the 1986 BIB assessment equivalent to the 1988 trend sample. Use of the estimated item parameters in 1986 mathematics trend is not appropriate for the 1988 assessment for age 17, because of the different mode of administration for the 1986 and the 1988 trend assessments for that age. For example, on all five items of a type referred to as "estimate" items, use of paper and pencil instead of a tape recorder had a dramatic effect. "Estimate" items ask the student to select an answer among several options, all of which are rounded so that none of them is exactly correct. The property of the response options is indicated by the word "about" being positioned before "how much" or "how many" in a question. When an "estimate" item was presented under taped administration, enough time was allowed for rough estimation of the (typically) large number, but not enough time was allowed for the numerical calculation of the answer. However, because under paper-and-pencil administration it is possible to spend more time to answer, the examinee may opt to perform the calculation rather than the estimation. In such a case, it is more appropriate to treat an "estimate" item as two different items under different modes of administration. Figure 15-1 presents the observed item regression curves of the 1986 BIB data and 1986 bridge data for one of the "estimate" items, along with the item response function estimated for the 1986 bridge data.

Therefore, for age 17, both equating methods, common-item (between the 1986 BIB and 1988 bridge samples) and common-population (between the 1986 BIB and 1986 bridge samples), were used to place the 1988 trend sample on a scale comparable to the 1986 reported scale. The procedure took place as follows. The item parameters for the total set of 73 items were estimated based on the two data sets: the 1986 BIB assessment and the 1988 bridge to 1986. Both samples included grade- and age-eligible students in order to maintain an adequate sample size for the estimation accuracy. This resulted in a second set of item parameters for age 17. The new item parameters are listed in Table F-10, Appendix F; the old parameters appear in Beaton (1988). The rationale for estimating parameters for all items instead of only "estimate" items comes from the main objective of the 1988 bridge to 1986, namely to examine the possibility of effects due to changes in assessment procedures. For each item, fit of the estimated item parameters was examined for some of the major subpopulations, such as male, female, White, Black, and Hispanic.

From the above estimated item parameters and background information for the appropriate sample, proficiency scores were imputed for each student in the 1986 BIB and 1988 bridge-to-1986 samples. Appendix C gives the conditioning variables (Table C-10) and the estimated conditioning effects (Table C-40) for age 17. Then the mean and standard deviation of the imputed scores of the age-only subsample of the 1986 BIB were calculated. Constants were calculated and used to match the means and standard deviations of the proficiency scores of the 1986 trend sample and the age-only subsample of the 1986 BIB sample. Subsequently, by applying the same linear transformation to the provisional imputed values of the 1988 trend age-only sample, the 1988

Figure 15-1

A Plot of Observed Proportion Correct of
the 1986 BIB Spiral and Trend Assessments with the Estimated
Item Response Function for an "Estimate" Item



trend point was aligned with the trend line up to 1986. The transformation constants for age 17 data are listed in Table 15-3.

15.3 SCALING OF THE SCIENCE TREND DATA

The 1988 science trend analysis followed procedures and methods similar to those for the mathematics analysis. From the item analysis, it was found that the 1988 response distributions of all response choices, including "omits," were quite similar to the 1986 data. The mean weighted proportion correct at a block level was computed; these values were compared with the 1986 results, and are presented in Table 15-4.

In 1986, item parameters were estimated for the age 9, 13, and 17 samples. The trend items for age 13 and age 17 were estimated together because the majority of the items were common to both ages. For the 1988 data, because of the change in the mode of administration for age 17, those items had to be estimated separately from the age 13 items. To obtain the best estimates of proficiencies for the two years, items for age 13 were reestimated using BILOG (Mislevy & Bock, 1982) on the 1986 and 1988 bridge data sets. It was found that, for one of 63 items in 1986, "I don't know" responses were treated as wrong when they should have been treated as "omit." This error was found only in the 1986 bridge data set for age 9. There are two reasons for judging that the effect of this error on the proficiency score is very small: It involved only 8 percent of the responses for a single item, and the subjects who selected the "I don't know" option had the lowest mean proportion correct among all options. In fact, using the trend item parameters from 1986 estimated on the incorrect data sets, we compared the means of the ability distributions of two data sets with and without correction of the 1986 age 9 trend and found that they differed by about .07 on the proficiency scale. In order to assess administration effects as accurately as possible, however, the item parameters for all items were estimated for age 9 based on the 1986 and 1988 corrected bridge data sets. For each item, fit of the estimated item parameters was examined for some of the major subpopulations, such as male, female, White, Black, and Hispanic. The estimated item parameters for three ages are listed in Tables F-11, F-12, and F-13 in Appendix F. Although a few items indicated that the estimated item parameters fit less well to a particular subpopulation, the examination of these items did not provide any clear explanations.

The imputed proficiency values of the 1988 sample were calculated from the responses on cognitive items and background questions based on the item parameters estimated on the trend samples of 1986 and 1988. The imputed values of the 1988 sample were transformed to become comparable to the trend scale of 1986. Note that the 1986 sample was used twice to obtain two separate sets of trend item parameters, the one for the data up to and including 1986 and the other for the data from 1986 and 1988. This design enabled us to use common-population equating based on the same sample, and also to express the difference in the distribution of proficiency between 1986 and 1988 in terms of the trend scale established in 1986. The linear transformations were derived separately for ages 9 and 13 to match, within

Table 15-4

Science Weighted Mean Proportion Correct

	Block	1986	(N)	1988	(N)	No. of Items [c]
Age 17 (paper)	1	60.5	(2223) [a]	60.6	(634)	27
	2	59.0	(1935) [a]	60.7	(619)	32
	3	53.7	(2282) [a]	56.3	(609)	23
	Total	58.0	(3611) [a]	59.5	(1862)	82
Age 17 (taped)	1	63.3	(1934) [b]			27
	2	63.4	(1934) [b]			32
	3	58.9	(1934) [b]			23
	Total	62.1	(3868) [b]			82
Age 13 (taped)	1	52.5	(2075)	53.8	(1405)	25
	2	54.2	(2054)	54.7	(1281)	31
	3	56.2	(2071)	57.8	(1256)	27
	Total	54.3	(6200)	55.5	(3942)	83
Age 9 (taped)	1	59.4	(2315)	62.6	(1274)	18
	2	52.5	(2361)	53.5	(1240)	25
	3	68.5	(2256)	69.0	(1197)	20
	Total	59.5	(6932)	61.0	(3711)	63

[a] Age-only BIB sample with at least one science trend block.
[b] 1986 age 17 trend sample blocks 1 and 2 were paired.
[c] Includes some items that were excluded from IRT scaling.

each age cohort, the two means and standard deviations of proficiencies of the 1986 bridge sample, one based on the item parameters estimated on the data until 1986 and the other based on the item parameters estimated on the 1986 and 1988 data. The linear constants derived from those transformations were applied to the 1988 data set to obtain trend points for 1988. For age 17, we applied an equating method identical to that used for the age 17 mathematics data. Appendix C gives the conditioning variables (Table C-11) and the estimated conditioning effects (Tables C-41, C-42, and C-43) for all three ages. Table 15-5 presents the linear coefficients used for the three ages.

Table 15-5
Coefficients of the Linear Transformation of the Trend Scale from Original Units to the Science Proficiency Scale

Age	Intercept	Slope
9	225.59	41.15
13	254.19	36.92
17	289.34	43.05

PART III

Statistical Summary of 1988 NAEP Data

Chapter 16

STATISTICAL SUMMARY OF THE 1988 NAEP SAMPLES AND ESTIMATES OF THE PROFICIENCIES OF AMERICAN STUDENTS[1]

Bruce A. Kaplan

Educational Testing Service

The analysis of the 1988 NAEP data has resulted in the production of many thousands of tables containing estimates of the proficiency of students, and various subgroups of students, in American schools. This chapter gives some selected results from the assessment as well as a statistical summary of the 1988 NAEP samples. The chapter assumes a general familiarity with the structure of NAEP as summarized in the Introduction and the overview Chapters 1 and 7.

Three of the many types of NAEP results are presented here:

- results of the instrument development process, including the sizes of the item pools and numbers of booklets;

- results of the sampling process, including the numbers of students in each sample by selected subgroups; and

- results of the parameter estimation process, including estimates of the proficiencies of several populations of students in reading, writing, civics, U.S. history, geography, mathematics, and science.

Interpretive results from the estimates presented here have been reported in the NAEP subject area trend and cross-sectional reports. The 1988 public-use data tapes and user guide (Rogers, Kline, Johnson, Mislevy, & Rust, 1990) are available for those who wish to estimate other parameters of student performance from the NAEP data or to search for possible explanations for the population characteristics that are reported here.

The technical details of the estimation process that underlie these tables are covered in previous parts of this report and not repeated here. A detailed discussion of how to read and use the tables of background and proficiency results is given by Zwick (1987b).

---

[1]The author is indebted to Albert Beaton for portions of this chapter. Information for various tables in this chapter was provided by John Ferris, David Freund, Lynn Jenkins, Edward Kulick, Michael Narcowich, Norma Norris, Kate Pashley, and Keith Rust.

16.1 MEASUREMENT INSTRUMENTS

In 1988 a total of 34 assessment booklets and questionnaires was printed for age class 9, 49 for age class 13, and 53 for age class 17. These booklets are enumerated by age level and by type of measurement instrument in Table 16-1. Some of the instruments were used at more than one age/grade level.

The item pool used to develop these booklets is described in Table 16-2. In general, there are two types of items, cognitive and noncognitive. The cognitive items are developed to measure proficiency in particular subject areas, such as reading and mathematics. Cognitive items may be open-ended or multiple-choice. The noncognitive items are usually questions about the student's or teacher's backgrounds and attitudes but may also probe other areas such as school policies or teaching methods. Because many items were used at more than one age class, the total number of items in an item pool is not the sum of the item pools used for the three age classes.

All of the items in the subject area pools were used for the main NAEP assessment, but not all could be used for the various bridge assessments. Table 16-3 shows the number of cognitive items in each subject area that were used in the separate samples.

The excluded student, teacher, and school questionnaires contained only noncognitive questions. The number of items in the noncognitive pools is the same as the number of items in the questionnaires. More information about the instruments that were developed is provided in Chapters 2 and 4.

16.2 SAMPLE CHARACTERISTICS

In this section, the characteristics of the final NAEP sample is described. The process by which the sample was selected is discussed in Chapter 3.

In the 1988 main assessment, NAEP contacted 1,29 schools, of which 1,030 contributed data to the assessment. The disposition of these schools is shown in Table 16-4. Some of the schools were unwilling to cooperate; others were believed to be eligible from the sampling frame, but were not. The cooperation rate is calculated as the sum of cooperating schools and the schools that were found to have no eligible students divided by the same sum plus the schools that refused or were from districts that refused to cooperate.

Table 16-4 also shows the number of schools in several categories: region of the country (northeast, southeast, central, west), school governance (public, private, Catholic, Bureau of Indian Affairs, Department of Defense), size and type of community, degree of urbanicity, grade span of school, number of teachers, and number of students.

For the 1988 bridge (trend) studies, NAEP contacted 558 schools, of which 441 contributed data to the various bridge assessments. Table 16-5 supplies the same information for the schools assessed for the bridge studies that the previous table supplies for the main assessment schools.

The numbers of respondents to the teacher questionnaire are summarized in Table 16 6. The first column in this table includes the number of teachers who responded by age class and subject area. The final column contains the number of students whose teachers responded to the questionnaire.

NAEP is administered in units called assessment sessions. If the number of students attending an assessment session is less than a predetermined number, the students missing from the session are assigned to a makeup session and then assessed. Table 16-7 shows the number of regular and makeup sessions in 1988 NAEP by age class for the main NAEP and two bridge samples.

Altogether, 133,542 students were involved in 1988 NAEP, including excluded students. The breakdown by age class and by sample is shown in Table 16-8

Tables 16-9 through 16-11 display the distribution of the students assessed in the main NAEP assessment in several basic categories for the three age classes: gender, racial/ethnic grouping, region of the country, parental education, and size and type of community. These tables have four columns:

- eligible by age, which means that the students were in an appropriate age group;
- eligible by grade, which means that the students were in an appropriate grade;
- eligible by age and by grade, which means that the students were of both an appropriate age and appropriate grade; and
- eligible by age or by grade, which is the total number of students for whom data were collected.

Tables 16-12 through 16-19 contain the distribution of students in the same categories by age class for the bridge samples. Tables 16-12 to 16-1 contain he distributions for the bridge to 1984 sample. Tables 16-15 to 16-17 display the distributions for the bridge to 1986 sample. Table 16-18 and Table 16-19 enumerate the students in the two age classes assessed as part of the civics bridge.

Similarly, Tables 16-20 through 16-23 contain the distribution of excluded students by age class. The distribution for the excluded students in the main and bridge samples combined are displayed in Tables 16-20 and 16-21 for grade 4/age 9 and grade 8/age 13, respectively. Table 16-22 contains the distribution of excluded students for the main sample grade 12/age 17 students, while Table 16-23 contains the distribution of excluded students for the bridge samples of grade 11/age 17. These two samples could not be combined because of different age definitions.

16.3 POPULATION ESTIMATES

The 1988 NAEP samples were designed for estimating the size and attributes of a number of different populations of students. The estimation procedures use sampling weights, developed by Westat, Inc., that are used in conjunction with the members of the sample (see Chapter 3). In this chapter, all estimates of population parameters use these sampling weights.

Table 16-24 shows the sizes of the various samples and the sums of their sampling weights by grade/age. In most cases, the sum of the weights for a given sample is an estimate of the number of students who are in the population represented by the sample. In other words, the sum of the weights is an estimate of the number of eligible stud s in the grade or age of interest.

The main assessment can be divided into winter and spring subsamples. These subsamples each have their own sets of weights, as well as a set of weights that allows the combination of the two subsamples. The sample sizes and the estimated population sizes for the winter and spring subsamples, as well as the combined main assessment sample are provided in Table 16-24. The combined set of weights will be used in subsequent tables.

Note that the samples for the main asse ment, the samples for all three age classes of the bridge to 1984, and the samples for the oldest age class for the bridge to 1986 are grade and age samples. The samples for the younger age classes of the bridge to 1986 and the samples for the civics bridge are age-only samples.

The sum of the weights of the excluded students estimates the number of ineligible students in the respective grade/age class. The 17-year-old excluded students are split into two samples, one for which the modal grade is 12 (for the main samples), and one for which the modal grade is 11 (for the bridge samples). This split was necessary because of differences in the time of testing and age definitions used for the bridge and main samples (see Chapter 3).

In most cases, the number of students in a grade/age combination is not of interest; a researcher will be interested in estimating the number of students at either a grade or an age level. For the samples that contain both grade- and age-eligible students, an estimate of the number of students at an age level can be made by summing the weights of only the age-eligible students and adding the corresponding sample of age-eligible excluded students. An estimate of the number of students in a grade sample can be made by summing the weights of grade-eligible students plus the weights of grade-eligible students from the appropriate excluded student sample.

From the main NAEP samples, the next group of tables estimates how many students are age-eligible and grade-eligible by age class. Tables 16-25 through 16-27 show how many students at a particular grade level are at, in, or above the modal age for that grade, and how many at a particular age level are at, in, or above the modal grade for that age. Along with the counts from these samples, the sum of the weights (Weighted N) for each category is presented, and these sums are estimates of the numbers of students in these categories in the population. The standard errors of these estimates and coefficients of variation are also given. (The coefficient of variation is defined as 100 times the standard error divided by the weighted N.)

Tables 16-28 through 16-34 contain the same type of information for the several bridge booklets, by age level. Tables 16-28 to 16-30 have the identical format as the main assessment tables, but contain the information for the bridge to 1984. Table 16-33 is also of the same format, and contains the data for grade 11/age 17 bridge to 1986. Table 16-31 and 16-32 are for the other two age-only samples of the bridge to 1986. Since these are age-only samples, the partitioning of the sample by modal age groupings provides no added information. Instead we supply information by booklet. This is useful because each booklet was weighted so that the sum of the weights of all students completing a booklet estimates the population size. Table 16-34 is constructed for the civics bridges. These are also age-only samples, with one booklet per age group, so the information can be provided on one table.

The next tables show the sizes of the estimated populations of assessable students and the weighted percentages for various NAEP reporting categories. These categories include gender, racial/ethnic grouping, region of the country, parents' education, and size and type of community. The estimated subpopulation percentages for the main NAEP samples are shown in Tables 16-35 through 16-37, separately by age eligibility, grade eligibility, and grade/age eligibility. Tables 16-38 to 16-45 show the same information for the bridge samples. In a similar manner, Tables 16-46 to 16-49 show the estimated total population of excluded students and the weighted percentages by demographic subgroups.

Students were assigned proficiency values in a subject area only if they received at least one assessment block in that area, and thus the sample sizes of students who have proficiency values vary from one subject area to another. Tables 16-50 through Table 16-60 show the number of students with proficiency values in each subject area by age and grade combinations.

Tables 16-61 to 16-70 contain population estimates of student proficiencies by grade and by the subpopulations of gender and race/ethnicity. The information about proficiency includes the mean and standard deviation of each subpopulation as well as the value of the 5th, 10th, 25th, 50th (median), 75th, 90th, and 95th percentiles. Results are shown separately for each subject area. Standard errors of the estimates are included in parentheses.

Tables 16-71 through 16-109 contain results for more finely defined subpopulations. The major reporting variables (gender, race/ethnicity, parental education) are cross-classified with one another. For example, Table 16-72 cross-classifies gender, racial/ethnic grouping, and parental education with the racial/ethnic grouping for fourth graders. Information included about these subpopulations is the actual sample size, the estimated population size (and its coefficient of variation), the proportion of students in each subpopulation (and its standard error), and the average proficiency of the students (and its standard error).

Table 16-1

Measurement Instruments Developed for 1988 NAEP

Instruments	Age Class 9	13	17
Student Assessment Booklets			
Main Sample	22	36	37
Bridge to 1984	6	6	6
Bridge to 1986	3	3	7
Civics Bridge	0	1	1
Total	31	46	51
Questionnaires			
Excluded Student Questionnaire	1	1	1
Teacher Questionnaire	1	1	0
School Characteristics Questionnaire	1	1	1
Total	3	3	2

Table 16-2

Numbers of Distinct Items Administered, by Age Class

	AGE CLASS 9	AGE CLASS 13	AGE CLASS 17	Total Distinct Items*
Common Background	66	68	98	110
Reading				
Background and Attitude	125	137	193	229
Cognitive	168	187	210	384
Writing				
Background and Attitude	45	80	86	89
Cognitive	16	17	17	35
Civics				
Background and Attitude	9	25	40	74
Cognitive	51	168	174	239
U.S. History				
Background and Attitude	9	27	62	98
Cognitive	45	161	211	290
Geography				
Background and Attitude	0	0	28	28
Cognitive	0	0	78	78
Document Literacy				
Background and Attitude	0	0	0	0
Cognitive	0	57	57	57
Mathematics				
Background and Attitude	3	29	69	79
Cognitive	68	84	142	222
Science				
Background and Attitude	16	28	45	61
Cognitive	63	58	110	195
Excluded Student Questionnaire	67	67	67	67
Teacher Questionnaire	67	91	0	131
School Characteristics and Policies Questionnaire	159	164	171	234
Total Distinct Items	989	1432	1844	2674

* Because many items were used at more than one age class, the total number of distinct items is not the sum of distinct items used for the three age classes.

Table 16-3

Numbers of Distinct Cognitive Items
by Age Class and Sample Type

	AGE CLASS 9	AGE CLASS 13	AGE CLASS 17	Total Distinct Items*
Reading				
Cross-sectional	83	100	110	215
Bridge to 1984	105	108	96	193
Bridge to 1986	31	35	72	91
Writing				
Cross-sectional	10	11	11	23
Bridge to 1984	6	6	6	12
Civics				
Cross-sectional	51	155	152	221
Civics Bridge	0	96	101	127
U.S. History				
Cross-sectional	45	161	162	241
Bridge to 1986	0	0	105	105
Geography				
Cross-sectional	0	0	78	78
Document Literacy				
Cross-sectional	0	57	57	57
Mathematics				
Bridge to 1986	68	84	142	222
Science				
Bridge to 1986	63	58	110	195

* Because many items were used at more than one age class, the total number of distinct items is not the sum of distinct items used for the three age classes.

Table 16-4

Characteristics of Schools in Main NAEP (Cross-sectional) Samples

	Grade/Age			
	4/9	8/13	12/17	Total
TOTAL ORIGINAL SAMPLE	384	520	394	1298
COOPERATING	327	399	304	1030
OUT-OF RANGE OR CLOSED	11	14	11	36
NO ELIGIBLES ENROLLED	4	39	13	56
DISTRICT REFUSED	33	40	46	119
SCHOOL REFUSED	9	28	20	57
COOPERATION RATE	88.7	86.6	82.8	86.1
REPLACEMENTS				
REPLACEMENTS FOR REFUSALS	15	18	23	56
REPLACEMENTS COOPERATING	9	14	8	31
TOTALS				
COOPERATING SCHOOLS	336	413	312	1061
COMPLETING QUESTIONNAIRES	316	389	289	994
REGION				
NORTHEAST	69	87	60	216
SOUTHEAST	86	105	85	276
CENTRAL	81	102	79	262
WEST	100	119	88	307
SCHOOL TYPE				
PUBLIC	282	321	272	875
PRIVATE	13	33	19	65
CATHOLIC	41	59	21	121
BUREAU OF INDIAN AFFAIRS	0	0	0	0
DEPT OF DEFENSE	0	0	0	0
SIZE AND TYPE OF COMMUNITY				
RURAL	34	42	30	106
LOW METRO	37	45	32	114
HIGH METRO	35	41	31	107
BIG CITY	47	56	44	147
FRINGE	43	68	35	146
MEDIUM CITY	49	48	29	126
SMALL PLACE	91	113	111	315

Table 16-4 (continued)

Characteristics of Schools in Main NAEP (Cross-sectional) Samples

	Grade/Age			
	4/9	8/13	12/17	Total
URBANICITY				
URBAN	115	137	87	339
SUBURBAN	125	149	106	380
RURAL	95	126	118	339
GRADE SPAN				
KINDERGARTEN TO GRADE 12	18	36	50	104
KINDERGARTEN TO GRADE 6	228	22	0	250
KINDERGARTEN TO GRADE 8	68	120	0	188
GRADE 6 OR 7 TO GRADE 8	6	144	0	150
GRADE 7 TO GRADE 9	0	46	3	49
GRADE 7 TO GRADE 12	0	30	26	56
GRADE 9 TO GRADE 12	0	14	187	201
GRADE 10 TO 12	0	0	45	45
KINDERGARTEN TO GRADE 3	15	0	0	15
NUMBER OF TEACHERS				
UNKNOWN	1	1	1	3
1 - 4	6	12	3	21
5 - 9	28	40	11	79
10 - 19	124	85	37	246
20 - 49	166	199	103	468
50 - 74	11	55	54	120
75 - 99	0	12	49	61
100 +	0	9	54	63
NUMBER OF STUDENTS				
UNKNOWN	1	1	1	3
1 - 99	10	18	9	37
100 - 299	93	103	52	248
300 - 499	121	73	42	236
500 - 749	71	87	38	196
750 - 999	29	69	26	124
1000 - 1499	10	45	61	116
1500 +	1	17	83	101

Table 16-5

Characteristics of Schools in NAEP Bridge Samples

	Age Class			
	9	13	17	Total
TOTAL ORIGINAL SAMPLE	190	223	155	568
COOPERATING	154	173	114	441
OUT-OF RANGE OR CLOSED	10	4	0	14
NO ELIGIBLES ENROLLED	3	30	7	40
DISTRICT REFUSED	12	10	21	43
SCHOOL REFUSED	11	6	13	30
COOPERATION RATE	87.2	92.7	78.1	86.8
REPLACEMENTS				
REPLACEMENTS FOR REFUSALS	7	12	11	30
REPLACEMENTS COOPERATING	2	8	6	16
TOTALS				
COOPERATING SCHOOLS	156	181	120	457
COMPLETING QUESTIONNAIRES	152	167	110	429
REGION				
NORTHEAST	41	38	26	105
SOUTHEAST	37	48	27	112
CENTRAL	38	57	25	120
WEST	40	38	42	120
SCHOOL TYPE				
PUBLIC	174	130	95	349
PRIVATE	12	23	10	45
CATHOLIC	20	28	15	63
BUREAU OF INDIAN AFFAIRS	0	0	0	0
DEPT OF DEFENSE	0	0	0	0
SIZE AND TYPE OF COMMUNITY				
RURAL	13	19	8	40
LOW METRO	16	17	2	35
HIGH METRO	24	24	22	70
BIG CITY	15	11	22	48
FRINGE	22	25	18	65
MEDIUM CITY	16	21	15	52
SMALL PLACE	50	64	33	147

Table 16-5 (continued)

Characteristics of Schools in NAEP Bridge Samples

	Age Class			
	9	13	17	Total
URBANICITY				
URBAN	48	46	39	133
SUBURBAN	66	73	50	189
RURAL	41	62	31	134
GRADE SPAN				
KINDERGARTEN TO GRADE 12	12	25	12	49
KINDERGARTEN TO GRADE 6	95	10	0	105
KINDERGARTEN TO GRADE 8	39	62	0	101
GRADE 6 OR 7 TO GRADE 8	3	44	0	47
GRADE 7 TO GRADE 9	0	12	0	12
GRADE 7 TO GRADE 12	0	13	10	23
GRADE 9 TO GRADE 12	0	15	81	96
GRADE 10 TO 12	0	0	17	17
KINDERGARTEN TO GRADE 3	7	0	0	7
NUMBER OF TEACHERS				
UNCLASSIFIED	0	0	0	0
1 - 4	4	5	2	11
5 - 9	20	20	5	45
10 - 19	49	43	14	106
20 - 49	77	79	35	191
50 - 74	6	23	28	57
75 - 99	0	5	17	22
100 +	0	6	19	25
NUMBER OF STUDENTS				
UNCLASSIFIED	0	0	0	0
1 - 99	5	8	3	16
100 - 299	44	56	21	121
300 - 499	56	40	10	106
500 - 749	34	27	10	71
750 - 999	12	16	17	45
1000 - 1499	4	25	18	47
1500 +	1	9	41	51

Table 16-6

Numbers of Responses to Teacher Questionnaire

Sample	Number of Teachers Responding	Number of Students with Responding Teachers
Grade 4 READING	769	3901
Grade 8 WRITING	756	3570

Table 16-7

Numbers of Assessment Sessions by Sample, Type of Session, and Age Class

	Age Class 9	13	17	Total
MAIN SAMPLE*				
REGULAR	582	635	511	1728
MAKEUP	0	15	68	83
BRIDGE TO 1984 SAMPLE				
REGULAR	184	185	110	479
MAKEUP	1	1	23	25
BRIDGE TO 1986 SAMPLE				
REGULAR	335	217	131	683
MAKEUP	1	0	24	25
CIVICS BRIDGE				
REGULAR	0	105	97	202
MAKEUP	0	0	20	20
TOTAL				
REGULAR	1101	1142	849	3092
MAKEUP	2	16	135	153
COMBINED	1103	1158	984	3245

* Includes focused-BIB, intercorrelation, and document literacy samples.

Table 16-8

Numbers of Students Assessed and Excluded
by Sample and Age Class

	Age Class			
	9	13	17	Total
ASSESSED				
MAIN NAEP	23012	31601	32710	87323
BRIDGE TO 1984	5188	5500	4622	15310
BRIDGE TO 1986	3711	3942	7052	14705
CIVICS BRIDGE (TO 1976 AND 1982)	0	1938	1786	3724
EXCLUDED				
MAIN NAEP	1554	2213	1527	5294
BRIDGES	699	851	538	2088
TOTAL	34164	46045	48235	128444

* Includes focused-BIB, intercorrelation, and document literacy samples.

Table 16-9

Numbers of Students in Main Sample
by Type of Eligibility and Subgroup Classification, Grade 4/Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	16366	17137	10491	23012
SEX				
Male	8227	8708	4966	11969
Female	8139	8429	5525	11043
RACE/ETHNICITY				
White	9335	9645	6138	12842
Black	2831	3104	1823	4112
Hispanic	3336	3476	1956	4856
Other	864	912	574	1202
REGION				
Northeast	3099	3061	2163	3997
Southeast	4559	4897	2732	6724
Central	3349	3518	2127	4740
West	5359	5661	3469	7551
PARENT'S EDUCATION				
Less than High School	797	942	438	1301
High School	2300	2453	1411	3342
Greater High School	1147	1333	831	1649
Graduate lege	5769	6156	4042	7883
Unknown	6273	6163	3712	8724
SIZE AND TYPE OF COMMUNITY				
Rural	1422	1432	859	1995
Disadvantaged Urban	1917	2042	1178	2781
Advantaged Urban	1990	2078	1411	2657
Big City	2332	2679	1609	3402
Fringe	2316	2385	1593	3108
Medium City	2442	2500	1523	3419
Small Places	3947	4021	2318	5650

Table 16-10

Numbers of Students in Main Sample
by Type of Eligibility and Subgroup Classification, Grade 8/Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	22471	23801	14671	31601
SEX				
Male	10923	11804	6608	16119
Female	11548	11997	8063	15482
RACE/ETHNICITY				
White	14008	14719	9643	19084
Black	3702	4153	2248	5607
Hispanic	3659	3769	2046	5382
Other	1102	1160	734	1528
REGION				
Northeast	4584	4843	3319	6108
Southeast	6022	6484	3695	8811
Central	4943	5102	3115	6930
West	6922	7372	4542	9752
PARENT'S EDUCATION				
Less than High School	1890	2153	1015	3028
High School	5937	6328	3685	8580
Greater than High School	3951	4356	2895	5412
Graduated College	8401	8788	5932	11257
Unknown	2224	2100	1102	3222
SIZE AND TYPE OF COMMUNITY				
Rural	1331	1399	790	1940
Disadvantaged Urban	2634	2738	1525	3847
Advantaged Urban	2098	2325	1580	2843
Big City	3179	3223	1996	4406
Fringe	4465	4768	3267	5966
Medium City	3275	3408	2088	4595
Small Places	5489	5940	3425	8004

Table 16-11

Numbers of Students in Main Sample
by Type of Eligibility and Subgroup Classification, Grade 12/Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	25531	24316	17137	32710
SEX				
Male	12361	11571	7527	16405
Female	13170	12745	9610	16305
RACE/ETHNICITY				
White	17273	16670	12224	21719
Black	4645	4224	2808	6061
Hispanic	2770	2614	1576	3808
Other	843	808	529	1122
REGION				
Northeast	4719	4514	3447	5786
Southeast	7857	7293	5170	9980
Central	5869	5831	3916	7784
West	7086	6678	4604	9160
PARENT'S EDUCATION				
Less than High School	2398	2187	1280	3305
High School	6522	6014	4041	8495
Greater than High School	6235	5974	4440	7769
Graduated College	9581	9469	7004	12046
Unknown	704	582	322	964
SIZE AND TYPE OF COMMUNITY				
Rural	1131	1090	675	1546
Disadvantaged Urban	2647	2300	1500	3447
Advantaged Urban	3152	3360	2450	4062
Big City	4572	4380	3203	5749
Fringe	3357	3148	2257	4248
Medium City	2705	2572	1878	3399
Small Places	7967	7466	5174	10259

Table 16-12

Numbers of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification, Grade 4/Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3782	3979	2573	5188
SEX				
Male	1863	1984	1167	2680
Female	1919	1995	1406	2508
RACE/ETHNICITY				
White	2245	2305	1530	3020
Black	693	782	481	994
Hispanic	688	733	456	965
Other	156	159	106	209
REGION				
Northeast	969	1048	763	1254
Southeast	993	1127	642	1478
Central	801	783	479	1105
West	1019	1021	689	1351
PARENT'S EDUCATION				
Less than High School	159	191	102	248
High School	561	663	382	842
Greater than High School	189	195	130	254
Graduated College	1545	1679	1136	2088
Unknown	1321	1242	820	1743
SIZE AND TYPE OF COMMUNITY				
Rural	283	304	186	401
Disadvantaged Urban	346	396	211	531
Advantaged Urban	617	610	424	803
Big City	380	412	273	519
Fringe	576	592	433	735
Medium City	424	444	289	579
Small Places	1156	1221	757	1620

Table 16-13

Number of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification, Grade 8/Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	4005	4133	2638	5500
SEX				
Male	1964	2024	1160	2828
Female	2041	2109	1478	2672
RACE/ETHNICITY				
White	2890	2958	1956	3692
Black	576	619	357	838
Hispanic	358	378	214	522
Other	181	178	111	248
REGION				
Northeast	884	904	651	1137
Southeast	840	887	532	1195
Central	1199	1206	746	1659
West	1082	1136	709	1509
PARENT'S EDUCATION				
Less than High School	299	325	168	456
High School	1227	1271	785	1713
Greater than High School	414	466	309	571
Graduated College	1686	1724	1173	2237
Unknown	369	333	197	505
SIZE AND TYPE OF COMMUNITY				
Rural	257	240	166	331
Disadvantaged Urban	357	371	207	521
Advantaged Urban	608	655	440	823
Big City	2[illegible]1	234	174	381
Fringe	715	715	469	961
Medium City	375	395	281	489
Small Places	1422	1473	901	1994

Table 16-14

Numbers of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification, Grade 11/Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3652	3664	2694	4622
SEX				
Male	1667	1653	1179	2141
Female	1985	2011	1515	2481
RACE/ETHNICITY				
White	2577	2620	1995	3202
Black	638	631	424	845
Hispanic	281	275	177	379
Other	156	138	98	196
REGION				
Northeast	792	821	584	1029
Southeast	944	959	682	1221
Central	642	660	504	798
West	1274	1224	924	1574
PARENT'S EDUCATION				
Less than High School	297	284	174	407
High School	1051	1041	755	1337
Greater than High School	636	654	498	792
Graduated College	1564	1580	1211	1933
Unknown	96	99	53	142
SIZE AND TYPE OF COMMUNITY				
Rural	159	153	126	186
Disadvantaged Urban	51	62	32	81
Advantaged Urban	663	684	521	826
Big City	702	736	499	939
Fringe	632	628	485	775
Medium City	483	473	355	601
Small Places	962	928	676	1214

Table 16-15

Numbers of Students in Bridge to 1986 Sample
by Type of Eligibility and Subgroup Classification, Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3711	2498	2498	3711
SEX				
Male	1837	1160	1160	1837
Female	1874	1338	1338	1874
RACE/ETHNICITY				
White	2225	1560	1560	2225
Black	586	397	397	586
Hispanic	726	431	431	726
Other	174	110	110	174
REGION				
Northeast	946	759	759	946
Southeast	1043	639	639	1043
Central	801	507	507	801
West	921	593	593	921
PARENT'S EDUCATION				
Less than High School	141	87	87	[illegible]
High School	545	378	378	545
Greater than High School	281	209	209	281
Graduated College	1571	1134	1134	1571
Unknown	1145	679	679	1145
SIZE AND TYPE OF COMMUNITY				
Rural	279	180	180	279
Disadvantaged Urban	349	213	213	349
Advantaged Urban	636	406	406	636
Big City	350	249	249	350
Fringe	535	408	408	535
Medium City	411	269	269	411
Small Places	1151	773	773	1151

Table 16-16

Numbers of Students in Bridge to 1986 Sample
by Type of Eligibility and Subgroup Classification, Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3942	2545	2545	3942
SEX				
Male	1963	1193	1193	1963
Female	1979	1352	1352	1979
RACE/ETHNICITY				
White	2756	1822	1822	2756
Black	639	400	400	639
Hispanic	388	216	216	388
Other	159	107	107	159
REGION				
Northeast	905	632	632	905
Southeast	843	482	482	843
Central	1104	705	705	1104
West	1090	726	726	1090
PARENT'S EDUCATION				
Less than High School	278	152	152	278
High School	992	602	602	992
Greater than High School	619	443	443	619
Graduated College	1719	1170	1170	1719
Unknown	327	175	175	327
SIZE AND TYPE OF COMMUNITY				
Rural	203	130	130	203
Disadvantaged Urban	391	230	230	391
Advantaged Urban	554	392	392	554
Big City	313	213	213	313
Fringe	726	482	482	726
Medium City	382	275	275	382
Small Places	1373	823	823	1373

Table 16-17

Numbers of Students in Bridge to 1986 Sample
by Type of Eligibility and Subgroup Classification, Grade 11/Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	5581	5638	4167	7052
SEX				
Male	2660	2671	1939	3392
Female	2921	2967	2228	3660
RACE/ETHNICITY				
White	4077	4106	3178	5005
Black	888	895	591	1192
Hispanic	420	419	264	575
Other	196	218	134	280
REGION				
Northeast	1320	1349	978	1691
Southeast	1428	1400	1010	1818
Central	969	1054	802	1221
West	1864	1835	1377	2322
PARENT'S EDUCATION				
Less than High School	440	411	242	609
High School	1284	1283	916	1651
Greater than High School	1326	1341	1028	1639
Graduated College	2363	2426	1871	2918
Unknown	150	163	98	215
SIZE AND TYPE OF COMMUNITY				
Rural	221	241	177	285
Disadvantaged Urban	98	116	71	143
Advantaged Urban	1146	1182	922	1406
Big City	1033	1033	730	1336
Fringe	943	942	699	1186
Medium City	649	670	500	819
Small Places	1491	1454	1068	1877

Table 16-18

Numbers of Students in Civics Bridge
by Type of Eligibility and Subgroup Classification, Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	1938	1298	1298	1938
SEX				
Male	988	594	594	988
Female	950	704	704	950
RACE/ETHNICITY				
White	1374	942	942	1374
Black	247	166	166	247
Hispanic	224	129	129	224
Other	93	61	61	93
REGION				
Northeast	487	354	354	487
Southeast	318	197	197	31[illegible]
Central	543	338	338	543
West	590	409	409	590
PARENT'S EDUCATION				
Less than High School	119	64	64	119
High School	554	358	358	554
Greater than High School	325	233	233	325
Graduated College	791	567	567	791
Unknown	142	72	72	142
SIZE AND TYPE OF COMMUNITY				
Rural	102	62	62	102
Disadvantaged Urban	176	105	105	176
Advantaged Urban	317	228	228	317
Big City	150	102	102	150
Fringe	364	252	252	364
Medium City	193	143	143	193
Small Places	636	406	406	636

Table 16-19

Numbers of Students in Civics Bridge
by Type of Eligibility and Subgroup Classification, Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	1786	1333	1333	1786
SEX				
Male	818	591	591	818
Female	968	742	742	968
RACE/ETHNICITY				
White	1263	992	992	1263
Black	310	199	199	310
Hispanic	142	91	91	142
Other	71	51	51	71
REGION				
Northeast	378	281	281	378
Southeast	459	321	321	459
Central	337	279	279	337
West	612	452	452	612
PARENT'S EDUCATION				
Less than High School	154	100	100	154
High School	427	291	291	427
Greater than High School	434	328	328	434
Graduated College	725	590	590	725
Unknown	40	20	20	40
SIZE AND TYPE OF COMMUNITY				
Rural	91	73	73	91
Disadvantaged Urban	32	23	23	32
Advantaged Urban	331	268	268	331
Big City	348	245	245	348
Fringe	303	236	236	303
Medium City	226	164	164	226
Small Places	455	324	324	455

Table 16-20

Numbers of Excluded Students in Main and Bridge Samples
by Type of Eligibility and Subgroup Classification, Grade 4/Age 9

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	1480	1315	542	2253
SEX				
Male	899	817	299	1417
Female	581	497	243	835
RACE/ETHNICITY				
White	506	410	133	783
Black	308	243	83	468
Hispanic	567	537	282	822
Other	99	125	44	180
REGION				
Northeast	269	195	82	382
Southeast	344	288	66	566
Central	200	138	44	294
West	667	694	350	1011
SIZE AND TYPE OF COMMUNITY				
Rural	67	52	10	109
Disadvantaged Urban	304	279	129	454
Advantaged Urban	95	74	34	135
Big City	305	306	152	459
Fringe	239	209	99	349
Medium City	223	190	68	345
Small Places	247	205	50	402

Ta'le 16-21

Numbers of Excluded Students in Main and Bridge Samples
by Type of Eligibility and Subgroup Classification, Grade 8/Age 13

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	1754	1852	542	3064
SEX				
Male	1125	1161	317	1969
Female	579	649	191	1037
RACE/ETHNICITY				
White	810	908	250	1468
Black	418	404	78	744
Hispanic	381	389	144	626
Other	145	151	70	226
REGION				
Northeast	419	427	137	709
Southeast	385	405	78	712
Central	369	417	91	695
West	581	603	236	948
SIZE AND TYPE OF COMMUNITY				
Rural	89	99	21	167
Disadvantaged Urban	331	309	89	551
Advantaged Urban	95	109	42	162
Big City	268	250	47	471
Fringe	276	325	113	488
Medium City	295	321	131	485
Small Places	400	439	99	740

Table 16-22

Numbers of Excluded Students in Main Sample
by Type of Eligibility and Subgroup Classification, Grade 12/Age 17

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	1055	708	236	1527
SEX				
Male	684	417	126	975
Female	371	290	110	551
RACE/ETHNICITY				
White	476	366	119	723
Black	300	207	68	439
Hispanic	214	95	35	274
Other	65	40	14	91
REGION				
Northeast	201	139	61	279
Southeast	287	198	55	430
Central	225	133	45	313
West	342	238	75	505
SIZE AND TYPE OF COMMUNITY				
Rural	26	30	5	51
Disadvantaged Urban	264	127	52	339
Advantaged Urban	41	33	14	60
Big City	160	108	36	232
Fringe	120	93	29	184
Medium City	102	62	20	144
Small Places	342	255	80	517

Table 16-23

Numbers of Excluded Students in Bridge Samples
by Type of Eligibility and Subgroup Classification, Grade 11/Age 17

	Eligible by --			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	361	310	133	538
SEX				
Male	247	207	88	366
Female	114	103	45	172
RACE/ETHNICITY				
White	203	183	88	298
Black	70	60	22	108
Hispanic	56	26	9	73
Other	32	41	14	59
REGION				
Northeast	79	74	39	114
Southeast	83	63	25	121
Central	62	60	27	95
West	137	113	42	208
SIZE AND TYPE OF COMMUNITY				
Rural	19	15	8	26
Disadvantaged Urban	7	15	4	18
Advantaged Urban	48	50	32	66
Big City	44	22	11	55
Fringe	84	52	22	114
Medium City	59	67	27	99
Small Places	100	89	29	160

Table 16-24

Numbers of Students by Sample and Age Class

Sample	Age Class 9		Age Class 13		Age Class 17	
	Total	Sum of Weights	Total	Sum of Weights	Total	Sum of Weights
Main Assessment[a]						
Winter and spring combined	23012	4477209	31601	3737197	32710	4024260
Winter only	12293	4497709	16489	3724549	18542	4044142
Spring only	10719	4460873	15112	3749449	14168	3999942
Civics Bridge (to 1976 and 1982)[b]	--	--	1938	3041846	1786	3439421
Bridge to 1984[c]	5188	4477355	5500	4312127	4622	4487990
Bridge to 1986, Ages 9 and 13[b]						
Booklet 91	1274	3211177	1405	3049468	--	--
Booklet 92	1240	3207450	1281	3039867	--	--
Booklet 93	1197	3210115	1256	3068702	--	--
Bridge to 1986, Age 17[c]	--	--	--	--	7052	4487744
Excluded Students						
Main and bridge combined, grade 4/age 9, grade 8/age 13	2253	231907	3064	246018	--	--
Main, grade 12/age 17	--	--	--	--	1527	134353
Bridge, grade 11/age 17	--	--	--	--	538	171553

[a] Grade/age sample using calendar-year age definitions
[b] Age-only sample using previous age definitions
[c] Grade/age sample using previous age definitions

Table 16-25

Numbers of Students Assessed in Main Assessment, Grade 4/Age 9
(Booklets 1-22)

	GRADE < 4	GRADE = 4	GRADE > 4	Total
AGE < 9				
UNWEIGHTED N	0	129	0	129
WEIGHTED N	0	23915	0	23915
STANDARD ERROR	-	3325	-	3325
COEFF. OF VAR.*	-	13.90	-	13.90
AGE = 9				
UNWEIGHTED N	5804	10491	71	16366
WEIGHTED N	1204331	2006376	13867	3224573
STANDARD ERROR	8252	4239	2393	9133
COEFF. OF VAR.*	0.69	0.21	17.26	0.28
AGE > 9				
UNWEIGHTED N	0	6517	0	6517
WEIGHTED N	0	1228720	0	1228720
STANDARD ERROR	-	8759	-	8759
COEFF. OF VAR.*	-	0.71	-	0.71
AGE TOTAL				
UNWEIGHTED N	5804	17137	71	23012
WEIGHTED N	1204331	3259011	13867	4477209
STANDARD ERROR	8252	11015	2393	14194
COEFF. OF VAR.*	0.69	0.34	17.26	0.32

---

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-26

Numbers of Students Assessed in Main Assessment, Grade 8/Age 13
(Booklets 1-36)

	GRADE			
	< 8	= 8	> 8	Total
AGE < 13				
UNWEIGHTED N	0	192	0	192
WEIGHTED N	0	21787	0	21787
STANDARD ERROR	-	2368	-	2368
COEFF. OF VAR.*	-	10.87	-	10.87
AGE = 13				
UNWEIGHTED N	7709	14671	91	22471
WEIGHTED N	1032989	1601566	16935	2651490
STANDARD ERROR	9786	4670	4550	10115
COEFF. OF VAR.*	0.95	0.29	26.87	0.38
AGE > 13				
UNWEIGHTED N	0	8938	0	8938
WEIGHTED N	0	1063920	0	1063920
STANDARD ERROR	-	7365	-	7365
COEFF. OF VAR.*	-	0.69	-	0.69
AGE TOTAL				
UNWEIGHTED N	7709	23801	91	31601
WEIGHTED N	1032989	2687273	16935	3737197
STANDARD ERROR	9786	8444	4550	14086
COEFF. OF VAR.*	0.95	0.31	26.87	0.38

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-27

Numbers of Students Assessed in Main Assessment, Grade 12/Age 17
(Booklets 1-37)

	GRADE			
	< 12	= 12	> 12	Total
AGE < 17				
UNWEIGHTED N	0	286	0	286
WEIGHTED N	0	35112	0	35112
STANDARD ERROR	-	2823	-	2823
COEFF. OF VAR.*	-	8.04	-	8.04
AGE = 17				
UNWEIGHTED N	8394	17137	0	25531
WEIGHTED N	1053292	2080066	0	3133358
STANDARD ERROR	5236	14474	-	15779
COEFF. OF VAR.*	0.50	0.70	-	0.50
AGE > 17				
UNWEIGHTED N	0	6893	0	6893
WEIGHTED N	0	855789	0	855789
STANDARD ERROR	-	28000	-	28000
COEFF. OF VAR.*	-	3.27	-	3.27
AGE TOTAL				
UNWEIGHTED N	33[illegible]4	24316	0	32710
WEIGHTED N	105[illegible]292	2970968	0	4024260
STANDARD ERROR	5236	19439	-	20747
COEFF. OF VAR.*	0.50	0.65	-	0.52

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-28

Numbers of Students Assessed in Bridge to 1984, Grade 4/Age 9
(Booklets 51-56)

	GRADE			
	< 4	= 4	> 4	Total
AGE < 9				
UNWEIGHTED N	0	27	0	27
WEIGHTED N	0	25050	0	25050
STANDARD ERROR	-	5112	-	5112
COEFF. OF VAR.*	-	20.41	-	20.41
AGE = 9				
UNWEIGHTED N	1189	2573	20	3782
WEIGHTED N	1179250	2004609	17095	3200954
STANDARD ERROR	17494	6920	5230	20730
COEFF. OF VAR.*	1.48	0.35	30.59	0.65
AGE > 9				
UNWEIGHTED N	0	1379	0	1379
WEIGHTED N	0	1251351	0	1251351
STANDARD ERROR	-	15706	-	15706
COEFF. OF VAR.*	-	1.26	-	1.26
AGE TOTAL				
UNWEIGHTED N	1189	3979	20	5188
WEIGHTED N	1179250	3281010	17095	4477355
STANDARD ERROR	17494	17845	5230	31964
COEFF. OF VAR.*	1.48	0.54	30.59	0.71

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-29

Numbers of Students Assessed in Bridge to 1984, Grade 8/Age 13
(Booklets 51-56)

	GRADE < 8	GRADE = 8	GRADE > 8	Total
AGE < 13				
UNWEIGHTED N	0	31	0	31
WEIGHTED N	0	29904	0	29904
STANDARD ERROR	-	7127	-	7127
COEFF. OF VAR.*	-	23.83	-	23.83
AGE = 13				
UNWEIGHTED N	1352	2638	15	4005
WEIGHTED N	1176048	1836364	20738	3033150
STANDARD ERROR	19496	7661	15482	21237
COEFF. OF VAR.*	1.66	0.42	74.66	0.70
AGE > 13				
UNWEIGHTED N	0	1464	0	1464
WEIGHTED N	0	1249074	0	1249074
STANDARD ERROR	-	12959	-	12959
COEFF. OF VAR.*	-	1.04	-	1.04
AGE TOTAL				
UNWEIGHTED N	1352	4133	15	5500
WEIGHTED N	1176048	3115341	20738	4312127
STANDARD ERROR	19496	17554	15482	29734
COEFF. OF VAR.*	1.66	0.56	74.66	0.69

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-30

Numbers of Students Assessed in Bridge to 1984, Grade 11/Age 17
(Booklets 51-56)

	GRADE			
	< 11	= 11	> 11	Total
AGE < 17				
UNWEIGHTED N	0	399	0	399
WEIGHTED N	0	370779	0	370779
STANDARD ERROR	-	26139	-	26139
COEFF. OF VAR.*	-	7.05	-	7.05
AGE = 17				
UNWEIGHTED N	623	2694	335	3652
WEIGHTED N	806680	2218862	395088	3420630
STANDARD ERROR	40941	5164	44128	12266
COEFF. OF VAR.*	5.08	0.23	11.17	0.36
AGE > 17				
UNWEIGHTED N	0	571	0	571
WEIGHTED N	0	696581	0	696581
STANDARD ERROR	-	24053	-	24053
COEFF. OF VAR.*	-	3.45	-	3.45
AGE TOTAL				
UNWEIGHTED N	623	3664	335	4622
WEIGHTED N	806680	3286221	395088	4487990
STANDARD ERROR	40941	10847	44128	16310
COEFF. OF VAR.*	5.08	0.33	11.17	0.36

---

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-31

Numbers of Students Assessed in Bridge to 1986, Age 9

AGE 9	GRADE < 4	GRADE = 4	GRADE > 4	Total
Booklet 91				
UNWEIGHTED N	407	863	4	1274
WEIGHTED N	1073424	2130791	6961	3211177
STANDARD ERROR	49361	45194	4085	22224
COEFF. OF VAR.*	4.60	2.12	58.68	0.69
Booklet 92				
UNWEIGHTED N	402	833	5	1240
WEIGHTED N	1055792	2142834	8824	3207450
STANDARD ERROR	64622	59518	4542	20597
COEFF. OF VAR.*	6.12	2.78	51.47	0.64
Booklet 93				
UNWEIGHTED N	387	802	8	1197
WEIGHTED N	1087649	2094985	27482	3210115
STANDARD ERROR	63643	61277	8101	23096
COEFF. OF VAR.*	5.85	2.92	29.48	0.72

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-32

Numbers of Students Assessed in Bridge to 1986, Age 13

AGE 13	GRADE < 8	= 8	> 8	Total
Booklet 91				
UNWEIGHTED N	486	914	5	1405
WEIGHTED N	1050983	1986691	11795	3049468
STANDARD ERROR	47648	49544	6286	17394
COEFF. OF VAR.*	4.53	2.49	53.29	0.57
Booklet 92				
UNWEIGHTED N	464	805	12	1281
WEIGHTED N	1160841	1841841	37185	3039867
STANDARD ERROR	75205	81211	24652	17387
COEFF. OF VAR.*	6.48	4.41	66.30	0.57
Booklet 93				
UNWEIGHTED N	421	826	9	1256
WEIGHTED N	1069538	1972011	27153	3068702
STANDARD ERROR	78205	80834	16131	16454
COEFF. OF VAR.*	7.31	4.10	59.41	0.54

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-33

Numbers of Students Assessed in Bridge to 1986, Grade 11/Age 17
(Booklets 61-67)

	GRADE			
	< 11	= 11	> 11	Total
AGE < 17				
UNWEIGHTED N	0	640	0	640
WEIGHTED N	0	414914	0	414914
STANDARD ERROR	-	30365	-	30365
COEFF. OF VAR.*	-	7.32	-	7.32
AGE = 17				
UNWEIGHTED N	893	4167	521	5581
WEIGHTED N	811417	2218503	385821	3415741
STANDARD ERROR	37457	4912	39407	15624
COEFF. OF VAR.*	4.62	0.22	10.21	0.46
AGE > 17				
UNWEIGHTED N	0	831	0	831
WEIGHTED N	0	657089	0	657089
STANDARD ERROR	-	26348	-	26348
COEFF. OF VAR.*	-	4.01	-	4.01
AGE TOTAL				
UNWEIGHTED N	893	5638	521	7052
WEIGHTED N	811417	3290506	385821	4487744
STANDARD ERROR	37457	10395	39407	17884
COEFF. OF VAR.*	4.62	0.32	10.21	0.40

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-34

Numbers of Students Assessed in Civics Bridge to 1976 and 1982,
Age 13 and Age 17
(Booklet 90)

	GRADE			
AGE 13	< 8	= 8	> 8	Total
UNWEIGHTED N	635	1298	5	1938
WEIGHTED N	1010355	2024944	6547	3041846
STANDARD ERROR	53035	51919	4636	16956
COEFF. OF VAR.*	5.25	2.56	70.81	0.56

	GRADE			
AGE 17	< 11	= 11	> 11	Total
UNWEIGHTED N	300	1333	153	1786
WEIGHTED N	550974	2624633	263814	3439421
STANDARD ERROR	40700	50701	28587	9556
COEFF. OF VAR.*	7.39	1.93	10.84	0.28

* Coefficient of variation is defined as (100 times Standard Error divided by Weighted N).

Table 16-35

Weighted Percentage of Students in Main Sample
by Type of Eligibility and Subgroup Classification, Grade 4/Age 9

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	50.38	50.66	47.15	52.03
Female	49.62	49.34	52.85	47.97
RACE/ETHNICITY				
White	71.08	69.83	72.91	69.35
Black	14.51	15.25	13.51	15.50
Hispanic	10.69	10.91	9.36	11.44
Other	3.72	4.01	4.21	3.71
REGION				
Northeast	22.50	22.58	25.20	21.35
Southeast	24.96	25.73	23.25	26.29
Central	23.95	23.86	23.15	24.25
West	28.60	27.82	28.41	28.12
PARENT'S EDUCATION				
Less than High School	4.36	4.82	3.65	5.02
High School	14.43	14.77	13.86	14.93
Greater than High School	6.96	7.96	8.04	7.20
Graduated College	36.86	38.04	40.26	36.19
Unknown	37.08	34.05	33.84	36.32
SIZE AND TYPE OF COMMUNITY				
Rural	11.31	10.32	10.14	11.11
Disadvantaged Urban	8.27	8.57	7.82	8.69
Advantaged Urban	14.10	14.22	15.60	13.52
Big City	8.81	9.63	9.42	9.13
Fringe	13.40	13.02	14.34	12.70
Medium City	15.36	15.47	15.53	15.36
Small Places	28.74	28.77	27.14	29.48
ESTIMATED TOTAL POPULATION	3224573	3259011	2006376	4477209

Table 16-36

Weighted Percentage of Students in Main Sample
by Type of Eligibility and Subgroup Classification, Grade 8/Age 13

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.46	50.17	45.46	51.68
Female	50.54	49.83	54.54	48.32
RACE/ETHNICITY				
White	70.86	70.86	74.69	69.22
Black	14.65	14.97	12.88	15.64
Hispanic	10.57	10.13	8.44	11.17
Other	3.92	4.03	4.00	3.96
REGION				
Northeast	23.08	22.84	24.74	22.20
Southeast	23.75	24.46	22.28	24.89
Central	25.04	24.70	24.98	24.82
West	28.13	28.00	28.00	28.09
PARENT'S EDUCATION				
Less than High School	7.56	7.92	5.78	8.58
High School	26.76	26.89	25.48	27.40
Greater than High School	17.65	16.65	19.96	17.38
Graduated College	38.73	38.70	42.21	37.22
Unknown	9.03	7.57	6.32	9.14
SIZE AND TYPE OF COMMUNITY				
Rural	7.87	7.69	7.20	8.03
Disadvantaged Urban	9.35	9.03	7.96	9.71
Advantaged Urban	9.76	10.21	11.65	9.28
Big City	10.59	9.73	9.55	10.42
Fringe	19.12	19.34	21.46	18.27
Medium City	13.76	13.62	13.77	13.65
Small Places	29.55	30.38	28.41	30.64
ESTIMATED TOTAL POPULATION	2651490	2687273	1601566	3737197

Table 16-37

Weighted Percentage of Students in Main Sample
by Type of Eligibility and Subgroup Classification, Grade 12/Age 17

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	48.32	47.65	43.76	50.18
Female	51.68	52.35	56.24	49.82
RACE/ETHNICITY				
White	73.96	74.82	77.46	72.78
Black	14.52	13.41	12.14	14.93
Hispanic	8.10	7.80	6.66	8.62
Other	3.43	3.97	3.74	3.67
REGION				
Northeast	25.40	25.63	27.92	24.27
Southeast	24.03	23.01	22.84	23.89
Central	24.37	25.29	23.74	25.3
West	26.20	26.07	25.50	26.47
PARENT'S EDUCATION				
Less than High School	8.18	7.61	6.28	8.74
High School	25.21	24.27	23.13	25.59
Greater than High School	24.04	24.17	25.23	23.52
Graduated College	39.80	41.43	43.39	39.15
Unknown	2.38	2.16	1.65	2.59
SIZE AND TYPE OF COMMUNITY				
Rural	4.81	4.80	4.28	5.08
Disadvantaged Urban	8.71	7.41	6.67	8.81
Advantaged Urban	16.66	18.29	19.21	16.54
Big City	12.66	12.40	12.80	12.40
Fringe	13.65	13.40	13.53	13.53
Medium City	10.84	11.34	11.91	10.66
Small Places	32.66	32.37	31.61	32.99
ESTIMATED TOTAL POPULATION	3133358	2970968	2080066	4024260

Table 16-38

Weighted Percentage of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification, Grade 4/Age 9

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	50.27	50.68	45.89	52.53
Female	49.73	49.32	54.11	47.47
RACE/ETHNICITY				
White	71.09	69.92	72.96	69.39
Black	14.54	15.22	13.55	15.48
Hispanic	10.72	10.78	9.31	11.39
Other	3.66	4.08	4.17	3.74
REGION				
Northeast	22.41	23.41	25.61	21.71
Southeast	25.32	26.05	23.59	26.63
Central	23.82	23.08	22.31	23.95
West	28.45	27.45	28.49	27.70
PARENT'S EDUCATION				
Less than High School	4.56	5.18	4.28	5.14
High School	15.91	17.97	15.78	17.47
Greater than High School	4.95	4.89	5.12	4.83
Graduated College	40.15	41.99	44.13	39.72
Unknown	34.20	29.76	30.55	32.58
SIZE AND TYPE OF COMMUNITY				
Rural	9.74	9.69	9.24	9.92
Disadvantaged Urban	7.39	8.30	6.63	8.40
Advantaged Urban	15.71	13.69	14.98	14.56
Big City	7.29	7.80	7.71	7.47
Fringe	11.77	11.83	13.40	11.06
Medium City	13.91	14.17	14.31	13.92
Small Places	34.18	34.52	33.67	34.66
ESTIMATED TOTAL POPULATION	3200954	3281010	2004609	4477355

Table 16-39

Weighted Percentage of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification, Grade 8/Age 13

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.52	49.40	43.30	52.08
Female	50.48	50.60	56.70	47.92
RACE/ETHNICITY				
White	71.18	70.61	74.67	69.28
Black	14.39	14.90	12.79	15.44
Hispanic	10.31	10.28	8.40	11.10
Other	4.12	4.20	4.14	4.17
REGION				
Northeast	22.75	22.76	24.79	21.89
Southeast	23.67	23.73	22.15	24.36
Central	23.64	25.90	25.88	25.73
West	27.94	27.61	27.18	28.02
PARENT'S EDUCATION				
Less than High School	7.92	8.45	6.35	8.97
High School	30.93	31.36	30.88	31.26
Greater than High School	9.96	10.81	11.25	10.02
Graduated College	41.65	41.05	44.11	40.17
Unknown	9.24	7.99	7.19	9.21
SIZE AND TYPE OF COMMUNITY				
Rural	6.07	5.88	6.47	5.76
Disadvantaged Urban	7.33	7.01	5.76	7.76
Advantaged Urban	13.30	13.87	14.27	13.29
Big City	8.69	8.67	8.40	8.80
Fringe	14.70	14.00	14.63	14.22
Medium City	11.41	11.49	13.00	10.79
Small Places	38.51	39.08	37.45	39.38
ESTIMATED TOTAL POPULATION	3033150	3115341	1836364	4312127

Table 16-40

Weighted Percentage of Students in Bridge to 1984 Sample
by Type of Eligibility and Subgroup Classification, Grade 11/Age 17

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	47.85	46.94	45.46	48.36
Female	52.15	53.06	54.54	51.64
RACE/ETHNICITY				
White	73.04	73.65	78.17	70.96
Black	15.08	14.91	12.37	16.30
Hispanic	8.23	7.99	6.37	8.97
Other	3.64	3.44	3.09	3.77
REGION				
Northeast	22.71	22.58	23.26	22.34
Southeast	22.87	22.04	20 38	23.50
Central	26.31	27.12	28.06	26.04
West	28.11	28.25	28.31	28.12
PARENT'S EDUCATION				
Less than High School	8.92	8.18	6.34	9.35
High School	30.20	30.39	29.25	30.80
Greater than High School	17.81	17.70	19.12	17.08
Graduated College	40.39	40.91	43.48	39.25
Unknown	2.43	2.69	1.68	2.99
SIZE AND TYPE OF COMMUNITY				
Rural	7.01	6.95	7.26	6.84
Disadvantaged Urban	0.83	1.13	0.67	1.13
Advantaged Urban	16.00	17.04	17.88	15.84
Big City	16.86	17.39	15.79	17.78
Fringe	14.14	14.04	15.10	13.59
Medium City	14.87	13.99	14.69	14.32
Small Places	30.28	29.46	28.61	30.50
ESTIMATED TOTAL POPULATION	3420630	3286221	2218862	4487990

Table 16-41

Weighted Percentage of Students in Bridge to 1986 Sample by Type of Eligibility and Subgroup Classification, Age 9

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.53	47.03	47.03	49.53
Female	50.47	52.97	52.97	50.47
RACE/ETHNICITY				
White	71.17	73.29	73.29	71.17
Black	14.47	13.88	13.88	14.47
Hispanic	10.73	9.42	9.42	10.73
Other	3.63	3.42	3.42	3.63
REGION				
Northeast	23.82	29.49	29.49	23.82
Southeast	25.18	23.57	23.57	25.18
Central	24.00	20.60	20.60	24.00
West	27.00	26.33	26.33	27.00
PARENT'S EDUCATION				
Less than High School	3.69	3.48	3.48	3.69
High School	15.44	16.06	16.06	15.44
Greater than High School	7.57	8.80	8.80	7.57
Graduated College	42.51	46.28	46.28	42.51
Unknown	30.03	24.94	24.94	30.03
SIZE AND TYPE OF COMMUNITY				
Rural	9.60	8.54	8.54	9.60
Disadvantaged Urban	7.88	7.68	7.68	7.88
Advantaged Urban	14.71	13.50	13.50	14.71
Big City	7.01	7.53	7.53	7.01
Fringe	11.63	13.37	13.37	11.63
Medium City	15.01	14.82	14.82	15.01
Small Places	34.16	34.57	34.57	34.16
ESTIMATED TOTAL POPULATION	9628742	6368610	6368610	9628742

Table 16-42

Weighted Percentage of Students in Bridge to 1986 Sample by Type of Eligibility and Subgroup Classification, Age 13

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	50.49	47.27	47.27	50.49
Female	49.51	52.73	52.73	49.51
RACE/ETHNICITY				
White	70.90	72.59	72.59	70.90
Black	14.59	14.03	14.03	14.59
Hispanic	10.43	8.93	8.93	10.43
Other	4.08	4.45	4.45	4.08
REGION				
Northeast	24.11	26.61	26.61	24.11
Southeast	23.07	19.92	19.92	23.07
Central	26.07	25.52	25.52	26.07
West	26.75	27.95	27.95	26.75
PARENT'S EDUCATION				
Less than High School	7.54	6.27	6.27	7.54
High School	25.37	23.60	23.60	25.37
Greater than High School	15.24	17.37	17.37	15.24
Graduated College	43.46	46.01	46.01	43.46
Unknown	8.18	6.59	6.59	8.18
SIZE AND TYPE OF COMMUNITY				
Rural	5.01	5.07	5.07	5.01
Disadvantaged Urban	7.73	6.84	6.84	7.73
Advantaged Urban	12.64	13.65	13.65	12.64
Big City	8.80	9.39	9.39	8.80
Fringe	15.02	15.46	15.46	15.02
Medium City	11.84	13.67	13.67	11 84
Small Places	38.96	35.93	35.93	38.96
ESTIMATED TOTAL POPULATION	9158037	5800543	5800543	9158037

Table 16-43

Weighted Percentage of Students in Bridge to 1986 Sample
by Type of Eligibility and Subgroup Classification, Grade 11/Age 17

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.03	48.65	47.01	49.75
Female	50.97	51.35	52.99	50.25
RACE/ETHNICITY				
White	73.05	73.56	78.19	70.88
Black	15.11	14.90	12.35	16.31
Hispanic	8.23	8.00	6.38	8.98
Other	3.62	3.54	3.08	3.83
REGION				
Northeast	22.74	21.95	22.42	22.32
Southeast	23.29	21.80	20.95	23.35
Central	25.75	27.98	28.37	26.09
West	28.22	28.27	28.26	28.24
PARENT'S EDUCATION				
Less than High School	8.83	7.60	5.53	9.56
High School	25.21	25.02	23.86	25.74
Greater than High School	23.76	24.04	25.72	23.00
Graduated College	38.98	39.98	42.30	38.08
Unknown	2.73	3.03	2.18	3.22
SIZE AND TYPE OF COMMUNITY				
Rural	5.96	6.77	6.22	6.43
Disadvantaged Urban	1.00	1.20	0.96	1.17
Advantaged Urban	15.99	16.10	17.18	15.48
Big City	17.42	16.56	15.71	17.63
Fringe	14.55	14.10	14.24	14.37
Medium City	13.58	14.67	15.46	13.46
Small Places	31.49	30.60	30.24	31.46
ESTIMATED TOTAL POPULATION	3415741	3290506	2218503	4487744

Table 16-44

Weighted Percentage of Students in Civics Bridge
by Type of Eligibility and Subgroup Classification, Age 13

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	51.10	45.93	45.93	51.10
Female	48.90	54.07	54.07	48.90
RACE/ETHNICITY				
White	71.07	73.21	73.21	71.07
Black	14.26	13.90	13.90	14.26
Hispanic	10.54	8.75	8.75	10.54
Other	4.12	4.14	4.14	4.12
REGION				
Northeast	24.36	26.72	26.72	24.36
Southeast	21.57	20.33	20.33	21.57
Central	27.10	25.59	25.59	27.10
West	26.98	27.35	27.35	26.98
PARENT'S EDUCATION				
Less than High School	6.21	4.88	4.88	6.21
High School	28.65	27.40	27.40	28.65
Greater than High School	16.80	18.23	18.23	16.80
Graduated College	40.76	43.48	43.48	40.76
Unknown	7.22	5.69	5.69	7.22
SIZE AND TYPE OF COMMUNITY				
Rural	5.23	4.85	4.85	5.23
Disadvantaged Urban	7.18	6.37	6.37	7.18
Advantaged Urban	15.64	16.64	16.64	15.64
Big City	9.79	9.87	9.87	9.79
Fringe	14.20	14.26	14.26	14.20
Medium City	11.09	12.57	12.57	11.09
Small Places	36.88	35.45	35.45	36.88
ESTIMATED TOTAL POPULATION	3041846	2024944	2024944	3041846

Table 16-45

Weighted Percentage of Students in Civics Bridge
by Type of Eligibility and Subgroup Classification, Age 17

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	45.90	44.35	44.35	45.90
Female	54.10	55.65	55.65	54.10
RACE/ETHNICITY				
White	73.03	75.73	75.73	73.03
Black	15.12	13.44	13.44	15.12
Hispanic	8.24	7.36	7.36	8.24
Other	3.61	3.48	3.48	3.61
REGION				
Northeast	22.41	21.61	21.61	22.41
Southeast	22.60	20.62	20.62	22.60
Central	26.85	30.48	30.48	26.85
West	28.14	27.30	27.30	28.14
PARENT'S EDUCATION				
Less than High School	9.09	8.03	8.03	9.09
High School	24.67	23.30	23.30	24.67
Greater than High School	25.41	26.19	26.19	25.41
Graduated College	38.65	40.97	40.97	38.65
Unknown	1.92	1.27	1.27	1.92
SIZE AND TYPE OF COMMUNITY				
Rural	6.86	7.27	7.27	6.86
Disadvantaged Urban	0.91	0.85	0.85	0.91
Advantaged Urban	16.35	17.35	17.35	16.35
Big City	16.85	16.51	16.51	16.85
Fringe	14.11	14.19	14.19	14.11
Medium City	14.47	14.27	14.27	14.47
Small Places	30.46	29.55	29.55	30.46
ESTIMATED TOTAL POPULATION	3439421	2624633	2624633	3439421

Table 16-46

Weighted Percentage of Excluded Students, by Type of Eligibility and Subgroup Classification in Grade 4/Age 9 Main and Bridge Samples

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	65.65	66.67	61.88	66.83
Female	34.35	33.25	38.12	33.13
RACE/ETHNICITY				
White	53.56	51.38	45.73	53.73
Black	24.46	22.01	19.26	24.04
Hispanic	16.98	18.73	26.25	16.34
Other	5.01	7.88	8.76	5.90
REGION				
Northeast	21.44	19.73	20.42	20.71
Southeast	26.81	26.18	16.47	28.22
Central	21.76	16.77	15.11	20.24
West	30.00	37.32	47.99	30.84
SIZE AND TYPE OF COMMUNITY				
Rural	7.08	7.44	3.96	7.79
Disadvantaged Urban	13.82	12.88	14.01	13.29
Advantaged Urban	8.42	8.94	10.00	8.43
Big City	14.67	16.61	19.21	14.93
Fringe	14.63	14.89	20.53	13.77
Medium City	15.19	15.73	15.72	15.39
Small Places	26.21	23.51	16.56	26.41
ESTIMATED TOTAL POPULATION	148361	22555	39008	231907

Table 16-47

Weighted Percentage of Excluded Students, by Type of Eligibility and Subgroup Classification in Grade 8/Age 13 Main and Bridge Samples

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	66.24	63.81	60.22	65.77
Female	31.64	34.44	34.44	32.80
RACE/ETHNICITY				
White	52.93	57.84	55.74	55.28
Black	25.65	21.10	14.84	24.72
Hispanic	15.12	14.64	18.71	14.31
Other	6.30	6.41	10.71	5.69
REGION				
Northeast	27.07	25.78	27.14	26.33
Southeast	22.64	21.45	14.23	23.24
Central	21.42	23.77	18.62	23.18
West	28.87	28.99	40.01	27.25
SIZE AND TYPE OF COMMUNITY				
Rural	6.43	7.81	4.56	7.49
Disadvantaged Urban	16.82	14.08	13.31	15.80
Advantaged Urban	6.21	6.11	8.94	5.74
Big City	16.76	12.12	7.46	15.55
Fringe	13.10	16.67	20.46	14.01
Medium City	14.35	14.92	22.42	13.45
Small Places	26.33	28.29	22.85	27.97
ESTIMATED TOTAL POPULATION	144182	139191	37354	246018

Table 16-48

Weighted Percentage of Excluded Students, by Type of Eligibility and Subgroup Classification in Grade 12/Age 17 Main Sample

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	63.62	57.14	51.77	62.47
Female	36.38	42.80	48.23	37.50
RACE/ETHNICITY				
White	54.33	60.04	61.22	55.94
Black	25.73	23.43	22.20	25.21
Hispanic	13.74	9.59	8.63	12.59
Other	6.21	6.94	7.95	6.27
REGION				
Northeast	25.83	24.33	32.34	24.01
Southeast	25.64	25.63	20.88	26.44
Central	22.94	20.06	22.29	21.65
West	25.59	29.98	24.49	27.90
SIZE AND TYPE OF COMMUNITY				
Rural	3.02	4.92	1.75	4.16
Disadvantaged Urban	19.23	14.38	17.53	17.17
Advantaged Urban	5.18	6.18	7.01	5.35
Big City	10.21	9.77	9.13	10.18
Fringe	13.25	13.75	14.09	13.35
Medium City	9.95	8.71	9.38	9.44
Small Places	39.16	42.29	41.10	40.35
ESTIMATED TOTAL POPULATION	91954	64970	22571	134353

Table 16-49

Weighted Percentage of Excluded Students, by Type of Eligibility and Subgroup Classification in Grade 11/Age 17 Bridge Sample

	ELIGIBLE BY			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	69.45	67.67	65.99	69.13
Female	30.55	32.33	34.01	30.87
RACE/ETHNICITY				
White	64.83	64.36	71.56	63.32
Black	16.47	17.10	13.26	17.41
Hispanic	13.61	8.22	7.71	11.79
Other	5.10	10.32	7.47	7.49
REGION				
Northeast	21.99	21.88	33.17	19.84
Southeast	21.72	17.67	17.18	20.37
Central	28.36	32.62	26.27	31.07
West	27.92	27.83	23.38	28.72
SIZE AND TYPE OF COMMUNITY				
Rural	10.64	8.45	11.09	9.36
Disadvantaged Urban	1.04	2.90	1.70	1.93
Advantaged Urban	10.94	13.29	21.97	10.15
Big City	7.29	3.54	4.18	5.84
Fringe	16.02	9.56	10.14	13.61
Medium City	19.53	22.95	23.60	20.62
Small Places	34.54	39.30	27.32	38.48
ESTIMATED TOTAL POPULATION	110521	93092	32060	171553

Table 16-50

Numbers of Students in Main Sample with Proficiency Scores by Type of Eligibility, Grade 4/Age 9

	ELIGIBLE BY			
Subject Area/Sample	Age	Grade	Age & Grade	Age or Grade
Reading Focused-BIB	4423	4534	2780	6177
Reading Intercorrelation	1882	1957	1201	2638
Writing Focused-BIB and Long Writing	6331	6679	4129	8881
Civics Focused-BIB	1858	1974	1180	2652
Civics Intercorrelation	1882	1957	1201	2638
History Focused-BIB	1872	1993	1201	2664
History Intercorrelation	1882	1957	1201	2638
Geography Focused-BIB	0	0	0	0
Geography Intercorrelation	0	0	0	0
TOTAL	16366	17137	10491	23012

Table 16-51

Numbers of Students in Main Sample with Proficiency Scores by Type of Eligibility, Grade 8/Age 13

Subject Area/Sample	ELIGIBLE BY Age	Grade	Age & Grade	Age or Grade
Reading Focused-BIB	4250	4404	2742	5912
Reading Intercorrelation	1819	1943	1172	2590
Writing Focused-BIB and Long Writing	6059	6525	3987	8597
Civics Focused-BIB	4249	4487	2755	5981
Civics Intercorrelation	1819	1943	1172	2590
History Focused-BIB	4303	4519	2834	5988
History Intercorrelation	1819	1943	1172	2590
Geography Focused-BIB	0	0	0	0
Geography Intercorrelation	0	0	0	0
TOTAL	20680	21878	13490	29068

Table 16-52

umbers of Students in Main Sample with Proficiency Scores
by Type of Eligibility, Grade 12/Age 17

Subject Area/Sample	ELIGIBLE BY Age	Grade	Age & Grade	Age or Grade
Reading Focused-BIB	4509	4250	2991	5768
Reading Intercorrelation	1879	1851	1292	2438
Writing Focused-BIB and Long Writing	6360	6069	4259	8170
Civics Focused-BIB	4416	4275	3008	5683
Civics Intercorrelation	1249	1233	861	1621
History Focused-BIB	4561	4268	3049	5780
History Intercorrelation	1266	1239	873	1632
Geography Focused-BIB	1912	1800	1266	2446
Geography Intercorrelation	1243	1230	850	1623
TCTAL	23637	22513	15865	30285

Table 16-53

Numbers of Students in Bridge to 1984 Sample with Proficiency Scores by Type of Eligibility, Grade 4/Age 9

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Reading	3782	3979	2573	5188
Writing	2154	3327	2154	3327
TOTAL	3782	3979	2573	5188

Table 16-54

Numbers of Students in Bridge to 1984 Sample with Proficiency Scores by Type of Eligibility, Grade 8/Age 13

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Reading	4005	4133	2638	5500
Writing	2638	4133	2638	4133
TOTAL	4005	4133	2638	5500

Table 16-55

Numbers of Students in Bridge to 1984 Sample with Proficiency Scores by Type of Eligibility, Grade 11/Age 17

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Reading	3652	3664	2694	4622
Writing	2694	3664	2694	[illegible]564
TOTAL	3652	3664	2694	4622

Table 16-56

Numbers of Students in Bridge to 1986 Sample with Proficiency Scores by Type of Eligibility, Age 9

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Reading	3711	2498	2498	3711
Mathematics	3711	2498	2498	3711
Science	3711	2498	2498	3711
TOTAL	3711	2498	2498	3711

Table 16-57

Numbers of Students in Bridge to 1986 Sample with Proficiency Scores by Type of Eligibility, Age 13

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Reading	3942	2545	2545	3942
Mathematics	3942	2545	2545	3942
Science	3942	2545	2545	3942
TOTAL	3942	2545	2545	3942

Table 16-58

Numbers of Students in Bridge to 1986 Sample with Proficiency Scores by Type of Eligibility, Grade 11/Age 17

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Reading	0	0	0	0
History	0	0	0	0
Mathematics	1852	1360	1360	1852
Science	1862	1389	1389	1862
TOTAL	3105	2304	2304	3105

Table 16-59

Numbers of Students in Civics Bridge with Proficiency Scores by Type of Eligibility, Age 13

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Civics	1938	1298	1298	1938
TOTAL	1938	1298	1298	1938

Table 16-60

Numbers of Students in Civics Bridge with Proficiency Scores
by Type of Eligibility, Age 17

	ELIGIBLE BY			
Subject Area	Age	Grade	Age & Grade	Age or Grade
Civics	1786	1333	1333	1786
TOTAL	1786	1333	1333	1786

Table 16-61

Weighted Proficiency Means, Standard Deviations, and Percentiles with Standard Errors for Main Focused-BIB Reading Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
TOTAL SAMPLE			
Mean	230.4 (1.1)	262.8 (1.0)	287.1 (0.8)
Standard Deviation	41.4 (0.7)	37.3 (0.6)	34.8 (0.6)
Percentiles			
5	156.4 (1.3)	195.0 (2.6)	224.9 (2.4)
10	174.2 (2.3)	211.6 (1.4)	241.1 (1.4)
25	204.3 (2.0)	239.8 (1.6)	265.9 (1.3)
50	233.6 (1.2)	266.0 (1.1)	290.0 (1.0)
75	259.0 (1.1)	288.6 (1.7)	310.8 (1.0)
90	280.3 (0.9)	308.2 (1.0)	329.1 (1.3)
95	293.2 (1.2)	320.1 (1.8)	340.0 (1.8)
MALE STUDENTS			
Mean	226.6 (1.5)	256.4 (1.3)	282.9 (1.1)
Standard Deviation	43.6 (1.0)	39.4 (0.8)	36.8 (0.8)
Percentiles			
5	148.7 (2.5)	186.0 (3.7)	214.6 (2.3)
10	166.5 (4.4)	201.0 (1.9)	232.8 (1.9)
25	198.7 (2.7)	230.7 (1.9)	259.7 (2.5)
50	229.5 (1.4)	259.9 (2.1)	286.2 (1.5)
75	257.1 (2.2)	284.2 (1.7)	307.9 (1.4)
90	279.3 (1.9)	304.7 (1.9)	327.3 (0.9)
95	293.2 (2.3)	317.3 (1.8)	338.1 (2.2)
FEMALE STUDENTS			
Mean	234.4 (1.2)	269.6 (1.1)	291.0 (1.1)
Standard Deviation	38.5 (0.6)	33.8 (0.7)	32.4 (0.8)
Percentiles			
5	167.2 (2.0)	211.3 (1.6)	235.6 (4.2)
10	182.8 (1.5)	225.0 (1.6)	249.3 (1.5)
25	210.1 (1.0)	248.5 (1.2)	270.6 (1.4)
50	237.6 (1.1)	271.6 (1.2)	292.8 (1.3)
75	260.6 (1.1)	292.2 (1.0)	313.2 (1.2)
90	281.3 (1.7)	310.6 (1.2)	330.5 (1.7)
95	292.9 (1.6)	322.4 (2.2)	341.3 (3.2)

Table 16-61 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles with Standard Errors for Main Focused-BIB Reading Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
WHITE STUDENTS			
Mean	238.1 (1.2)	269.2 (1.3)	292.6 (1.0)
Standard Deviation	39.3 (0.8)	36.2 (0.7)	33.4 (0.6)
Percentiles			
5	167.1 (2.9)	202.3 (5.7)	233.3 (3.6)
10	186.0 (3.6)	219.8 (2.7)	249.2 (1.9)
25	213.9 (1.3)	248.1 (1.7)	272.6 (1.2)
50	240.8 (1.7)	272.6 (1.3)	295.1 (1.0)
75	264.4 (1.1)	293.6 (0.7)	315.3 (1.3)
90	285.2 (1.6)	312.6 (1.3)	332.5 (0.7)
95	297.8 (2.7)	323.7 (1.5)	343.6 (2.2)
BLACK STUDENTS			
Mean	210.6 (1.9)	245.7 (2.0)	270.3 (1.6)
Standard Deviation	39.1 (1.0)	32.8 (1.2)	31.1 (1.2)
Percentiles			
5	143.0 (4.2)	187.4 (4.0)	213.9 (3.5)
10	158.4 (4.7)	201.2 (3.7)	228.8 (3.2)
25	184.7 (3.4)	225.5 (2.3)	250.3 (3.5)
50	212.0 (2.5)	248.2 (2.0)	272.9 (1.4)
75	238.8 (2.0)	267.0 (2.8)	292.0 (1.7)
90	258.0 (1.9)	284.9 (2.6)	307.1 (2.5)
95	270.5 (3.4)	297.2 (3.8)	317.5 (4.4)
HISPANIC STUDENTS			
Mean	209.9 (2.4)	243.6 (2.1)	267.1 (2.4)
Standard Deviation	41.5 (1.3)	37.6 (1.3)	35.7 (1.5)
Percentiles			
5	134.8 (7.6)	177.6 (4.3)	200.8 (7.0)
10	153.2 (10.4)	190.9 (4.3)	218.1 (3.3)
25	182.5 (4.0)	217.6 (4.3)	244.3 (4.3)
50	213.4 (1.9)	246.9 (3.4)	269.9 (2.3)
75	240.3 (2.9)	271.5 (2.5)	292.7 (1.9)
90	260.1 (6.4)	288.2 (2.5)	310.8 (6.5)
95	272.0 (4.1)	300.6 (3.1)	321.0 (2.4)

Table 16-62

Weighted Proficiency Means, Standard Deviations, and Percentiles with Standard Errors for Bridge to 1984 Reading Samples, by Age and Subgroup

	Age 9	Age 13	Age 17
TOTAL SAMPLE			
Mean	211.8 (1.2)	257.5 (0.9)	290.1 (1.1)
Standard Deviation	41.2 (1.0)	34.7 (0.4)	37.1 (0.7)
Percentiles			
5	142.0 (3.6)	199.7 (1.6)	226.2 (1.3)
10	156.8 (2.0)	213.0 (1.2)	241.5 (2.4)
25	184.3 (1.7)	234.3 (1.2)	265.8 (1.8)
50	213.7 (1.4)	257.9 (1.0)	291.1 (1.8)
75	240.1 (1.3)	281.4 (1.4)	316.0 (1.4)
90	263.0 (1.7)	301.6 (1.0)	336.9 (2.1)
95	277.5 (1.9)	313.7 (1.3)	348.7 (1.7)
MALE STUDENTS			
Mean	207.5 (1.5)	251.8 (1.2)	286.0 (1.5)
Standard Deviation	42.7 (1.1)	35.3 (0.6)	37.5 (1.1)
Percentiles			
5	136.6 (2.9)	192.7 (2.7)	222.0 (2.6)
10	151.1 (2.3)	206.8 (1.6)	236.3 (3.6)
25	178.4 (1.7)	227.8 (2.0)	261.6 (1.7)
50	209.8 (1.8)	252.2 (2.1)	287.0 (2.2)
75	237.1 (1.7)	276.5 (2.0)	312.0 (3.4)
90	260.4 (2.0)	297.2 (1.5)	333.4 (2.0)
95	275.1 (2.3)	309.4 (2.8)	345.6 (4.1)
FEMALE STUDENTS			
Mean	216.3 (1.4)	263.0 (1.0)	293.8 (1.6)
Standard Deviation	39.2 (1.1)	33.1 (0.5)	36.3 (0.9)
Percentiles			
5	149.4 (5.2)	207.4 (3.8)	231.8 (3.3)
10	164.4 (4.8)	221.1 (1.4)	246.6 (4.9)
25	190.6 (2.4)	240.1 (1.6)	270.2 (2.0)
50	217.5 (1.9)	263.0 (1.3)	294.6 (2.2)
75	242.6 (1.0)	285.8 (1.0)	319.4 (1.5)
90	265.3 (2.1)	305.2 (1.1)	339.8 (1.7)
95	279.2 (3.3)	317.7 (3.2)	351.7 (2.7)

Table 16-62 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Bridge to 1984 Reading Samples, by Age and Subgroup

	Age 9	Age 13	Age 17
WHITE STUDENTS			
Mean	217.7 (1.5)	261.3 (1.0)	294.7 (1.3)
Standard Deviation	39.3 (1.0)	33.9 (0.5)	36.0 (0.7)
Percentiles			
5	150.4 (3.4)	204.2 (1.4)	232.7 (1.0)
10	165.1 (3.9)	217.2 (1.9)	247.4 (3.8)
25	191.8 (2.4)	238.4 (1.0)	271.4 (1.7)
50	219.1 (1.2)	262.2 (1.1)	295.4 (1.6)
75	244.3 (1.8)	285.1 (0.9)	319.9 (1.9)
90	266.8 (2.2)	304.2 (1.4)	339.8 (1.5)
95	280.6 (2.5)	315.8 (1.1)	351.6 (2.9)
BLACK STUDENTS			
Mean	188.5 (2.6)	242.9 (2.3)	274.4 (2.6)
Standard Deviation	39.4 (1.5)	32.1 (1.3)	35.9 (1.3)
Percentiles			
5	124.7 (6.3)	190.6 (3.1)	214.5 (9.5)
10	138.3 (3.3)	202.2 (3.1)	227.8 (4.2)
25	161.8 (2.3)	222.0 (2.4)	250.5 (2.3)
50	188.3 (3.9)	242.4 (2.7)	274.3 (3.5)
75	216.5 (2.8)	263 (4.4)	299.[illegible] (3.0)
90	238.2 (3.7)	283.6 (4.7)	321.0 (3.8)
95	252.2 (4.3)	298.9 (2.1)	333.1 (4.8)
HISPANIC STUDENTS			
Mean	193.7 (3.9)	240.1 (3.5)	270.8 (4.0)
Standard Deviation	41.5 (2.6)	34.6 (2.2)	37.7 (2.0)
Percentiles			
5	121.9 (10.8)	181.7 (8.7)	204.2 (11.5)
10	140.3 (7.3)	194.6 (3.7)	218.0 (6.9)
25	164.9 (5.0)	213.9 (6.0)	246.4 (5.5)
50	196.0 (3.3)	240.3 (3.9)	273.6 (5.0)
75	222.0 (6.0)	262.0 (5.2)	297.9 (7.0)
90	246.7 (7.9)	284.0 (8.4)	315.9 (18.0)
95	258.6 (11.3)	297.3 (9.9)	328.0 (8.6)

Table 16-63

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Main Writing Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
TOTAL SAMPLE			
Mean	190.9 (1.0)	209.5 (0.9)	224.2 (1.3)
Standard Deviation	42.3 (0.7)	43.8 (0.6)	49.2 (0.8)
Percentiles			
5	119.3 (1.2)	137.1 (2.7)	142.3 (2.0)
10	136.3 (1.3)	153.7 (1.2)	161.1 (2.0)
25	163.5 (1.0)	180.3 (1.1)	191.9 (1.5)
50	192.0 (1.3)	209.9 (1.3)	224.9 (1.3)
75	219.7 (1.2)	238.9 (1.5)	257.4 (1.9)
90	243.9 (1.4)	264.9 (1.3)	286.4 (2.6)
95	258.4 (1.5)	280.4 (2.0)	303.7 (2.7)
MALE STUDENTS			
Mean	184.4 (1.3)	200.9 (1.4)	212.5 (1.4)
Standard Deviation	41.4 (0.9)	43.4 (0.9)	48.6 (1.1)
Percentiles			
5	113.5 (2.7)	129.3 (4.2)	131.7 (2.3)
10	130.2 (2.0)	145.6 (2.6)	151.0 (1.6)
25	157.5 (1.6)	172.1 (1.5)	180.6 (1.9)
50	186.1 (1.1)	201.5 (1.3)	213.7 (1.5)
75	212.9 (1.7)	230.4 (1.4)	245.5 (1.7)
90	236.5 (1.8)	255.8 (2.3)	273.2 (2.3)
95	249.8 (1.8)	271.1 (2.5)	290.4 (3.7)
FEMALE STUDENTS			
Mean	197.5 (1.3)	218.5 (1.2)	234.8 (1.8)
Standard Deviation	42 2 (0.7)	42.4 (0.8)	47.3 (1.1)
Percentiles			
5	127.2 (2.6)	148.8 (2.2)	156.0 (3.3)
10	143.0 (1.6)	164.5 (1.5)	174.7 (2.5)
25	169.9 (1.5)	190.2 (1.3)	203.4 (1.1)
50	198.1 (1.5)	219.0 (1.8)	235.6 (2.8)
75	226.3 (1.7)	247.2 (1.6)	267.1 (1.9)
90	251.0 (1.6)	272.1 (2.3)	295.2 (4.0)
95	265.4 (4.7)	287.7 (4.4)	311.5 (5.2)

Table 16-63 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Main Writing Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
WHITE STUDENTS			
Mean	197.6 (1.3)	216.0 (1.0)	230.5 (1.5)
Standard Deviation	40.2 (0.7)	42.2 (0.7)	47.9 (1.1)
Percentiles			
5	130.5 (2.1)	145.9 (3.2)	150.8 (2.8)
10	146.0 (1.4)	162.0 (1.7)	169.5 (3.2)
25	171.7 (1.2)	188.1 (1.3)	199.5 (1.7)
50	198.1 (1.6)	216.4 (1.4)	231.4 (1.2)
75	224.8 (2.0)	244 8 (1.1)	262.9 (1.7)
90	248.4 (2.7)	269.4 (1.8)	290.5 (3.0)
95	261.8 (1.9)	284.2 (2.8)	307.4 (2.1)
BLACK STUDENTS			
Mean	168.[illegible] (1.9)	187.5 (2.2)	200.7 (2.3)
Standard Deviation	[illegible] (1.1)	41.3 (1.3)	45.7 (1.8)
Percentiles			
5	[illegible] 3 (3.2)	119.3 (4.1)	125.3 (6.2)
10	1[illegible].0 (3.2)	136.4 (3.5)	142.1 (3.0)
25	141.5 (2.3)	159.8 (2.8)	169.9 (3.8)
50	169.6 (2.1)	186.8 (3.6)	200.4 (3.9)
75	196.8 (2.4)	215.8 (3.7)	231.9 (3.1)
90	221.5 (3.4)	239.8 (3.1)	258.1 (3.6)
95	237.0 (3.5)	255.6 (4.8)	275.1 (9.2)
HISPANIC STUDENTS			
Mean	178.2 (2.0)	192.4 (2.1)	204.9 (3.5)
Standard Deviation	42.3 (1.3)	44.1 (1.4)	50.2 (2.8)
Percentiles			
5	106.0 (4.5)	118.4 (4.9)	121.5 ([illegible].8)
10	123.1 (3.6)	135.5 (3.5)	141.0 (5.7)
25	150.6 (2.4)	163.3 (2.5)	172.9 (4.9)
50	179.6 (1.9)	192.9 (2.3)	204.8 (3.9)
75	206.9 (3.2)	223.2 (3.3)	237.5 (6.9)
90	231.0 (3.8)	248.5 (5.3)	269.6 (8.3)
95	244.6 (4.1)	263.5 (4.0)	290.4 (11.8)

Table 16-64

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Bridge to 1984 Writing Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 11
TOTAL SAMPLE			
Mean	173.3 (1.3)	208.2 (0.8)	220.7 (1.2)
Standard Deviation	43.4 (1.1)	40.1 (0.9)	39.1 (1.2)
Percentiles			
5	101.4 (2.5)	140.9 (1.5)	154.6 (2.7)
10	116.9 (2.0)	156.6 (2.8)	170.7 (2.1)
25	143.6 (1.6)	181.5 (1.2)	195.8 (1.8)
50	174.4 (2.2)	208.9 (1.1)	221.6 (1.4)
75	203.5 (2.0)	235.6 (1.1)	247.3 (2.1)
90	228.5 (2.5)	258.7 (1.7)	269.6 (2.1)
95	243.3 (3.2)	273.3 (1.9)	283.2 (2.6)
MALE STUDENTS			
Mean	164.3 (1.9)	197.9 (1.4)	211.[illegible] (1.6)
Standard Deviation	42.5 (1.2)	39.5 (1.2)	39.1 (1.5)
Percentiles			
5	93.7 (2.5)	131.5 (3.5)	145.0 (3.4)
10	109.7 (2.5)	146.3 (1.8)	161.3 (4.2)
25	135.0 (1.7)	171.7 (2.0)	186.5 (1.9)
50	165.6 (2.4)	198.3 (1.7)	2[illegible]1 (1.5)
75	194.4 (2.6)	225.1 (2.4)	2[illegible].7 (2.2)
90	218.7 (3.5)	248.3 (2.0)	260.4 (4.0)
95	232.7 (4.9)	261.2 (2.5)	274.0 (5.3)
FEMALE STUDENTS			
Mean	182.4 (1.6)	218.2 (1.1)	229.2 (1.4)
Standard Deviation	42.5 (1.4)	38.2 (1.3)	37.1 (1.2)
Percentiles			
5	110.6 (4.0)	154.4 (4.4)	167.8 (3.8)
10	127.3 (2.4)	169.5 (3.4)	182.9 (2.8)
25	154.1 (2.1)	192.8 (2.6)	205.2 (2.1)
50	183.0 (1.8)	218.6 (2.3)	229.9 (0.9)
75	211.7 (2.0)	244.0 (1.5)	254.5 (2.9)
90	236.6 (2.3)	267.1 (2.5)	275.6 (2.9)
95	251.9 (5.6)	280.0 (3.9)	288.3 (3.6)

Table 16-64 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Bridge to 1984 Writing Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 11
WHITE STUDENTS			
Mean	180.0 (1.6)	213.1 (1.0)	225.3 (1.3)
Standard Deviation	41.8 (1.1)	39.5 (0.9)	37.9 (1.4)
Percentiles			
5	110.2 (2.7)	146.4 (1.7)	161.5 (5.0)
10	126.0 (2.1)	162.3 (2.8)	176.8 (2.1)
25	151.6 (1.7)	187.6 (1.7)	200.8 (1.9)
50	181.0 (2.5)	213.7 (1.3)	226.2 (1.2)
75	208.7 (2.0)	240.1 (1.6)	250.7 (2.1)
90	232.6 (3.4)	262.2 (2.4)	273.0 (2.8)
95	247.6 (4.5)	277.2 (2.5)	286.0 (2.3)
BLACK STUDENTS			
Mean	150.7 (3.1)	190.1 (2.3)	206.9 (2.6)
Standard Deviation	42.4 (1.6)	37.7 (1.8)	38.0 (1.7)
Percentiles			
5	81.1 (8.4)	127.2 (3.8)	143.2 (6.5)
10	96.0 (5.8)	141.4 (5.8)	158.0 (3.2)
25	121.8 (5.1)	165.2 (2.4)	182.6 (4.1)
50	150.5 (2.8)	189.9 (2.4)	206.7 (2.5)
75	179.2 (2.8)	215.6 (2.8)	232.1 (4.1)
90	206.2 (4.4)	238.5 (3.7)	257.1 (4.5)
95	220.4 (4.9)	251.7 (6.1)	267.7 (8.0)
HISPANIC STUDENTS			
Mean	162.2 (3.6)	197.2 (3.2)	202.0 (3.2)
Standard Deviation	43.1 (1.4)	38.7 (1.9)	41.1 (3.3)
Percentiles			
5	93.5 (5.4)	130.6 (8.2)	132.7 (8.1)
10	106.9 (3.9)	147.2 (4.4)	148.4 (13.8)
25	131.2 (4.6)	172.9 (4.9)	176.6 (3.3)
50	161.7 (5.9)	197.9 ([illegible])	201.8 (3.2)
75	191.5 (7.4)	223.4 (6.6)	229.4 (4.5)
90	217.9 (5.5)	244.4 (3.9)	253.4 (5.2)
95	234.5 (6.1)	258.8 (5.1)	268.1 (5.2)

Table 16-65

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Main Focused-BIB Civics Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
TOTAL SAMPLE			
Mean	214.0 (0.9)	259.7 (0.9)	296.3 (1.1)
Standard Deviation	8.6 (0.6)	35.7 (0.6)	37.8 (0.6)
Percentiles			
5	163.7 (1.6)	198.1 (2.2)	228.4 (2.4)
10	176.5 (3.8)	212.8 (1.5)	247.0 (1.6)
25	195.9 (1.4)	237.1 (1.3)	273.0 (1.3)
50	215.7 (1.2)	261.2 (1.6)	299.0 (1.0)
75	233.3 (1.3)	284.2 (1.3)	322.1 (1.4)
90	248.9 (2.4)	304.1 (1.3)	342.0 (1.2)
95	258.4 (2.4)	315.3 (1.1)	353.6 (1.2)
MALE STUDENTS			
Mean	214.8 (1.3)	258.7 (1.1)	298.6 (1.6)
Standard Deviation	29.1 (0.9)	38.7 (1.0)	41.1 (1.0)
Percentiles			
5	164.1 (4.1)	190.0 (3.3)	221.8 (4.7)
10	177.5 (2.9)	207.6 (2.0)	244.0 (3.0)
25	196.2 (2.0)	233.9 (1.6)	273.8 (2.4)
50	216.0 (1.3)	260.7 (1.7)	302.5 (1.5)
75	234.0 (1.9)	286.1 (1.3)	326.8 (2.4)
90	250.7 (3.7)	307.0 (2.4)	347.7 (2.7)
95	261.8 (2.3)	318.0 (2.6)	360.1 (3.2)
FEMALE STUDENTS			
Mean	213.3 (1.1)	260.6 (0.9)	294.1 (1.1)
Standard Deviation	28.1 (0.8)	32.6 (0.6)	34.3 (0.8)
Percentiles			
5	163.4 (2.1)	205.2 (3.5)	234.1 (2.4)
10	175.6 (2.4)	218.3 (1.2)	249.1 (1.7)
25	195.6 (1.9)	239.4 (0.8)	272.3 (1.6)
50	215.4 (2.1)	261.6 (0.8)	296.0 (1.5)
75	232.8 (1.4)	282.7 (1.4)	317.9 (0.9)
90	247.7 (1.8)	301.2 (1.8)	336.4 (1.9)
95	255.6 (2.3)	312.6 (3.1)	346.5 (2.3)

Table 16-65 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Main Focused-BIB Civics Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
WHITE STUDENTS			
Mean	220.0 (1.0)	266.3 (1.2)	301.9 (1.2)
Standard Deviation	26.5 (0.8)	34.2 (0.8)	35.7 (0.7)
Percentiles			
5	173.[illegible] ([illegible].9)	206.4 (2.2)	240.8 (1.5)
10	185.[illegible] (1.4)	222.0 (1.8)	255 7 (2.2)
25	203.9 (1.0)	245.1 (2.3)	279.9 (1.4)
50	221.3 (2.1)	268.4 (1.4)	304.3 (1.1)
75	237.8 (1.7)	289.7 (1.1)	326.1 (2.2)
90	252.5 (1.9)	308.6 (1.4)	344.8 (2.2)
95	261.4 (3.0)	319.7 (2.8)	356.1 (1.8)
BLACK STUDENTS			
Mean	198.1 (2.2)	243.6 (1.9)	273.8 (1.9)
Standard Deviation	27.0 (1.4)	33.4 (1.0)	36.3 (1.1)
Percentiles			
5	151.2 (4.8)	188.0 (3.9)	209.4 (3.5)
10	163.2 (2.9)	202.1 (4.0)	225.1 (2.7)
25	180.5 (4.2)	221.3 (3.8)	250.4 (3.2)
50	200.0 (1.9)	244.2 (2.7)	276.2 (2.1)
75	216.8 (3.6)	267.0 (2.3)	298.9 (2.4)
90	231.1 (1.7)	285.6 (1.8)	318.3 (3.1)
95	239.2 (5.4)	297.4 (9.7)	330.0 (3.6)
HISPANIC STUDENTS			
Mean	199.5 (1.9)	240.6 (1.7)	279.2 (2.3)
Standard Deviation	28.8 (1.3)	33.8 (1.4)	37.6 (1.8)
Percentiles			
5	150.5 (5.4)	180.9 (7.3)	209.4 (6.9)
10	164.0 (2.9)	196.6 (4.0)	226.9 (5.9)
25	180.4 (2.2)	219.3 (2.1)	256.4 (4.6)
50	201.0 (1.9)	243.0 (1.4)	282.8 (2.6)
75	217.6 (2.0)	263.1 (1.6)	305.8 (2.9)
90	235.3 (6.1)	282.0 (2.2)	323.5 (3.2)
95	246.4 (6.3)	293.0 (3.4)	336.3 (2.9)

440

Table 16-66

Weighted Proficiency Means, Standard Deviations, and Percentiles with Standard Errors for Civics Bridge Samples, by Age and Subgroup

	Age 13	Age 17
TOTAL SAMPLE		
Mean	50.0 (0.4)	59.6 (0.5)
Standard Deviation	10.0 (0.2)	12.3 (0.4)
Percentiles		
5	34.0 (0.7)	39.1 (1.6)
10	37.5 (0.7)	44.3 (0.6)
25	43.3 (0.4)	52.0 (0.6)
50	49.7 (0.5)	59.8 (0.7)
75	56.4 (0.7)	68.1 (0.8)
90	62.9 (0.9)	74.8 (0.6)
95	67.2 (1.5)	79.1 (0.9)
MALE STUDENTS		
Mean	50.5 (0.6)	61.2 (0.7)
Standard Deviation	10.4 (0.3)	13.0 (0.5)
Percentiles		
5	33.6 (0.9)	39.2 (2.7)
10	37.5 (0.9)	44.9 (1.0)
25	43.5 (0.6)	52.8 (0.8)
50	50.1 (0.7)	61.7 (1.4)
75	57.3 (0.5)	70.4 (0.8)
90	64.2 (0.9)	77.3 (1.3)
95	68.2 (1.0)	81.0 (1.4)
FEMALE STUDENTS		
Mean	49.5 (0.4)	58.2 (0.6)
Standard Deviation	9.6 (0.3)	11.5 (0.5)
Percentiles		
5	34.2 (0.9)	39.1 (2.1)
10	37.5 (1.1)	43.7 (0.7)
25	43.1 (0.4)	51.4 (0.8)
50	49.3 (0.4)	58.7 (0.4)
75	55.6 (0.5)	65.8 (1.0)
90	61.8 (1.0)	71.9 (0.8)
95	65.4 (1.9)	75.9 (1.2)

Table 1 66 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Civics Bridge Samples, by Age and Subgroup

	Age 13	Age 17
WHITE STUDENTS		
Mean	51.2 (0.5)	61.4 (0.6)
Standard Deviation	9.8 (0.3)	11.7 (0.4)
Percentiles		
5	35.5 (1.0)	42.2 (1.2)
10	39.0 (0.6)	47.7 (1.5)
25	44.6 (0.5)	54.1 (0.8)
50	50.8 (0.5)	61.6 (1.0)
75	57.4 (0.6)	69.2 (0.6)
90	64.1 (0.9)	76.1 (0.7)
95	68.2 (0.8)	80.0 (1.0)
BLACK STUDENTS		
Mean	45.7 (0.6)	53.1 (1.0)
Standard Deviation	8.6 (0.4)	12.4 (0.6)
Percentiles		
5	31.7 (1.0)	32.2 (4.6)
10	34.7 (1.4)	38.5 (1.5)
25	39.7 (0.8)	45.2 (0.7)
50	45.3 (0.7)	52.4 (1.0)
75	51.9 (0.8)	61.5 (1.4)
90	57.2 (1.0)	69.6 (1.3)
95	60.4 (1.1)	73.6 (1.3)
HISPANIC STUDENTS		
Mean	45.5 (1.8)	53.8 (1.7)
Standard Deviation	10.5 (0.9)	11.5 (1.0)
Percentiles		
5	28.8 (1.7)	35.0 (5.5)
10	32.0 (2.5)	39.3 (6.4)
25	38.7 (2.0)	46.1 (3.2)
50	45.3 (2.3)	53.8 (3.5)
75	51.7 (2.2)	62.1 (1.3)
90	60.2 (3.5)	68.2 (2.2)
95	63.4 (3.4)	71.4 (5.3)

Table 16-67

Weighted Proficiency Means, Standard Deviations, and Percentiles with Standard Errors for Main U.S. History Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
TOTAL SAMPLE			
Mean	220.6 [illegible]	263.9 (0.7)	295.0 (1.0)
Standard Deviation	30.0 (0.5)	34.1 (0.4)	36.1 (0.6)
Percentiles			
5	169.7 (1.2)	204.3 (2.4)	231.8 (1.0)
10	181.0 (1.2)	219.0 (1.1)	246.7 (2.3)
25	201.3 (1.1)	240.9 (1.1)	270.4 (1.2)
50	221.5 (1.0)	265.0 (0.6)	295.0 (1.1)
75	240.3 (0.9)	288.8 (0.6)	318.9 (0.8)
90	258.7 (1.7)	306.6 (1.2)	341.5 (0.8)
95	268.9 (1.7)	317.2 (0.7)	348.6 (1.4)
MALE STUDENTS			
Mean	222.9 (1.2)	266.2 (1.0)	298.5 (1.3)
Standard Deviation	30.9 (0.7)	36.0 (0.7)	38.6 (0.8)
Percentiles			
5	170.3 (1.5)	202.5 (2.7)	229.2 (3.1)
10	182.0 (2.0)	218.2 (2.1)	244.7 (2.1)
25	202.6 (1.4)	242.1 (1.3)	273.3 (2.6)
50	225.1 (1.8)	267.3 (0.9)	302.0 (4.3)
75	242.7 (1.4)	291.4 (0.8)	324.4 (2.5)
90	262.4 (1.6)	311.6 (1.6)	344.6 (0.8)
95	272.1 (1.8)	320.1 (1.1)	353.9 (2.3)
FEMALE STUDENTS			
Mean	218.2 (1.0)	261.6 (0.8)	291.8 (1.1)
Standard Deviation	28.8 (0.6)	31.9 (0.5)	33.4 (0.6)
Percentiles			
5	169.0 (1.8)	206.2 (3.1)	234.1 (1.8)
10	180.1 (1.2)	219.8 (1.9)	248.3 (2.0)
25	199.7 (1.7)	240.1 (1.0)	268.7 (0.9)
50	218.6 (1.2)	263.3 (0.5)	291.8 (0.8)
75	238.2 (1.3)	285.7 (1.2)	316.2 (0.6)
90	253.7 (1.9)	299.8 (1.0)	335.8 (2.2)
95	264.6 (2.6)	313.4 (1.5)	344.6 (0.7)

Table 16-67 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Main U.S. History Samples, by Grade and Subgroup

	Grade 4	Grade 8	Grade 12
WHITE STUDENTS			
Mean	227.5 (1.0)	270.4 (0.8)	301.1 (1.2)
Standard Deviation	27.0 (0.6)	32.2 (0.6)	34.1 (0.6)
Percentiles			
5	181.8 (2.3)	212.9 (1.5)	240.5 (4.7)
10	192.5 (1.9)	229.5 (1.4)	255.8 (1.7)
25	209.5 (1.1)	251.4 (2.0)	281.5 (2.0)
50	229.2 (1.0)	270.8 (1.2)	304.0 (1.6)
75	244.4 (1.0)	291.8 (0.6)	322.3 (1.4)
90	262.6 (1.7)	311.2 (1.0)	343.5 (1.1)
95	271.8 (1.8)	319.3 (0.9)	350.9 (1.4)
BLACK STUDENTS			
Mean	199.5 (1.9)	246.0 (1.5)	274.4 (1.7)
Standard Deviation	27.7 (1.1)	30.4 (1.0)	33.7 (0.9)
Percentiles			
5	154.5 (5.1)	195.2 (3.9)	218.0 (4.5)
10	163.4 (4.7)	207.5 (3.0)	230.9 (2.7)
25	180.5 (1.3)	227.5 (2.1)	252.1 (2.2)
50	199.5 (3.0)	245.1 (1.6)	273.6 (3.4)
75	217.5 (1.2)	265.3 (1.9)	296.3 (2.5)
90	236.9 (3.2)	286.7 (1.6)	317.6 (0.9)
95	244.9 (2.8)	295.0 (2.4)	327.8 (4.6)
HISPANIC STUDENTS			
Mean	202.7 (1.7)	244.3 (1.9)	273.9 (1.8)
Standard Deviation	28.7 (0.9)	33.7 (0.9)	34.6 (1.2)
Percentiles			
5	153.4 (5.9)	187.7 (5.9)	217.1 (4.4)
10	165.7 (2.8)	202.1 (3.2)	229.7 (2.9)
25	182.7 (1.3)	221.8 (2.2)	252.3 (3.0)
50	204.1 (2.5)	244.4 (2.5)	272.7 (3.2)
75	222.5 (1.6)	266.7 (1.3)	295.8 (2.9)
90	239.5 (0.9)	288.5 (2.0)	318.6 (1.0)
95	247.3 (4.0)	296.7 (1.9)	329.1 (7.9)

Table 16-68

Weighted Proficiency Means, Standard Deviations, and Percentiles with Standard Errors for Main Geography Sample, by Grade and Subgroup

	Grade 12
TOTAL SAMPLE	
Mean	293.1 (1.0)
Standard Deviation	37.0 (0.8)
Percentiles	
5	227.5 (2.7)
10	242.2 (1.6)
25	268.7 (2.0)
50	295.5 (1.5)
75	319.1 (0.9)
90	338.6 (1.3)
95	349.1 (3.3)
MALE STUDENTS	
Mean	301.2 (1.6)
Standard Deviation	38.3 (1.0)
Percentiles	
5	231.6 (2.2)
10	247.1 (2.8)
25	275.6 (2.8)
50	305.6 (2.0)
75	328.4 (1.6)
90	346.6 (2.3)
95	358.3 (4.8)
FEMALE STUDENTS	
Mean	285.7 (1.2)
Standard Deviation	34.2 (0.9)
Percentiles	
5	224.3 (2.5)
10	238.9 (2.2)
25	263.7 (1.3)
50	289.2 (1.2)
75	310.0 (1.2)
90	327.6 (2.2)
95	337.9 (3.6)

Table 16-68 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles with Standard Errors for Main Geography Samples, by Grade and Subgroup

Grade 12

WHITE STUDENTS	
Mean	301.1 (1.1)
Standard Deviation	33.7 (0.9)
Percentiles	
5	242.3 (2.8)
10	256.4 (2.9)
25	279.7 (1.4)
50	302.7 (1.4)
75	323.7 (1.4)
90	341.9 (2.1)
95	352.9 (2.1)
BLACK STUDENTS	
Mean	258.4 (2.0)
Standard Deviation	32.0 (1.2)
Percentiles	
5	207.3 (3.7)
10	216.8 (3.5)
25	234.7 (2.9)
50	259.4 (3.5)
75	282.2 (3.8)
90	298.8 (2.5)
95	310.5 (7.8)
HISPANIC STUDENTS	
Mean	271.8 (3.9)
Standard Deviation	35.0 (1.6)
Percentiles	
5	214.4 (8.6)
10	225.0 (7.4)
25	245.8 (3.4)
50	273.1 (5.6)
75	296.9 (2.8)
90	317.8 (8.7)
95	331.1 (3.0)

Table 16-69

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Bridge to 1986 Mathematics Samples, by Age and Subgroup

	Age 9	Age 13	Age 17
TOTAL SAMPLE			
Mean	229.0 (1.1)	273.3 (0.8)	305.4 (1.2)
Standard Deviation	33.1 (0.7)	31.7 (0.4)	29.7 (0.7)
Percentiles			
5	171.8 (1.4)	221.2 (1.6)	255.5 (1.8)
10	185.3 (2.0)	233.0 (1.4)	266.1 ([illegible].3)
25	207.0 (1.6)	252.1 (0.9)	284.5 (1.6)
50	230.6 (1.0)	273.4 (1.0)	306.0 (1.4)
75	252.0 (1.6)	294.6 (1.4)	326.1 (1.9)
90	270.1 (1.1)	314.5 (2.5)	344.4 (1.1)
95	280.5 (1.4)	325.4 (1.8)	353.7 (3.4)
MALE STUDENTS			
Mean	229.1 (1.6)	275.3 (1.1)	306.7 (1.8)
Standard Deviation	33.8 (1.0)	31.8 (0.6)	30.2 (1.1)
Percentiles			
5	171.7 (2.3)	223.6 (2.3)	256.7 (3.7)
10	184.3 (2.4)	235.4 (1.3)	266.8 (0.8)
25	206.3 (3.2)	253.8 (1.0)	284.9 (2.1)
50	230.5 (1.8)	275.3 (1.1)	306.9 (1.9)
75	252.3 (2.2)	296.6 (1.4)	327.6 (3.3)
90	271.4 (2.1)	316.9 (1.9)	346.7 (3.8)
95	282.8 (2.3)	328.1 (2.3)	355.9 (4.2)
FEMALE STUDENTS			
Mean	229.0 (1.1)	271.2 (1.0)	304.2 (1.4)
Standard Deviation	32.4 (0.8)	31.5 (0.8)	29.2 (0.6)
Percentiles			
5	172.2 (3.4)	219.3 (2.5)	253.9 (3.0)
10	186.2 (2.8)	230.5 (2.2)	264.9 (3.7)
25	207.4 (1.4)	250.0 (1.2)	284.0 (2.8)
50	230.7 (1.1)	271.6 (1.4)	305.2 (1.9)
75	251.8 (1.8)	292.3 (1.2)	324.5 (2.2)
90	269.1 (1.5)	311.5 (1.4)	341.9 (1.8)
95	278.4 (2.4)	322.8 (1.8)	350.8 (3.1)

Table 16-69 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Bridge to 1986 Mathematics Samples, by Age and Subgroup

	Age 9	Age 13	Age 17
WHITE STUDENTS			
Mean	234.5 (1.2)	279.1 (0.9)	309.5 (1.4)
Standard Deviation	30.4 (0.7)	29.4 (0.4)	28.9 (0.8)
Percentiles			
5	182.4 (3.3)	231.6 (1.5)	259.3 (1.7)
10	194.4 (1.8)	241.9 (1.4)	271.0 (1.2)
25	214.4 (2.0)	259.1 (1.3)	290.2 (2.7)
50	235.9 (1.3)	279.0 (1.4)	310.1 (1.3)
75	255.6 (1.7)	299.1 (1.5)	329.5 (2.7)
90	272.6 (1.4)	317.3 (1.3)	346.6 (2.0)
95	282.4 (1.3)	327.8 (1.4)	356.1 (2.9)
BLACK STUDENTS			
Mean	206.3 (2.6)	250.3 (1.2)	289.2 (2.1)
Standard Deviation	32.7 (1.3)	27.9 (1.0)	27.1 (1.3)
Percentiles			
5	152.6 (2.2)	203.7 (1.7)	243.9 (5.9)
10	164.3 (4.0)	214.5 (1.8)	254.5 (3.4)
25	183.8 (4.8)	231.9 (2.5)	270.3 (1.6)
50	206.6 (3.6)	250.9 (2.3)	288.6 (1.3)
75	229.2 (2.1)	269.5 (1.4)	307.4 (3.4)
90	247.7 (2.2)	284.8 (2.0)	325.1 (3.6)
95	258.2 (2.0)	294.1 (2.1)	334.1 (6.4)
HISPANIC STUDENTS			
Mean	215.9 (3.4)	254.7 (3.9)	294.3 (3.5)
Standard Deviation	33.1 (1.7)	30.3 (2.2)	27.6 (2.6)
Percentiles			
5	162.1 (6.8)	201.1 (7.6)	249.1 (9.7)
10	173.7 (7.0)	214.9 (4.3)	259.3 (11.4)
25	193.3 (4.0)	234.4 (5.4)	274.9 (6.9)
50	216.3 (5.4)	257.2 (4.6)	292.8 (4.8)
75	239.1 (5.4)	274.2 (3.0)	313.9 (2.9)
90	259.8 (4.7)	289.1 (5.0)	330.1 (2.9)
95	269.2 (9.7)	299.4 (8.4)	341.9 (9.3)

Table 16-70

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Bridge to 1986 Science Samples, by Age and Subgroup

	Age 9	Age 13	Age 17
TOTAL SAMPLE			
Mean	228.9 (1.3)	257.3 (0.9)	294.2 (1.5)
Standard Deviation	41.0 (0.7)	37.2 (0.3)	41.4 (1.2)
Percentiles			
5	160.5 (1.9)	194.4 (2.4)	225.1 (4.1)
10	175.8 (1.5)	208.8 (1.1)	240.9 (1.7)
25	201.3 (1.3)	232.1 (1.3)	266.5 (1.7)
50	230.0 (1.6)	258.1 (1.0)	294.5 (3.0)
75	257.0 (0.9)	282.9 (1.3)	322.2 (2.1)
90	280.7 (2.2)	304.3 (1.4)	347.7 (3.3)
95	293.9 (2.0)	317.7 (2.4)	362.7 (3.9)
MALE STUDENTS			
Mean	232.1 (1.6)	262.2 (1.2)	302.5 (2.3)
Standard Deviation	41.2 (1.0)	37.5 (0.8)	41.6 (1.7)
Percentiles			
5	163.5 (2.8)	198.6 (2.2)	232.2 (4.0)
10	178.7 (2.0)	213.1 (1.9)	249.7 (3.3)
25	204.6 (1.8)	237.0 (1.6)	275.4 (2.8)
50	233.3 (1.9)	262.9 (1.2)	302.7 (3.7)
75	259.8 (2.5)	287.7 (1.7)	331.1 (3.3)
90	284.0 (2.6)	309.8 (2.3)	357.3 (4.3)
95	298.3 (4.5)	322.9 (3.0)	370.5 (5.4)
FEMALE STUDENTS			
Mean	225.7 (1.6)	252.4 (1.0)	285.6 (1.9)
Standard Deviation	40.5 (0.8)	36.3 (0.9)	39.4 (1.5)
Percentiles			
5	157.5 (1.9)	190.8 (2.9)	220.0 (6.1)
10	173.2 (3.1)	205.2 (2.6)	235.1 (3.5)
25	198.7 (1.5)	228.3 (2.1)	258.9 (3.3)
50	227.0 (2.7)	253.5 (0.8)	287.0 (5.0)
75	253.9 (1.9)	277.4 (1.3)	312.7 (2.6)
90	276.3 (2.1)	297.7 (2.1)	335.3 (3.3)
95	289.4 (2.6)	310.2 (2.8)	349.1 (2.5)

Table 16-70 (continued)

Weighted Proficiency Means, Standard Deviations, and Percentiles
with Standard Errors
for Bridge to 1986 Science Samples, by Age and Subgroup

	Age 9	Age 13	Age 17
WHITE STUDENTS			
Mean	237.4 (1.3)	265.2 (0.9)	301.9 (1.7)
Standard Deviation	37.4 (0.7)	34.[illegible] (0.7)	38.0 (1.5)
Percentiles			
5	176.0 (1.6)	209.1 (1.1)	239.1 (4.5)
10	189.1 (1.1)	221.9 (1.8)	254.3 (5.0)
25	211.8 (1.4)	242.3 (1.1)	276.4 (3.0)
50	238.3 (1.6)	265.3 (0.9)	301.9 (2.0)
75	262.7 (1.3)	287.9 (1.5)	327.5 (2.4)
90	285.2 (1.8)	308.6 (2.0)	351.3 (3.1)
95	298.3 (4.2)	320.9 (1.9)	364.5 (2.6)
BLACK STUDENTS			
Mean	200.1 (2.5)	229.4 (1.2)	260.0 (3.4)
Standard Deviation	38.7 (1.5)	33.5 (1.2)	38.6 (2.7)
Percentiles			
5	135.7 (3.0)	175.0 (3.3)	200.2 (11.1)
10	149.3 (5.4)	186.7 (4.1)	212.3 (8.2)
25	173.5 (3.8)	206.8 (1.9)	236.0 (3.2)
50	200.2 (2.6)	228.6 (3.0)	259.0 (2.8)
75	226.3 (2.9)	252.7 (2.6)	283.9 (5.4)
90	251.3 (4.5)	272.7 (2.5)	309.2 (4.4)
95	263.8 (4.1)	283.6 (4.2)	323.6 (9.6)
HISPANIC STUDENTS			
Mean	201.0 (6.1)	229.3 (4.2)	281.8 (5.2)
Standard Deviation	42.0 (2.8)	35.5 (2.1)	37.3 (2.8)
Percentiles			
5	132.3 (6.4)	172.1 (8.9)	225.8 (6.8)
10	146.8 (9.1)	181.9 (5.9)	236.6 (4.9)
25	172.0 (8.6)	202.0 (6.8)	254.7 (6.7)
50	201.0 (8.6)	229.8 (4.0)	282.6 (11.6)
75	228.7 (6.5)	254.7 (5.5)	303.2 (9.3)
90	255.6 (5.4)	276.9 (4.6)	334.1 (13.8)
95	270.2 (4.8)	289.2 (8.4)	350.6 (7.5)

Table 16-71

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 4, by Gender, for Main Focused-BIB Reading Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	4531	860,553(1%)	51.7(1.1) 226.6(1.5)	48.3(1.1) 234.4(1.2)	0.0
GENDER					
MALE	2345	445,048(3%)	100.0(0.0) 226.6(1.5)	0.0(0.0)	0.1
FEMALE	2186	415,506(3%)	0.0(0.0)	100.0(0.0) 234.4(1.2)	0.0
RACE/ETHNICITY					
WHITE	2557	599,759(2%)	51.5(1.6) 234.8(1.8)	48.[illegible](1.6) 241.6(1.4)	0.0
BLACK	854	137,164(3%)	50.7(1.9) 206.6(2.8)	49.3(1.9) 214.6(2.4)	0.1
HISPANIC	869	88,350(3%)	53.5(2.0) 202.7(3.4)	46.5(2.0) 218.2(2.1)	0.0
OTHER	251	35,281(7%)	54.6(3.4) 226.4(5.4)	45.4(3.4) 229.0(4.4)	0.3
PARENTAL EDUCATION					
LESS THAN H.S.	228	38,671(10%)	57.2(3.2) 196.9(4.8)	42.8(3.2) 214.4(3.8)	0.0
GRADUATED H.S.	67[illegible]	132,632(5%)	50.3(2.7) 222.9(2.7)	49.7(2.7) 228.9(2.6)	0.0
SOME EDUC AFTER H.S.	358	67,315(6%)	49.4(3.6) 232.3(3.0)	50.6(3.6) 249.[illegible](3.1)	0.0
GRADUATED COLLEGE	1623	327,691(4%)	53.4(1.4) 238.5(2.1)	46.[illegible](1.4) 243.7(1.6)	0.0
UNKNOWN	1628	292,006(3%)	50.1(1.5) 217.7(1.7)	49.9(1.5) 226.2(1.6)	0.1

* CV is the coefficient of variation for the sum of the weights.

Table 16-72

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 4, by Derived Race/ethnicity, for Main Focused-BIB Reading Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	4534	860,914(1%)	69.7(0.6)	15.9(0.5)	10.3(0.3)	2.3(0.2)	1.7(0.2)	0.0(0.0)	0.0
			238.1(1.2)	210.6(1.9)	209.9(2.4)	234.5(5.9)	218.6(5.1)	213.0(17.0)	
GENDER									
MALE	2347	445,278(3%)	69.4(1.1)	15.6(0.7)	10.6(0.6)	2.5(0.3)	1.8(0.3)	0.1(0.0)	0.0
			234.8(1.8)	206.6(2.8)	202.7(3.4)	232.0(7.9)	219.6(6.7)	193.0(37.5)	
FEMALE	2187	415,636(3%)	70.0(1.2)	16.3(0.9)	9.9(0.5)	2.1(0.3)	1.7(0.3)	0.0(0.0)	0.0
			241.6(1.4)	214.6(2.4)	218.2(2.1)	237.7(6.3)	217.6(6.9)	251.1(9.5)	
RACE/ETHNICITY									
WHITE	2558	599,890(2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			238.1(1.2)						
BLACK	855	137,296(3%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				210.6(1.9)					
HISPANIC	869	88,350(3%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					209.9(2.4)				
OTHER	252	35,378(7%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	56.8(4.1)	42.1(4.2)	1.1(0.7)	0.0
						234.5(5.9)	218.6(5.1)	213.0(17.0)	
PARENTAL EDUCATION									
LESS THAN H.S.	228	38,671(10%)	65.8(3.9)	15.1(2.5)	16.7(2.9)	0.4(0.4)	2.0(0.7)	0.0(0.0)	0.0
			210.0(4.3)	187.2(7.4)	197.6(6.2)	204.5(****)	204.5(7.6)		
GRADUATED H.S.	675	132,632(5%)	73.7(1.7)	14.6(1.5)	9.7(0.9)	1.0(0.3)	0.9(0.3)	0.0(0.0)	0.0
			231.0(2.6)	211.3(3.5)	210.6(5.1)	239.6(18.2)	194.7(16.2)		
SOME EDUC AFTER H.S.	358	67,315(6%)	74.0(2.1)	13.7(1.9)	9.0(1.2)	1.9(0.7)	1.4(0.6)	0.0(0.0)	0.0
			248.0(2.3)	219.9(6.0)	223.5(5.7)	216.6(20.4)	209.5(26.6)		
GRADUATED COLLEGE	1624	327,824(4%)	72.0(1.3)	16.0(1.0)	7.5(0.6)	2.4(0.4)	2.0(0.3)	0.1(0.1)	0.0
			248.3(1.7)	217.9(3.1)	220.0(3.8)	247.8(4.9)	230.8(6.6)	213.0(17.0)	
UNKNOWN	1630	292,234(3%)	65.3(1.4)	16.6(1.1)	13.0(0.7)	3.2(0.5)	1.9(0.4)	0.0(0.0)	0.0
			230.3(1.5)	203.9(3.0)	203.4(2.9)	225.6(8.7)	213.2(7.7)		

* CV is the coefficient of variation for the sum of the weights.
(****) Standard error is greater than 99.9.

Table 16-73

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 4, by Parental Education, for Main Focused-BIB Reading Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	4515	858,677(1%)	4.5(0.4)	15 4(0.8)	7.8(0.5)	38.2(1.5)	34.0(1.1)	0.3
			204.4(3.2)	225.9(2.1)	240.8(2.1)	240.9(1.3)	221.9(1.4)	
GENDER								
MALE	2333	443,602(3%)	5.0(0.6)	15.0(1.0)	7.5(0.7)	39.5(1.6)	33.0(1.3)	0 4
			196.9(4.8)	222.9(2.7)	232.3(3.0)	238.4(2.1)	217.7(1.7)	
FEMALE	2182	415,075(3%)	4.0(0.5)	15.9(1.1)	8.2(0.7)	36.8(1.8)	35.1(1.4)	
			214.4(3.8)	228.9(2.6)	249.2(3.1)	243.7(1.6)	226.2(1.6)	
RACE/ETHNICITY								
WHITE	2557	599,750(2%)	4.2(0.6)	16.3(1.0)	8.3(0.6)	39.3(1.9)	31.8(1.4)	0.0
			210.0(4.3)	231.0(2.6)	248.0(2.3)	248.3(1.7)	230.3(1.5)	
BLACK	842	135,579(3%)	4.3(0.8)	14.3(1.6)	6.8(1.0)	38.8(1.9)	35.8(1.9)	1.3
			187.2(7.4)	211.3(3.5)	219.9(6.0)	217.9(3.1)	203.9(3.0)	
HISPANIC	864	87,970(3%)	7.3(1.2)	14.7(1.2)	6.9(1.0)	27.8(2.2)	43.3(2.1)	0.4
			197.6(6.2)	210.6(5.1)	223.5(5.7)	220.0(3.8)	203.4(2.9)	
OTHER	252	35,378(7%)	2.6(0.8)	7.2(1.6)	6.3(1.6)	41.7(4.4)	42.2(4.0)	0.0
			204.5(9.3)	218.1(12.0)	213.7(14.9)	239.5(3.9)	221.0(5.9)	
PARENTAL EDUCATION								
LESS THAN H.S.	228	38,671(10%)	100.0(0 0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			204.4(3.2)					
GRADUATED H.S.	675	132,632(5%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				225 9(2.1)				
SOME EDUC AFTER H.S.	358	67,315(6%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					240.8(2.1)			
GRADUATED COLLEGE	1624	327,824(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
						240.9(1.3)		
UNKNOWN	1630	292,234(3%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0
							221.9(1.4)	

* CV is the coefficient of variation for the sum of the weights.

Table 16-74

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 8, by Gender, for Main Focused-BIB Reading Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	4404	491,575(1%)	51.5(1.1)	48.5(1.1)	0.0
			256.4(1.3)	269.6(1.1)	
GENDER					
MALE	2242	253,319(3%)	100.0(0.0)	0.0(0.0)	0.0
			256.4(1.3)		
FEMALE	2162	238,257(2%)	0.0(0.0)	100.0(0.0)	0.0
				269.6(1.1)	
RACE/ETHNICITY					
WHITE	2694	343,252(2%)	51.5(1.2)	48.5(1.2)	0.0
			262.8(1.8)	276.1(1.4)	
BLACK	772	74,650(5%)	49.3(3.3)	50.7(3.3)	0.0
			238.9(2.5)	252.4(2.2)	
HISPANIC	711	52,413(4%)	54.8(2.3)	45.2(2.3)	0.0
			237.7(2.3)	250.7(2.7)	
OTHER	227	21,260(8%)	51.8(3.2)	48.2(3.2)	0.0
			260.7(4.4)	271.9(3.3)	
PARENTAL EDUCATION					
LESS THAN H.S.	375	36,427(8%)	40.4(2.9)	59.6(2.9)	0.0
			236.8(3.9)	250.2(2.2)	
GRADUATED H.S.	1193	136,783(3%)	51.7(1.6)	48.3(1.6)	0.0
			249.9(2.0)	263.2(1.7)	
SOME EDUC AFTER H.S.	731	89,177(4%)	47.3(1.6)	52.7(1.6)	0.0
			261.2(2.1)	275.3(1.8)	
GRADUATED COLLEGE	1 39	189,833(4%)	53.8(1.6)	46.2(1.6)	0.0
			266.4(1.8)	280.2(1.5)	
UNKNOWN	403	38,033(7%)	59.6(2.8)	40.4(2.8)	0.0
			237.6(2.7)	248.6(2.3)	

* CV is the coefficient of variation for the sum of the weights.

Table 16-75

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 8, by Derived Race/ethnicity, for Main Focused-BIB Reading Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	4404	491,575(1%)	69.8(0.7)	15.2(0.7)	10.7(0.4)	3.0(0.3)	1.3(0.2)	0.1(0.0)	0.0
			269.2(1.3)	245.7(2.0)	243.6(2.1)	274.9(2.7)	248.1(5.5)	206.1(26.7)	
GENDER									
MALE	2242	253,319(3%)	69.8(1.0)	14.5(0.8)	11.3(0.6)	3.0(0.4)	1.3(0.3)	0.1(0.1)	0.0
			262.8(1.8)	238.9(2.5)	237.7(2.3)	273.0(3.1)	236.7(7.8)	206.1(26.7)	
FEMALE	2162	238,257(2%)	69.9(1.3)	15.9(1.3)	10.0(0.7)	3.0(0.4)	1.3(0.3)	0.0(0.0)	0.0
			276.1(1.4)	252.4(2.2)	250.7(2.7)	277.0(4.5)	260.2(5.3)		
RACE/ETHNICITY									
WHITE	2694	343,252(2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			269.2(1.3)						
BLACK	772	74,650(5%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				245.7(2.0)					
HISPANIC	711	52,413(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					243.6(2.1)				
OTHER	227	21,260(8%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	68.9(4.9)	28.9(4.9)	1.2(0.7)	0.0
						274.9(2.7)	248.1(5.5)	206.1(26.7)	
PARENTAL EDUCATION									
LESS THAN H.S.	375	36,427(8%)	51.9(3.5)	16.8(2.9)	27.2(2.1)	2.5(1.1)	1.4(0.7)	0.2(0.2)	0.0
			246.9(3.2)	241.8(5.2)	241.2(3.1)	262.6(12.7)	240.4(14.1)	239.3(****)	
GRADUATED H.S.	1193	136,783(3%)	71.9(1.7)	15.7(1.2)	9.8(0.8)	1.5(0.3)	1.0(0.3)	0.0(0.0)	0.0
			260.7(1.7)	242.4(3.0)	244.6(2.9)	271.3(7.9)	251.1(9.6)		
SOME EDUC AFTER H.S.	781	89,177(4%)	73.9(1.9)	14.9(1.7)	8.7(0.9)	1.3(0.3)	1.2(0.3)	0.0(0.0)	0.0
			272.4(1.8)	257.5(2.7)	251.4(3.9)	287.0(8.0)	276.9(10.0)		
GRADUATED COLLEGE	1639	189,833(4%)	75.8(1.4)	13.0(1.1)	5.8(0.6)	4.2(0.8)	1.1(0.3)	0.0(0.0)	0.0
			278.6(1.3)	248.5(2.7)	248.2(4.5)	282.6(3.3)	251.3(7.2)		
UNKNOWN	403	38,033(7%)	41.7(2.6)	21.6(2.2)	26.8(1.9)	6.8(1.7)	2.9(1.0)	0.3(0.3)	0.0
			251.3(4.0)	234.2(3.3)	233.5(3.3)	253.2(5.3)	221.4(11.7)	216.1(****)	

* CV is the coefficient of variation for the sum of the weights.
(****) Standard error is greater than 99.9.

Tabl 16-76

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 8, by Parental Education, for Mai[illegible] Focused-BIB Reading Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	4391	490,253(1%)	7.4(0.5) 244.8(2.2)	27.9(0.9) 256.3(1.5)	18.2(0.8) 268.6(1.4)	38.7(1.3) 272.7(1.2)	7.8(0.5) 242.0(2.0)	0 3
GENDER								
MALE	2234	252,546(3%)	5.8(0.6) 236.8(3.9)	28.0(1.0) 24[illegible].9(2.0)	16.7(0.8) 261.2(2.1)	40.5(1.3) 266 [illegible] 1.8)	9.0(0.7) 237.6(2.7)	0.3
FEMALE	[illegible]157	237,707(2%)	9.1(0.9) 250.2(2.2)	27.8(1.3) 263.2(1.7)	[illegible].8(1.1) 275.3(1.8)	36 [illegible] 7) 280 [illegible])	6.5(0.5) 248.6(2.3)	0.2
RACE/ETHNICITY								
WHITE	2691	342,965(2%)	5.[illegible](0.6) 246.9(3.2)	28.7(1.3) 260.7(1.7)	19.2(0.9) 272.4(1.8)	42.0(1.6) 278.6(1.3)	4.6(0.4) 251.3(4.0)	0.1
BLACK	76[illegible]	73,869(5%)	8.3(1.4) 241.8(5.2)	29.1(2.0) 242.4(3.0)	18.0(2.1) 257.5(2.7)	33.5(2.1) 248.5(2.7)	11.1(1.6) 234.2(3.3)	1.0
HISPANIC	709	52,347(4%)	18.9(2.2) 241.2(3.1)	[illegible] 6(1.5) [illegible]44.6(2.9)	14.8(1.3) 251.4(3.9)	21.2(1.9) 248.2(4.5)	19.5(1.4) 233.5(3.3)	0.1
OTHER	225	21,073(8%)	7.0(2.2) 253.9(9.2)	16.6([illegible].2) 26[illegible].2(6.7)	10.4(2.1) 282.2(6.6)	47.9(6.6) 275.9(2.6)	18.0(3.1) 243.0(6.7)	0.9
PARENTAL EDUCATION								
LESS THAN H.S.	375	36,42[illegible](8%)	100.0(0.0) 244.8(2.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED H.S.	1193	136,783(3%)	0.0(0.0)	100.0(0.[illegible]) 256.3(1.5)	0.0(0.0)	0.0(0.0)	0.0(0.0)	G 0
SOME EDUC AFTER H.S.	781	89,177(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 268.6(1.4)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED COLLEGE	1639	189,833(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 272.7(1.2)	0.0(0.0)	0.0
UNKNOWN	403	38,033(7%)	0.0(0 0)	0.0(0.0)	[illegible] 0(0.0)	0.0(0 0)	100.0(0 0) 242.0(2 0)	0.0

* CV is the coefficient of variation for the sum of the weights.

Table 16-77

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 12, by Gender, for Main Focused-BIB Reading Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	4250	521,501(2%)	47.5(1.1)	52.5(1.1)	0.0
			282.9(1.1)	291.0(1.1)	
GENDER					
MALE	2002	247,866(3%)	100.0(0.0)	0.0(0.0)	0.0
			282.9(1.1)		
FEMALE	2248	273,635(3%)	0.0(0.0)	100.0(0.0)	0.0
				291.0(1.1)	
RACE/ETHNICITY					
WHITE	2887	385,848(2%)	47.1(1.4)	52.9(1.4)	0.0
			288.5(1.1)	296.2(1.2)	
BLACK	748	72,396(6%)	45.4(2.4)	54.6(2.4)	0.0
			264.4(2.5)	275.2(2.6)	
HISPANIC	473	41,853(8%)	46.7(2.3)	53.3(2.3)	0.0
			257.6(3.1)	275.3(3.2)	
OTHER	142	21,404(10%)	63.5(4.9)	36.5(4.9)	0.0
			287.8(4.5)	279.9(5.4)	
PARENTAL EDUCATION					
LESS THAN H.S.	385	39,295(8%)	40.0(3.0)	60.0(3.0)	0.0
			263.9(4.0)	276.4(2.1)	
GRADUATED H.S.	1060	128,097(5%)	45.7(2.0)	54.3(2.0)	0.0
			273.0(1.8)	282.5(1.7)	
SOME EDUC AFTER H.S.	1057	127,961(4%)	45.0(1.8)	55.0(1.8)	0.0
			285.2(1.7)	291.3(1.5)	
GRADUATED COLLEGE	1643	214,061(3%)	51.2(1.6)	48.8(1.6)	0.0
			291.0(1.5)	301.0(1.7)	
UNKNOWN	97	10,952(10%)	51.3(6.8)	48.7(6.8)	0.0
			253.8(5.1)	265.4(9.0)	

* CV is the coefficient of variation for the sum of the weights.

Table 16-78

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 12, by Derived Race/ethnicity, for Main Focused-BIB Reading Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	4250	521,501(2%)	74.0(1.0)	13.9(0.7)	8.0(0.5)	3.2(0.3)	0.8(0.3)	0.1(0.1)	0.0
			292.6(1.0)	270.3(1.6)	267.1(2.4)	290.3(3.7)	267.9(7.4)	246.0(28.5)	
GENDER									
MALE	2002	247,866(3%)	73.4(1.2)	13.3(0.9)	7.9(0.6)	4.5(0.5)	0.9(0.3)	0.1(0.1)	0.0
			288.5(1.1)	264.4(2.5)	257.6(3.1)	292.4(5.3)	271.2(10.0)	235.3(****)	
FEMALE	2248	273,635(3%)	74.6(1.5)	14.4(1.0)	8.2(0.8)	2.0(0.4)	0.8(0.4)	0.0(0.0)	0.0
			296.2(1.2)	275.2(2.6)	275.3(3.2)	286.2(5.3)	264.5(8.9)	268.2(6.8)	
RACE/ETHNICITY									
WHITE	2887	385,848(2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			292.6(1.0)						
BLACK	748	72,396(6%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				270.3(1.6)					
HISPANIC	473	41,853(8%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					267.1(2.4)				
OTHER	142	21,404(10%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	77.6(5.8)	20.6(5.8)	1.8(1.3)	0.0
						290.3(3.7)	267.9(7.4)	246.0(28.5)	
PARENTAL EDUCATION									
LESS THAN H.S.	385	39,295(8%)	48.6(4.2)	18.7(2.9)	28.8(3.2)	2.0(0.9)	1.9(1.7)	0.0(0.0)	0.0
			276.8(3.1)	265.7(3.2)	266.1(3.9)	261.1(4.2)	279.8(29.5)		
GRADUATED H.S.	1060	128,097(5%)	73.1(1.8)	17.0(1.6)	7.3(0.7)	2.0(0.5)	0.6(0.3)	0.0(0.0)	0.0
			283.0(1.4)	265.3(2.9)	260.7(3.7)	275.3(14.0)	270.5(5.6)		
SOME EDUC AFTER H.S.	1057	127,961(4%)	74.3(1.7)	15.9(1.3)	6.3(0.7)	2.7(0.6)	0.7(0.3)	0.1(0.1)	0.0
			293.1(1.4)	275.[illegible](2.8)	273.2(4.5)	279.2(6.4)	288.2(9.9)	268.2(6.8)	
GRADUATED COLLEGE	1543	214,061(3%)	81.0(1.1)	9.[illegible](0.7)	4.8(0.5)	4.0(0.6)	0.6(0.3)	0.1(0.1)	0.0
			299.9(1.3)	274.1(2.8)	271.[illegible](3.8)	302.9(5.0)	257.0(12.5)	235.3(****)	
UNKNOWN	97	10,952(10%)	39.2(7.0)	22.6(4.0)	24.0(5.3)	6.2(3.2)	5.9(3.0)	0.0(0.0)	0.0
			262.3(12.7)	255.4(7.6)	254.4(4.6)	282.0(7.9)	245.2(20.0)		

* CV is the coefficient of variation for the sum of the weights.
(****) Standard error is greater than 99.9.

460

461

Table 16-79

Weighted Response Percentages and Reading Proficiency Means with Standard Errors
Grade 12, by Parental Education, for Main Focused-BIB Reading Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	4242	520,364(2%)	7.6(0.6) 271.4(2.1)	24.6(1.0) 278.2(1.1)	24.6(0.9) 288.6(1.2)	41.1(1.5) 295.9(1.2)	2.1(0.2) 259.5(6.0)	0.2
GENDER								
MALE	1997	247,126(3%)	6.4(0.7) 263.9(4 0)	23.7(1.3) 273.0(1.8)	2[illegible] 3(1.1) 28[illegible].2(1.7)	44.4(1.9) 291.0(1.5)	2.3(0.4) 253.8(5.1)	0.3
FEMALE	2245	273,238(3%)	8.6(0.8) 275.4(2.1)	25.4(1.2) 282.5(1.7)	25.[illegible](1.1) 291.3(1.5)	38.2(1.6) 301.0(1.7)	2.0(0.3) 265.4(9.0)	0.1
RACE/ETHNICITY								
WHITE	2884	385,571(2%)	5.0(0.6) 276.8(3.1)	2[illegible].3(1.2) 283.0(1.4)	24.7(1.0) 293.1(1.4)	45.0(1.7) 299.9(1.3)	1.1(0.2) 262.3(12.7)	0.1
BLACK	746	72,150(6%)	10.2(1.5) 265.7(3.2)	30.1(2.0) 265.3(2.9)	28.2(1.7) 275.3(2.8)	28.1(2.6) 274.1(2.8)	3.4(0.7) 255.4(7.6)	0.3
HISPANIC	471	41,512(8%)	27.3(2.2) 266.1(3.9)	22.4(1.8) 260.7(3.7)	19.3(1.5) 273.2(4.5)	24.7(2.3) 271.1(3.8)	6.3(1.3) 254.4(4.6)	0.8
OTHER	141	21,131(10%)	7.2(3.3) 270.2(5.7)	16.1(2.9) 274.1(10.9)	21.4(3.9) 280.7(5.1)	47.9(6.4) 295.2(5.5)	7.3(2.3) 266.6(11.8)	1.3
PARENTAL EDUCATION								
LESS THAN H.S.	385	39,295(8%)	100.0(0.0) 271.4(2.1)	0.[illegible] 0.0)	0.0(0.0)	0.0(0 0)	0.0(0.0)	0.0
GRADUATED H.S.	1069	128,097(5%)	0.0(0.0)	100.0(0.0) 278.2(1.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
SOME EDUC AFTER H.S.	1057	127,961(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 288.6(1.2)	0.0(0.0)	0.[illegible] 0.0)	0.0
GRADUATED COLLEGE	1643	214,061(3%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 295.9(1.2)	0.0(0.0)	0.0
UNKNOWN	97	10,952(10%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 259.5(6.0)	0.0

* CV is the coefficient of variation for the sum of the weights.

462

Table 16-80

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors
Grade 4, by Gender, for Main Writing Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)**	MALE	FEMALE	MISSING
--TOTAL--	6677	1,270,697(1%)	50.4(0.9)	49.6(0.9)	0.0
			184.4(1.3)	197.5(1.3)	
GENDER					
MALE	3356	640,162(2%)	100.0(0.0)	0.0(0.0)	0.0
			184.4([illegible].3)		
FEMALE	3321	630,535(2%)	0.0(0.0)	100.0(0.0)	0.0
				197.5(1.3)	
RACE/ETHNICITY					
WHITE	3707	881,758(1%)	50.4(1.1)	49.6(1.1)	0.0
			191.1(1.6)	204.3(1.7)	
BLACK	1206	192,142(2%)	45.3(1.7)	54.7(1.7)	0.0
			161.9(3.3)	174.5(1.8)	
HISPANIC	1405	144,578(2%)	56.9(1.3)	43.1(1.3)	0.0
			172.1(2.0)	186.2(2.8)	
OTHER	359	52,218(5%)	50.2(2.6)	49.8(2.6)	0.0
			183.9(4.9)	204.2(4.4)	
PARENTAL EDUCATION					
LESS THAN HS	358	62,539(7%)	46.3(2.9)	53.7(2.9)	0.0
			172.9(4.4)	181.9(3.7)	
GRADUATED HS	978	194,058(4%)	49.5(1.8)	50.5(1.8)	0.0
			178.1(2.2)	194.3(2.5)	
SOME EDUC AFTER HS	508	99,842(5%)	49.1(2.1)	50.9(2.[illegible])	0.0
			195.7(3.3)	208.2(3.3)	
GRADUATED COLLEGE	2403	479,780(4%)	52.2(1.2)	47.8(1.2)	0.0
			194.0(1.5)	208.0(1.8)	
UNKNOWN	2385	428,174(4%)	49.5(1.2)	50.5(1.2)	0.0
			175.4(1.9)	188.0(2.0)	

* Average response method proficiency score.
** CV is the coefficient of variation for the sum of the weights.

Table 16-81

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors
Grade 4, by Derived Race/ethnicity, for Main Writing Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)**	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
--TOTAL--	6679	1,270,904(1%)	69.4(0.4)	15.1(0.3)	11.4(0.3)	1.9(0.2)	2.1(0.2)	0.1(0.0)	0.0
			197.6(1.3)	168.8(1.9)	178.2(2.0)	209.0(6.9)	181.0(4.5)	188.3(13.3)	
GENDER									
MALE	3357	640,295(2%)	69.4(0.7)	13.6(0.5)	12.9(0.4)	1.7(0.2)	2.3(0.2)	0.0(0.0)	0.0
			191.1(1.6)	161.9(3.2)	172.1(2.0)	197.7(6.7)	173.4(7.2)	187.2(34.3)	
FEMALE	3322	630,610(2%)	69.3(0.7)	16.7(0.6)	9.9(0.4)	2.0(0.3)	2.0(0.2)	0.1(0.1)	0.0
			204.3(1.7)	174.5(1.8)	186.2(2.8)	218.7(9.2)	190.1(5.0)	188.6(17.2)	
RACE/ETHNICITY									
WHITE	3707	881,758(1%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			197.6(1.3)						
BLACK	1207	192,275(2%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				168.8(1.9)					
HISPANIC	1406	144,65[illegible](2%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					178.2(2.0)				
OTHER	359	52,218(5%)	0.0(0.0)	0.0(0.0)	0.[illegible](0.0)	46.1(4.1)	52.2(4.0)	1.8(1.0)	0.0
						209.0(6.9)	181.0(4.5)	188.3(13.3)	
PARENTAL EDUCATION									
LESS THAN HS	358	62,539(7%)	60.3(2.5)	19.6(2.5)	16.9(2.1)	1.6(1.1)	1.6(0.6)	0.0(0.0)	0.0
			183.3(4.2)	165.5(6.5)	174.6(6.9)	173.8(61.6)	152.4(12.1)		
GRADUATED HS	978	194,058(4%)	71.1(1.3)	17.8(1.3)	8.1(0.8)	0.8(0.3)	2.2(0.3)	0.0(0.0)	0.0
			191.8(2.1)	166.9([illegible].7)	177.9(4.1)	229.0(24.8)	183.6(10.0)		
SOME EDUC AFTER HS	509	99,917(5%)	75.1(2.0)	10.8(1.6)	10.3(1.1)	1.4(0.5)	2.2(0.6)	0.3(0.3)	0.0
			206.6(3.1)	176.2(7.8)	192.4(5.2)	231.5(1[illegible])	205.2(13.6)	178.8(5.7)	
GRADUATED COLLEGE	2404	479,913(4%)	73.3(1.2)	14.0(0.9)	9.0(0.6)	1.8(0.3)	1.8(0.3)	0.1(0.0)	0.0
			207.2(1.5)	174.5([illegible])	186.9(3.2)	219.1(6.2)	187.6(8.2)	189.2(18.4)	
UNKNOWN	2335	428,174(4%)	64.5(1.0)	15.2(0.9)	15.0(0.7)	2.7(0.4)	2.6(0.3)	0.1(0.0)	0.0
			188.2(1.8)	163.4(2.3)	171.0(2.6)	199.0(7.[illegible])	172.9(6.0)	195.9(37.1)	

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.

Table 16-82

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors
Grade 4, by Parental Education, for Main Writing Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)**	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
--TOTAL--	6679	1,264,600(1%)	4.9(0.4) 177.7(2.9)	15.3(0.6) 186.3(1.9)	7.9(0.4) 202.1(2.5)	37.9(1.4) 200.7(1.3)	33.9(1.1) 181.7(1.1)	0.5
GENDER								
MALE	3357	636,711(2%)	4.5(0.5) 172.9(4.4)	15.1(0.7) 178.1(2.2)	7.7(0.6) 195.7(3.3)	39.4(1.5) 194.0(1.4)	33.3(1.1) 175.4(1.9)	0.6
FEMALE	3322	627,890(2%)	5.3(0.4) 181.9(3.7)	15.6(0.7) 194.3(2.5)	8.1(0.5) 208.2(3.3)	36.5(1.6) 208.0(1.8)	34.5(1.4) 188.0(2.0)	0.4
RACE/ETHNICITY								
WHITE	3707	878,914(1%)	4.3(0.4) 183.3(4.2)	15.7(0.7) 191.8(2.1)	8.5(0.5) 206.6(3.1)	40.0(1.7) 207.2(1.5)	31.4(1.3) 188.2(1.8)	0.3
BLACK	1207	189,810(2%)	6.5(1.0) 165.5(6.5)	18.2(1.2) 166.9(3.7)	5.7(0.8) 176.2(7.8)	35.4(1.9) 174.5(4.2)	34.2(2.2) 163.4(2.3)	1.3
HISPANIC	1406	143,751(3%)	7.3(0.8) 174.6(6.9)	11.0(1.2) 177.9(4.1)	7.2(0.7) 192.4(5.2)	29.9(1.7) 186.9(3.2)	44.6(1.4) 171.0(2.6)	0.6
OTHER	359	52,126(5%)	3.8(1.3) 163.0(16.6)	11.1(2.2) 195.5(11.2)	7.3(1.6) 212.9(9.8)	33.9(2.8) 203.2(4.9)	43.9(3.1) 186.1(4.5)	0.2
PARENTAL EDUCATION								
LESS THAN HS	358	62,539(7%)	100.0(0.0) 177.7(2.9)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED HS	978	194,058(4%)	0.0(0.0)	100.0(0.0) 186.3(1.9)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
SOME EDUC AFTER HS	509	99,917(5%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 202.1(2.5)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED COLLEGE	2404	479,913(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 200.7(1.3)	0.0(0.0)	0.0
UNKNOWN	2385	428,[illegible]74(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 181.7(1.1)	0.0

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.

Table 16-83

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors
Grade 8, by Gender, for Main Writing Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)**	MALE	FEMALE	MISSING
--TOTAL--	6525	743,791(1%)	51.1(0.8) 200.9(1.4)	48.9(0.8) 218.5(1.2)	0.0
GENDER					
MALE	3294	379,715(2%)	100.0(0.0) 200.9(1.4)	0.0(0.0)	0.0
FEMALE	3231	364,076(2%)	0.0(0.0)	100.0(0.0) 218.5(1.2)	0.0
RACE/ETHNICITY					
WHITE	4143	537,337(2%)	50.8(0.9) 207.2(1.6)	49.2(0.9) 225.0(1.1)	0.0
BLACK	1119	108,142(3%)	49.9(2.0) 179.4(2.5)	50.1(2.0) 195.7(2.6)	0.0
HISPANIC	966	69,963(3%)	53.6(2.1) 184.5(2.6)	46.4(2 1) 201.5(3.4)	0.0
OTHER	297	28,349(4%)	53.4(2.3) 205.0(4.2)	46.6(2.3) 221.7(5.7)	0.0
PARENTAL EDUCATION					
LESS THAN HS	603	59,584(6%)	39.8(2.3) 185.2(4.1)	60.2(2.3) 204.1(2.4)	0.0
GRADUATED HS	1706	198,188(3%)	52.0(1.4) 196.3(1.9)	48.0(1.4) 214.6(2.4)	0.0
SOME EDUC AFTER HS	1178	135,639(4%)	47.9(1.8) 207.3(2.6)	52.1(1.8) 224.2(2.0)	0.0
GRADUATED COLLEGE	2444	292,035(3%)	52.9(1.1) 208.3(2.1)	47.1(1.1) 225.8(1.9)	0.0
UNKNOWN	573	56,300(6%)	57.8(2.3) 179.3(3.0)	42.2(2.3) 196.7(3.6)	0.0

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.

Table 16-84

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors Grade 8, by Derived Race/ethnicity, for Main Writing Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)**	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
--TOTAL--	6525	743,791(1%)	72.2(0.5)	14.5(0.4)	9.4(0.2)	2.5(0.2)	1.2(0.2)	0.1(0.0)	0.0
			216.0(1.0)	187.5(2.2)	192.4(2.1)	222.5(5.0)	194.1(5.1)	193.8(17.5)	
GENDER									
MALE	3294	379,715(2%)	71.9(0.9)	14.2(0.7)	9.9(0.5)	2.5(0.2)	1.4(0.2)	0.1(0.0)	0.0
			207.2(1.6)	179.4(2.5)	184.5(2.6)	216.3(5.7)	186.1([illegible].6)	191.2(24.7)	
FEMALE	3231	364,076(2%)	72.6(0.8)	14.9(0.6)	8.9(0.4)	2.5(0.2)	1.0(0.2)	0.1(0.1)	0.0
			225.0(1.1)	195.7(2.6)	201.5(3.4)	228.9(7.4)	206.3(8.5)	195.6(19.3)	
RACE/ETHNICITY									
WHITE	4143	537,337(2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			216.0(1.0)						
BLACK	1119	108,142(3%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				187.5(2.2)					
HISPANIC	966	69,963(3%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					192.4(2.1)				
OTHER	297	28,349(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	65.8(4.4)	31.9(4.2)	2.3(1.2)	0.0
						222.5(5.0)	194.1(5.1)	193.8(17.5)	
PARENTAL EDUCATION									
LESS THAN HS	603	59,584(6%)	59.6(2.5)	15.2(1.7)	23.2(1.8)	1.2(0.5)	0.8(0.5)	0.0(0.0)	0.0
			203.3(2.8)	183.5(6.1)	187.8(4.8)	204.5(13.2)	182.0(16.4)		
GRADUATED HS	1706	198,188(3%)	72.1(1.3)	16.7(1.1)	8.6(0.6)	1.2(0.3)	1.4(0.3)	0.1(0.1)	0.0
			211.9(2.2)	182.7(4.9)	192.2(3.7)	214.9(10.9)	194.7(11.5)	209.8(22.9)	
SOME EDUC AFTER HS	1178	135,639(4%)	75.7(1.0)	14.0(0.8)	6.9(0.6)	1.6(0.4)	1.8(0.4)	0.0(0.0)	0.0
			220.7(1.8)	200.4(4.0)	199.5(6.2)	222.6(10.6)	207.1(7.3)	183.5(22.6)	
GRADUATED COLLEGE	2444	292,035(3%)	78.5(1.0)	11.7(0.8)	5.4(0.4)	3.6(0.4)	0.6(0.2)	0.0(0.0)	0.0
			220.7(1.5)	190.7(2.6)	204.0(3.9)	230.8(7.5)	203.5(14.3)	240.8(11.0)	
UNKNOWN	573	56,300(6%)	46.1(3.2)	22.1(1.8)	24.0(2.1)	5.1(0.8)	2.6(0.8)	0.1(0.1)	0.0
			195.5(3.4)	175.5(4.7)	178.8(3.9)	202.7(7.1)	164.9(13.0)	164.4(16.5)	

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.

Table 16-85

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors Grade 8, by Parental Education for Main Writing Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)**	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
--TOTAL--	6525	741,746(1%)	8.0(0.4)	26.7(0.8)	18.3(0.6)	39.4(1.0)	7.6(0.5)	0.3
			196.6(2.1)	205.1(1.8)	216.1(1.6)	216.5(1.5)	186.6(2.2)	
GENDER								
MALE	3294	378,715(2%)	6.3(0.5)	27.2(1.1)	17.2(0.9)	40.8(1.1)	8.0(0.5)	0.3
			185.2(4.1)	196.3(1.9)	207.3(2.6)	208.3(2.1)	179.3(3.0)	
FEMALE	3231	363,032(2%)	9.9(0.6)	26.2(1.0)	19.5(0.7)	37.9(1.3)	6.5(0.6)	0.3
			204.1(2.4)	214.6(2.4)	224.2(2.0)	225.8(1.9)	196.7(3.6)	
RACE/ETHNICITY								
WHITE	4143	536,319(2%)	6.6(0.5)	26.6(1.0)	19.1(0.7)	42.8(1.2)	[illegible].8(0.5)	0.2
			203.3(2.8)	211.9(2.2)	220.7(1.8)	220.7(1.5)	195.5(3.4)	
BLACK	1119	107,806(3%)	8.4(0.8)	30.6(1.7)	17.6(1.1)	31.8(2.0)	11.5(0.9)	0.3
			183.5(6.1)	182.7(4.9)	200.4(4.0)	190.7(2.6)	175.5(4.7)	
HISPANIC	966	69,604(3%)	19.8(1.6)	24.6(1.7)	13.4(1.2)	22.8(1.8)	19.4(1.9)	0.5
			187.8(4.8)	182.2(3.7)	199.5(6.2)	204.0(3.9)	178.[illegible](3.[illegible])	
OTHER	297	28,018(4%)	4.2(1.4)	18.7(3.0)	16.7(2.4)	44.6(3.7)	15.7(2.0)	1.2
			195.6(11.5)	204.1(8.8)	213.9(5.7)	226.9(7.2)	189.4(6.2)	
PARENTAL EDUCATION								
LESS THAN HS	603	59,5[illegible]4(6%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			196.6(2.1)					
GRADUATED HS	1706	198,188(3%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				205.1(1.8)				
SOME EDUC AFTER HS	1178	135,639(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					216.1(1.6)			
GRADUATED COLLEGE	2444	292,035(3%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
						216.5(1.5)		
UNKNOWN	573	56,300(6%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0
							186.6(2.2)	

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.

Table 16-86

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors
Grade 12, by Gender, for Main Writing Samples

GENDER OF SUBJECT (NAEP ID: [illegible])

	N	WEIGHTED N (CV)**	MALE	FEMALE	MISSING
--TOTAL--	6069	729,819(2%)	47.7(1.2) 212.5(1.4)	52.3(1.2) 234.8(1.8)	0.0
GENDER					
MALE	2934	347,787(3%)	100.0(0.0) 212.5(1.4)	0.0(0.0)	0.0
FEMALE	3135	382,032(3%)	0.0(0.0)	100.0(0.0) 234.8(1.8)	0.0
RACE/ETHNICITY					
WHITE	4178	548,476(2%)	47.5(1.4) 218.0(1.7)	52.5(1.4) 241.8(1.8)	0.0
BLACK	1095	103,870(4%)	46.8(2.1) 191.9(2.6)	53.2(2.1) 208.5(3.4)	0.0
HISPANIC	612	52,802(5%)	48.5(3.0) 193.2(4.3)	51.5(3.0) 216.0(4.5)	0.0
OTHER	184	24,671(8%)	53.3(4.8) 218.1(6.6)	46.7(4.8) 228.1(6.9)	0.0
PARENTAL EDUCATION					
LESS THAN HS	541	54,507(7%)	41.9(2.2) 198.1(3.6)	58.1(2.2) 216.6(4.5)	0.0
GRADUATED HS	1524	179,919(4%)	49.3(1.6) 208.5(2.5)	50.7(1.6) 229.0(2.7)	0.0
SOME EDUC AFTER HS	1475	175,858(3%)	44.1(1.6) 215.6(2.6)	55.9(1.6) 235.7(2.3)	0.0
GRADUATED COLLEGE	2356	301,753(3%)	49.1(1.6) 218.1(1.8)	50.9(1.8) 243.4(2.6)	0.0
UNKNOWN	144	14,413(12%)	62.5(4.5) 178.8(7.7)	37.5(4.5) 188.5(9.3)	0.0

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.

Table 16-87

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors
Grade 12, by Derived Race/ethnicity, for Main Writing Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)**	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
--TOTAL--	6069	729,819(2%)	75.2(0.7) 230.5(1.5)	14.2(0.6) 200.7(2.3)	7.2(0.3) 204.9(3.5)	2.8(0.3) 225.1(5.4)	0.5(0.1) 212.1(17.0)	0.1(0.0) 214.8(34.1)	0.0
GENDER									
MALE	2934	347,787(3%)	74.9(1.0) 218.0(1.7)	14.0(1.0) 191.9(2.6)	7.4(0.5) 193.2(4.3)	3.0(0.4) 220.7(5.8)	0.7(0.2) 209.2(18.5)	0.1(0.1) 203.6(50.3)	0.0
FEMALE	3135	382,032(3%)	75.4(0.9) 241.8(1.8)	14.5(0.6) 208.5(3.4)	7.1(0.6) 216.0(4.5)	2.6(0.4) 229.8(8.2)	0.4(0.2) 216.7(23.6)	0.0(0.0) *****(****)	0.0
RACE/ETHNICITY									
WHITE	4178	548,476(2%)	100.0(0.0) 230.5(1.5)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
BLACK	1095	103,870(4%)	0.0(0.0)	100.0(0.0) 200.7(2.3)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
HISPANIC	612	52,802(5%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 204.9(3.5)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
OTHER	184	24,671(8%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	81.8(4.2) 225.1(5.4)	16.2(3.9) 212.1(17.0)	2.0(1.1) 214.8(34.1)	0.0
PARENTAL EDUCATION									
LESS THAN HS	541	54,507(7%)	55.1(3.1) 220.2(4.3)	18.7(2.8) 189.4(5.4)	22.2(2.0) 197.0(4.7)	2.9(0.9) 212.7(21.3)	0.9(0.5) 196.7(34.0)	0.2(0.2) *****(****)	0.0
GRADUATED HS	1524	179,919(4%)	74.1(1.6) 225.2(2.2)	16.7(1.3) 196.2(4.7)	6.6(0.7) 207.8(5.5)	1.9(0.5) 220.2(10.3)	0.7(0.3) 198.1(33.1)	0.0(0.0) *****(****)	0.0
SOME EDUC AFTER HS	1475	175,858(3%)	77.2(1.3) 231.3(1.9)	14.6(1.0) 206.9(4.4)	6.4(0.8) 216.8(5.8)	1.3(0.4) 222.3(10.7)	0.4(0.2) 256.6(24.8)	0.1(0.1) *****(****)	0.0
GRADUATED COLLEGE	2356	301,753(3%)	80.1(1.2) 235.5(2.1)	11.3(0.9) 207.3(3.3)	4.0(0.5) 208.2(6.3)	4.1(0.5) 230.6(8.4)	0.4(0.2) 220.3(22.4)	0.0(0.0)	0.0
UNKNOWN	144	14,413(12%)	37.5(4.0) 193.6(12.2)	22.6(4.2) 166.1(14.5)	35.3(4.0) 182.1(6.0)	2.6(1.3) 180.9(28.6)	2.0(1.5) 166.0(40.0)	0.0(0.0)	0.0

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.
*****(****) The sample size is insufficient to provide reliable estimates.

473

472

Table 16-88

Weighted Response Percentages and Writing Proficiency* Means with Standard Errors
Grade 12, by Parental Education, for Main Writing Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)**	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
--TOTAL--	6069	726,451(2%)	7.5(0.5)	24.8(0.8)	24.2(0.7)	41.5(1.1)	2.0(0.2)	0.5
			208.9(3.0)	218.9(1.9)	226.8(1.7)	231.0(1.8)	182.4(6.0)	
GENDER								
MALE	2934	346,235(3%)	6.6(0.5)	25.6(1.1)	22.4(0.9)	42.8(1.4)	2.6(0.3)	0.4
			198.1(3.6)	208.5(2.5)	215.6(2.6)	218.1(1.8)	178.8(7.7)	
FEMALE	3135	380,216(3%)	8.3(0.6)	24.0(0.9)	25.9(1.0)	40.4(1.5)	1.4(0.3)	0.5
			216.6(4.5)	229.0(2.7)	235.7(2.3)	243.4(2.6)	188.5(9.3)	
RACE/ETHNICITY								
WHITE	4178	546,337(2%)	5.5(0.5)	24.4(1.1)	24.9(0.8)	44.2(1.2)	1.0(0.2)	0.4
			220.2(4.3)	225.2(2.2)	231.3(1.9)	235.5(2.1)	193.6(12.2)	
BLACK	1095	103,257(4%)	9.9(1.6)	29.0(1.6)	24.9(1.7)	33.1(2.5)	3.1(0.8)	0.6
			189.4(5.4)	196.2(4.7)	206.9(4.4)	207.3(3.3)	166.1(14.5)	
HISPANIC	612	52,577(5%)	23.0(2.1)	22.6([illegible].9)	21.6(2.6)	23.1(3.0)	9.7(1.0)	0.4
			197.0(4.7)	207.8(5.5)	216.8(5.8)	208.2(6.3)	182.1(6.0)	
OTHER	184	24,280(8%)	8.8(2.3)	19.4(3.5)	12.5(2.9)	56.6(5.1)	2.7(1.2)	1.6
			209.3(17.7)	213.9(9.9)	230.1(9.5)	229.6([illegible].6)	174.5(23.1)	
PARENTAL EDUCATION								
LESS THAN HS	541	54,507(7%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			208.9(3.0)					
GRADUATED HS	1524	179,919(4%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				218.9(1.9)				
SOME EDUC AFTER HS	1475	175,858(3%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					226.8(1.7)			
GRADUATED COLLEGE	2356	301,753(3%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
						231.0(1.8)		
UNKNOWN	144	14,413(12%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0
							182.4(6.0)	

---

* Average response method proficiency score
** CV is the coefficient of variation for the sum of the weights.

Table 16-89

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 4, by Gender, for Main Focused-BIB Civics Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	1974	375,529(1%)	49.1(1.3) 214.8(1.3)	50.9(1.3) 213.3(1.1)	0.0
GENDER					
MALE	978	184,226(3%)	100.0(0.0) 214.8(1.3)	0.0(0.0)	0.0
FEMALE	996	191,303(3%)	0.0(0.0)	100.0(0.0) 213.3(1.1)	0.0
RACE/ETHNICITY					
WHITE	1122	262,686(2%)	47.8(1.5) 220.6(1.5)	52.2(1.5) 219.5(1.2)	0.0
BLACK	360	58,616(5%)	48.0(3.6) 199.7(2.8)	52.0(3.6) 196.7(2.9)	0.0
HISPANIC	390	40,037(5%)	56.4(2.4) 202.1(2.4)	43.6(2.4) 196.1(2.3)	0.0
OTHER	102	14,190(8%)	56.5(4.9) 213.6(4.9)	43.5(4.9) 205.4(7.6)	0.0
PARENTAL EDUCATION					
LESS THAN H.S.	116	17,487(11%)	39.7(4.9) 202.9(4.5)	60.3(4.9) 210.6(4.7)	0.0
GRADUATED H.S.	269	50,576(7%)	56.2(3.1) 209.2(2.5)	43.8(3.1) 213.8(2.6)	0.0
SOME EDUC AFTER H.S.	152	31,922(10%)	45.3(4.6) 221.9(4.2)	54.7(4.6) 221.0(3.2)	0.0
GRADUATED COLLEGE	712	146,718(5%)	51.4(2.1) 222.8(1.8)	48.6(2.1) 222.3(1.7)	0.0
UNKNOWN	716	127,468(4%)	45.4(2.3) 207.7(1.7)	54.6(2.3) 202.5(2.2)	0.0

* CV is the coefficient of variation for the sum of the weights.

Table 16-90

Weighted Response Percentages and Civics Proficiency Means 1th Standard Errors
Grade 4, by Derived Race/ethnicity, for Main Focused-BIB Civics Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	1974	375,529(1%)	70.0(0.9)	15.6(0.8)	10.7(0.6)	2.0(0.3)	1.7(0.3)	0.1(0.1)	0.0
			220.0(1.0)	198.1(2.2)	199.5(1.9)	210.3(8.3)	209.6(4.5)	211.7(21.0)	
GENDER									
MALE	978	184,226(3%)	68.1(1.3)	15.3(1.0)	12.3(1.0)	2.4(0.5)	2.0(0.4)	0.0(0.0)	0.0
			220.6(1.5)	199.7(2.8)	202.1(2.4)	215.7(7.0)	211.0(6.9)		
FEMALE	996	191,303(3%)	71.7(1.5)	15.9(1.5)	9.1(0.7)	1.6(0.4)	1.5(0.4)	0.1(0.1)	0.0
			219.5(1.2)	196.7(2.9)	196.1(2.3)	202.2(14.5)	207.9(6.9)	211.7(21.0)	
RACE/ETHNICITY									
WHITE	1122	262,686(2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			220.0(1.0)						
BLACK	360	58,616(5%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				198.1(2.2)					
HISPANIC	390	40,037(5%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					199.5(1.9)				
OTHER	102	14,190(8%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	51.8(6.4)	46.3(6.5)	1.9(1.4)	0.0
						210.3(8.3)	209.6(4.5)	211.7(21.0)	
PARENTAL EDUCATION									
LESS THAN H.S.	116	17,487(11%)	59.8(6.0)	16.7(3.7)	17.8(3.6)	2.5(2.5)	2.2(1.3)	1.0(1.0)	0.0
			214.2(4.1)	197.1(8.8)	202.2(4.5)	179.4(****)	184.3(19.1)	200.7(16.3)	
GRADUATED H.S.	269	50,576(7%)	68.2(2.8)	18.5(1.8)	10.4(1.7)	1.2(0.6)	1.7(0.8)	0.0(0.0)	0.0
			216.3(2.0)	199.4(4.8)	200.3(4.4)	212.3(4.9)	200.6(12.0)		
SOME EDUC AFTER H.S.	152	31,922(10%)	73.2(3.6)	15.3(2.8)	7.9(2.0)	0.3(0.3)	3.3(1.6)	0.0(0.0)	0.0
			226.8(3.0)	198.3(6.5)	215.1(6.5)	207.5(****)	224.5(7.3)		
GRADUATED COLLEGE	712	146,718(5%)	74.8(1.6)	13.4(1.2)	7.8(1.1)	2.4(0.5)	1.6(0.4)	0.0(0.0)	0.0
			227.3(1.3)	204.8(2.7)	205.9(3.2)	227.0(6.5)	221.3(9.0)		
UNKNOWN	715	127,468(4%)	66.1(1.9)	16.3(1.6)	13.9(1.1)	2.1(0.5)	1.5(0.4)	0.1(0.1)	0.0
			211.1(1.6)	192.4(3.4)	192.4(2.6)	192.6(10.4)	196.4(4.7)	231.3(****)	

* CV is the coefficient o variation for the sum of the weights.
(****) Standard error is eater than 99.9.

Table 16-91

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 4, by Parental Education, for Main Focused-BIB Civics Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	1985	374,170(1%)	4.7(0.5) 207.5(3.5)	13.5(1.0) 211.2(1.8)	8.5(0.9) 221.4(2.8)	39.2(1.6) 222.5(1.3)	34.1(1.3) 204.8(1.4)	0.4
GENDER								
MALE	971	183,131(3%)	3.8(0.8) 202.9(4.5)	15.5(1.5) 209.2(2.5)	7.9(1.1) 221.9(4.2)	41.2(2.0) 222.8(1.8)	31.6(1.4) 207.7(1.7)	0.6
FEMALE	994	191,039(3%)	5.5(0.8) 210.6(4.7)	11.6(1.1) 213.8(2.6)	9.1(1.1) 221.0(3.2)	37.3(2.2) 222.3(1.7)	36.4(2.0) 202.5(2.2)	0.1
RACE/ETHNICITY								
WHITE	1120	262,280(2%)	4.0(0.7) 214.2(4.1)	13.1(1.3) 216.3(2.0)	8.9(1.0) 226.8(3.0)	41.8(1.9) 227.3(1.3)	32.1(1.6) 211.1(1.6)	0.2
BLACK	353	57,663(6%)	5.1(1.2) 197.1(8.8)	16.3(1.6) 199.4(4.8)	8.5(1.8) 198.3(6.5)	34.1(2.6) 204.8(2.7)	36.1(3.0) 192.4(3.4)	1.6
HISPANIC	390	40,037(5%)	7.8(1.6) 202.2(4.5)	13.1(2.1) 200.3(4.4)	6.3(1.6) 215.1(6.5)	28.5(3.5) 205.9(3.2)	44.3(3.2) 192.4(2.6)	0.0
OTHER	102	14,190(8%)	7.0(3.3) 185.1(10.4)	10.2(3.3) 205.4(8.2)	8.1(3.6) 223.1(7.1)	41.6(6.4) 224.7(5.0)	33.1(4.9) 195.0(6.3)	0.0
PARENTAL EDUCATION								
LESS THAN H.S.	116	17,487(11%)	100.0(0.0) 207.5(3.5)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED H.S.	269	50,576(7%)	0.0(0.0)	100.0(0.0) 211.2(1.8)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
SOME EDUC AFTER H.S.	152	31,922(10%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 221.4(2.8)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED COLLEGE	712	146,718(5%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 222.5(1.3)	0.0(0.0)	0.0
UNKNOWN	716	127,468(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 204.8(1.4)	0.0

* CV is the coefficient of variation for the sum of the weights.

478

479

Table 16-92

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 8, by Gender, for Main Focused-BIB Civics Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	4486	504,400(1%)	48.8(0.8) 258.7(1.1)	51.2(0.8) 260.6(0.9)	0.0
GENDER					
MALE	2155	246,135(2%)	100.0(0.0) 258.7(1.1)	0.0(0.0)	0.0
FEMALE	2331	258,265(2%)	0.0(0.0)	100.0(0.0) 260.6(0.9)	0.0
RACE/ETHNICITY					
WHITE	2708	351,704(2%)	48.5(1.1) 265.9(1.6)	51.5(1.1) 266.7(1.3)	0.0
BLACK	805	78,364(4%)	47.7(1.5) 242.2(2.7)	52.3(1.5) 245.0(1.9)	0.1
HISPANIC	736	52,422(4%)	51.3(2.1) 238.1(2.4)	48.7(2.1) 243.2(2.0)	0.0
OTHER	237	21,909(7%)	50.7(4.0) 253.3(5.0)	49.3(4.0) 258.2(3.9)	0.0
PARENTAL EDUCATION					
LESS THAN H.S.	410	39,746(5%)	44.5(2.8) 234.8(3.2)	55.5(2.8) 240.3(2.8)	0.0
GRADUATED H.S.	1225	136,360(4%)	47.4(1.8) 250.9(1.9)	52.6(1.8) 253.9(1.3)	0.0
SOME EDUC AFTER H.S.	844	97,514(4%)	46.6(1.4) 262.2(1.9)	53.4(1.4) 264.9(1.5)	0.0
GRADUATED COLLEGE	1630	193,521(4%)	49.6(1.4 272.4(1.6)	50.4(1.4) 272.0(1.4)	0.0
UNKNOWN	357	35,605(6%)	59.7(3.6) 235.8(3.9)	40.3(3.6) 235.9(3.2)	0.2

* CV is the coefficient of variation for the sum of the weights.

Table 16-93

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 8, by Derived Race/ethnicity, for Main Focused-BIB Civics Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	4487	504,472(1%)	69.7(0.6) 266.3(1.2)	15.5(0.5) 243.6(1.9)	10.4(0.4) 240.6(1.7)	2.8(0.3) 262.7(4.1)	1.4(0.2) 242.3(4.4)	0.1(0.0) 254.5(18.5)	0.0
GENDER									
MALE	2156	246,208(2%)	69.4(0.9) 265.9(1.6)	15.2(0.7) 242.1(2.7)	10.9(0.7) 238.1(2.4)	2.6(0.3) 262.5(6.2)	1.7(0.3) 239.8(7.2)	0.2(0.1) 246.6(19.5)	0.0
FEMALE	2331	258,265(2%)	70.1(0.9) 266.7(1.3)	15.9(0.7) 245.0(1.9)	9.9(0.5) 243.2(2.0)	3.0(0.5) 262.9(4.5)	1.2(0.2) 245.7(4.4)	0.0(0.0) 292.6(****)	0.0
RACE/ETHNICITY									
WHITE	2708	351,704(2%)	100.0(0.0) 266.3(1.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
BLACK	806	78,437(4%)	0.0(0.0)	100.0(0.0) 243.6(1.9)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
HISPANIC	736	52,422(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 240.6(1.7)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
OTHER	237	21,909(7%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	64.5(4.5) 262.7(4.1)	33.4(4.5) 242.3(4.4)	2.1(1.0) 254.5(18.5)	0.0
PARENTAL EDUCATION									
LESS THAN H.S.	410	39,746(5%)	56.5(3.5) 243.4(2.8)	11.3(1.7) 231.5(4.5)	28.2(2.9) 228.9(3.3)	1.9(0.6) 249.4(20.1)	2.0(0.7) 231.4(11.1)	0.0(0.0)	0.0
GRADUATED H.S.	1228	136,360(4%)	67.7(1.5) 258.0(1.6)	17.8(1.3) 238.9(2.6)	10.5(1.0) 240.7(2.5)	1.7(0.3) 247.2(6.8)	2.3(0.4) 251.5(5.4)	0.0(0.0)	0.0
SOME EDUC AFTER H.S.	844	97,514(4%)	73.6(1.8) 267.5(1.3)	15.6(1.5) 252.7(2.9)	8.7(0.7) 251.3(3.1)	1.2(0.4) 270.2(9.1)	0.8(0.3) 248.3(7.7)	0.1(0.1) 283.7(16.2)	0.0
GRADUATED COLLEGE	1630	193,521(4%)	75.6(1.2) 277.7(1.4)	14.1(0.9) 250.7(2.9)	5.6(0.6) 251.8(3.5)	3.9(0.5) 273.8(5.2)	0.6(0.2) 256.6(6.8)	0.1(0.1) 269.2(10.8)	0.0
UNKNOWN	358	35,677(6%)	49.8(3.5) 244.1(3.6)	19.2(2.0) 223.1(4.6)	20.6(2.1) 229.1(4.4)	6.5(2.6) 242.5(11.6)	3.6(1.3) 207.2(10.3)	0.3(0.3) 201.1(15.5)	0.0

* CV is the coefficient of variation for the sum of the weights.
(****) Standard error is greater than 99.9.

481 482

Table 1b 94

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 8, by Parental Education, for Main Focused-BIB Civics Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	4470	502,817(1%)	7.9(0.4)	27.1(1.0)	19.4(0.7)	38.5(1.4)	7.1(0.4)	0.3
			237.8(2.2)	252.5(1.3)	263.7(1.3)	272.2(1.1)	235.4(2.7)	
GENDER								
MALE	2145	245,061(2%)	7.2(0.6)	26.4(1.3)	18.6(0.9)	39.2(1.8)	8.7(0.7)	0.5
			234.3(3.2)	250.9(1.9)	262.2(1.9)	272.4(1.6)	235.7(3.9)	
FEMALE	2325	257,756(2%)	8.6(0.6)	27.8(1.3)	20.2(0.9)	37.8(1.4)	5.6(0.6)	0.2
			240.3(2.8)	253.9(1.3)	264.9(1.5)	272.0(1.4)	235.0(3.2)	
RACE/ETHNICITY								
WHITE	2700	350,678(2%)	6.4(0.6)	26.3(1.2)	20.5(0.9)	41.7(1.8)	5.1(0.5)	0.3
			243.4(2.8)	258.0(1.6)	267.5(1.3)	277.7(1.4)	244.1(3.6)	
BLACK	801	77,984(4%)	5.8(1.0)	31.1(2.0)	19.5(1.8)	34.9(1.8)	8.8(1.0)	0.6
			231.5(4.5)	238.9(2.6)	252.7(2.9)	250.7(2.9)	223.1(4.6)	
HISPANIC	732	52,246(4%)	21.5(1.9)	27.4(2.2)	16.2(1.5)	20.9(2.0)	14.1(1.4)	0.3
			228.9(3.3)	240.7(2.5)	251.3(3.1)	251.8(3.5)	229.1(4.4)	
OTHER	237	21,909(7%)	7.2(1.6)	25.1(3.3)	9.4(2.5)	41.3(4.7)	17.0(4.6)	0.0
			240.2(10.4)	250.3(4.6)	262.2(6.9)	271.3(4.7)	228.9(11.3)	
PARENTAL EDUCATION								
LESS THAN H.S.	410	39,746(5%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			237.8(2.2)					
GRADUATED H.S.	1228	136,360(4%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				252.5(1.3)				
SOME EDUC AFTER H.S.	844	97,514(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					263.7(1.3)			
GRADUATED COLLEGE	1630	193,521(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
						272.2(1.1)		
UNKNOWN	358	35,677(6%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0
							235.4(2.7)	

* CV is the coefficient of variation for the sum of the weights.

Table 16-95

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 12, by Gender, for Main Focused-BIB Civics Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	4275	525,727(1%)	48.4(1.6) 298.6(1.6)	51.6(1.6) 294.1(1.1)	0.0
GENDER					
MALE	2049	254,193(3%)	100.0(0.0) 298.6(1.6)	0.0(0.0)	0.0
FEMALE	2226	271,533(4%)	0.0(0.0)	100.0(0.0) 294.1(1.1)	0.0
RACE/ETHNICITY					
WHITE	2960	397,203(2%)	47.8(1.8) 305.0(1.7)	52.2(1.8) 299.1(1.2)	0.0
BLACK	678	62,248(5%)	47.5(2.9) 278.1(3.1)	52.5(2.9) 269.9(2.3)	0.0
HISPANIC	487	43,903(4%)	52.1(2.3) 277.6(3.8)	47.9(2.8) 281.0(2.9)	0.0
OTHER	150	22,370(10%)	53.9(3.8) 289.4(7.2)	46.1(3.8) 297.5(8.0)	0.0
PARENTAL EDUCATION					
LESS THAN H.S.	374	38,467(7%)	44.9(3.4) 272.4(3.9)	55.1(3.4) 273.4(3.5)	0.0
GRADUATED H.S.	1026	125,616(4%)	47.2(2.4) 286.3(2.4)	52.8(2.4) 283.8(1.6)	0.0
SOME EDUC AFTER H.S.	1056	125,510(5%)	46.7(1.9) 300.9(2.5)	53.3(1.9) 296.6(1.5)	0.0
GRADUATED COLLEGE	1708	222,853(4%)	49.4(1.8) 311.2(1.9)	50.6(1.8) 303.8(1.6)	0.0
UNKNOWN	97	11,246(13%)	67.5(5.1) 266.8(7.9)	32.5(5.1) 263.7(7.7)	0.0

* CV is the coefficient of variation for the sum of the weights.

Table 16-96

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 12, by Derived Race/ethnicity, for Main Focused-BIB Civics Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	4275	525,727(1%)	75.6(0.8)	11.8(0.6)	8.4(0.4)	3.1(0.4)	1.0(0.3)	0.1(0.1)	0.0
			301.9(1.2)	273.8(1.9)	279.2(2.3)	302.5(5.5)	274.2(8.6)	218.2(22.5)	
GENDER									
MALE	2049	254,193(3%)	74.6(1.1)	11.6(0.8)	9.0(0.6)	3.3(0.5)	1.2(0.4)	0.3(0.2)	0.0
			305.0(1.7)	278.1(3.1)	277.6(3.6)	302.2(7.6)	271.1(10.1)	218.2(22.5)	
FEMALE	2226	271,533(4%)	76.4(1.1)	12.0(0.8)	7.7(0.5)	3.0(0.5)	0.8(0.4)	0.0(0.0)	0.0
			299.1(1.2)	269.9(2.3)	281.0(2.9)	302.9(6.3)	278.4(9.9)		
RACE/ETHNICITY									
WHITE	2960	397,203(2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			301.9(1.2)						
BLACK	678	62,248(5%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				273.8(1.9)					
HISPANIC	487	43,906(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					279.2(2.3)				
OTHER	150	22,370(10%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	73.0(7.1)	23.9(6.9)	3.1(2.0)	0.0
						302.5(5.5)	274.2(8.6)	218.2(22.5)	
PARENTAL EDUCATION									
LESS THAN H.S.	374	38,467(7%)	55.2(3.7)	14.5(2.6)	26.8(2.4)	2.8(1.0)	0.7(0.6)	0.0(0.0)	0.0
			277.4(3.4)	261.0(5.0)	270.8(4.1)	276.8(13.1)	238.6(24.5)		
GRADUATED H.S.	1026	125,616(4%)	74.4(1.9)	14.7(1.4)	7.9(0.8)	1.6(0.7)	1.4(0.6)	0.1(0.1)	0.0
			290.1(1.7)	265.5(2.8)	276.6(3.9)	288.5(19.4)	263.5(15.0)	270.6(24.3)	
SOME EDUC AFTER H.S.	1056	125,510(5%)	77.3(1.6)	13.0(1.1)	7.3(0.8)	1.3(0.5)	1.0(0.3)	0.0(0.0)	0.0
			303.4(1.6)	279.8(2.7)	279.9(5.3)	318.4(14.6)	285.3(12.2)		
GRADUATED COLLEGE	1708	222,853(4%)	80.5(1.1)	8.5(0.7)	5.3(0.5)	4.9(0.7)	0.7(0.4)	0.1(0.1)	0.0
			311.2(1.6)	284.7(2.9)	291.4(4.1)	308.0(5.5)	280.5(6.3)	203.4(9.7)	
UNKNOWN	97	11,246(13%)	45.6(6.8)	21.1(4.2)	24.5(4.9)	5.6(2.0)	3.1(2.2)	0.0(0.0)	0.0
			275.8(10.2)	241.2(6.3)	265.7(8.5)	266.7(32.2)	285.0(16.7)		

* CV is the coefficient of variation for the sum of the weights.

485

486

Table 16-97

Weighted Response Percentages and Civics Proficiency Means with Standard Errors
Grade 12, by Parental Education, for Main Focused-BIB Civics Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	4261	523,692(1%)	7.3(0.5)	24.0(1.1)	24.0(1.1)	42.6(1.6)	2.1(0.3)	0.4
			273.0(2.5)	285.0(1.6)	298.6(1.4)	307.4(1.5)	265.8(5.7)	
GENDER								
MALE	2041	252,871(3%)	6.8(0.6)	23.5(1.4)	23.2(1.4)	43.5(2.0)	3.0(0.5)	0.5
			272.4(3.9)	286.3(2.4)	300.9(2.5)	311.2(1.9)	266.8(7.9)	
FEMALE	2220	270,821(4%)	7.8(0.7)	24.5(1.1)	24.7(1.1)	41.6(1.7)	1.3(0.2)	0.3
			273.4(3.5)	283.8(1.6)	296.6(1.5)	303.8(1.6)	263.7(7.7)	
RACE/ETHNICITY								
WHITE	2952	396,245(2%)	5.4(0.6)	23.6(1.3)	24.5(1.3)	45.3(1.8)	1.3(0.3)	0.2
			277.4(3.4)	290.1(1.7)	303.4(1.6)	311.2(1.6)	275.8(10.2)	
BLACK	674	61,628(5%)	9.0(1.6)	29.9(2.0)	26.4(2.0)	30.9(2.4)	3.9(0.8)	1.0
			261.0(5.0)	265.5(2.8)	279.8(2.7)	284.7(2.9)	241.2(6.3)	
HISPANIC	487	43,906(4%)	23.5(2.1)	22.6(1.9)	20.9(2.2)	26.7(2.5)	6.3(1.3)	0.0
			270.8(4.1)	276.6(3.9)	279.9(5.3)	291.4(4.1)	265.7(8.5)	
OTHER	148	21,913(10%)	6.2(2.0)	17.5(4.3)	13.7(2.9)	58.1(6.6)	4.5(1.5)	2.0
			268.9(11.7)	276.5(10.2)	304.0(11.1)	302.8(5.8)	273.2(16.5)	
PARENTAL EDUCATION								
LESS THAN H.S.	374	38,467(7%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			273.0(2.5)					
GRADUATED H.S.	1026	125,616(4%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				285.0(1.6)				
SOME EDUC AFTER H.S.	1056	125,510(5%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					298.6(1.4)			
GRADUATED COLLEGE	1708	222,853(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
						307.4(1.5)		
UNKNOWN	97	11,246(13%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0
							265.8(5.7)	

* CV is the coefficient of variation for the sum of the weights.

457

Table 16-98

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 4, by Gender, for Main U.S. History Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	3950	751,664(1%)	50.7(0.9)	49.3(0.9)	0.0
			222.9(1.2)	218.2(1.0)	
GENDER					
MALE	2026	381,243(2%)	100.0(0.0)	0.0(0.0)	0.0
			222.9(1.2)		
FEMALE	1924	370,420(2%)	0.0(0.0)	100.0(0.0)	0.0
				218.2(1.0)	
RACE/ETHNICITY					
WHITE	2258	531,472(1%)	49.9(1.2)	50.1(1.2)	0.0
			230.8(1.3)	224.4(1.2)	
BLACK	682	108,783(3%)	47.8(2.1)	52.2(2.1)	0.0
			200.1(2.6)	198.9(2.3)	
HISPANIC	811	82,452(3%)	55.4(1.9)	44.6(1.9)	0.0
			204.5(2.1)	200.5(2.4)	
OTHER	199	28,956(8%)	63.4(3.1)	36.6(3.1)	0.0
			219.7(4.9)	227.2(4.8)	
PARENTAL EDUCATION					
LESS THAN H.S.	240	38,427(10%)	47.7(3.2)	52.3(3.2)	0.0
			206.7(3.1)	199.0(2.6)	
GRADUATED H.S.	531	104,142(6%)	49.8(2.5)	50.2(2.5)	0.0
			215.6(2.1)	212.6(2.0)	
SOME EDUC AFTER H.S.	314	60,114(6%)	52.9(3.2)	47.1(3.2)	0.0
			230.0(3.7)	225.8(2.4)	
GRADUATED COLLEGE	1416	285,166(4%)	53.6(1.5)	46.4(1.5)	0.0
			233.3(1.8)	229.1(1.6)	
UNKNOWN	1432	261,658(3%)	48.0(1.7)	52.0(1.7)	0.0
			213.1(1.4)	211.2(1.3)	

* CV is the coefficient of variation for the sum of the weights.

Table 16-99

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 4, by Derived Race/ethnicity, for Main U.S. History Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	3950	751,664(1%)	70.7(0.6)	14.5(0.4)	11.0(0.4)	2.0(0.2)	1.8(0.2)	0.1(0.0)	0.0
			227.5(1.0)	199.5(1.9)	202.7(1.7)	231.3(5.7)	212.7(4.7)	208.6(10.7)	
GENDER									
MALE	2026	381,243(2%)	69.6(0.9)	13.6(0.7)	12.0(0.6)	2.1(0.3)	2.7(0.5)	0.1(0.1)	0.0
			230.8(1.3)	200.1(2.6)	204.5(2.1)	226.0(7.6)	214.9(6.5)	214.5(7.3)	
FEMALE	1924	370,420(2%)	71.9(0.9)	15.3(0.7)	9.9(0.4)	2.0(0.3)	0.8(0.2)	0.1(0.1)	0.0
			224.4(1.2)	198.9(2.3)	200.5(2.4)	237.0(6.0)	205.7(8.2)	202.5(****)	
RACE/ETHNICITY									
WHITE	2258	531,472(1%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			227.5(1.0)						
BLACK	682	108,783(3%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				199.5(1.9)					
HISPANIC	811	82,452(3%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					202.7(1.7)				
OTHER	199	28,956(8%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	52.8(4.8)	45.8(5.0)	1.4(1.0)	0.0
						231.3(5.7)	212.7(4.7)	208.6(10.7)	
PARENTAL EDUCATION									
LESS THAN H.S.	240	38,427(10%)	59.0(3.3)	19.3(3.2)	17.9(2.6)	2.3(1.0)	1.6(0.8)	0.0(0.0)	0.0
			211.4(3.2)	184.4(5.0)	196.3(3.5)	197.0(9.6)	183.8(16.9)		
GRADUATED H.S.	531	104,142(6%)	70.2(2.0)	18.5(1.7)	8.2(0.9)	0.8(0.3)	2.3(0.7)	0.0(0.0)	0.0
			220.5(1.7)	197.7(2.4)	200.7(4.6)	204.6(23.3)	201.3(15.1)		
SOME EDUC AFTER H.S.	314	60,114(6%)	75.0(2.7)	11.9(1.9)	10.9(1.6)	0.4(0.3)	1.8(0.6)	0.0(0.0)	0.0
			234.9(2.4)	205.2(5.5)	204.8(6.0)	252.7(14.8)	225.2(9.0)		
GRADUATED COLLEGE	1416	285,166(4%)	73.9(1.1)	13.1(0.9)	8.8(0.7)	2.6(0.4)	1.4(0.4)	0.1(0.1)	0.0
			237.7(1.4)	206.8(2.8)	211.3(2.3)	247.9(5.7)	223.2(8.0)	208.6(10.7)	
UNKNOWN	1432	261,658(3%)	68.5(1.6)	13.9(1.3)	13.3(0.9)	2.2(0.6)	2.0(0.4)	0.0(0.0)	0.0
			218.7(1.3)	195.2(2.9)	197.9(2.3)	218.2(5.4)	210.6(6.3)		

* CV is the coefficient of variation for the sum of the weights.
(****) Standard error is greater than 99.9.

Table 16-100

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 4, by Parental Education, for Main U.S. History Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	3933	749,507(1%)	5.1(0.5)	13.9(0.8)	8.0(0.5)	38.0(1.3)	34.9(0.9)	0.3
			202.7(2.2)	214.1(1.3)	228.0(2.4)	231.4(1.5)	212.5(1.0)	
GENDER								
MALE	2017	380,328(2%)	4.8(0.5)	13.6(0.9)	8.4(0.6)	40.3(1.6)	33.0(1.3)	0.2
			206.7(3.1)	215.6(2.1)	230.0(3.7)	233.3(1.8)	213.9(1.4)	
FEMALE	1916	369,179(2%)	5.4(0.7)	14.2(1.1)	7.7(0.7)	35.9(1.6)	36.9(1.3)	0.3
			199.0(2.6)	212.6(2.0)	225.8(2.4)	229.1(1.6)	211.2(1.3)	
RACE/ETHNICITY								
WHITE	2256	530,946(1%)	4.3(0.5)	13.8(1.0)	8.5(0.6)	39.7(1.5)	33.8(1.1)	0.1
			211.4(3.2)	220.5(1.7)	234.9(2.4)	237.7(1.4)	218.7(1.3)	
BLACK	673	107,607(3%)	6.9(1.5)	17.9(1.5)	6.7(1.1)	34.7(2.5)	33.9(3.0)	1.1
			184.4(5.0)	197.7(2.4)	205.2(5.5)	206.8(2.8)	195.2(2.9)	
HISPANIC	805	81,998(3%)	8.4(1.1)	10.4(1.1)	8.0(1.1)	30.7(2.5)	42.5(2.1)	0.6
			196.3(3.5)	200.7(4.6)	204.8(6.0)	211.3(2.3)	197.9(2.3)	
OTHER	199	28,956(8%)	5.1(1.5)	11.2(3.0)	4.5(1.5)	41.1(4.8)	38.1(4.9)	0.0
			191.5(9.0)	202.2(11.7)	230.5(7.8)	238.2(5.5)	214.7(4.7)	
PARENTAL EDUCATION								
LESS THAN H.S.	240	38,427(10%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			202.7(2.2)					
GRADUATED H.S.	531	104,142(6%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				214.1(1.3)				
SOME EDUC AFTER H.S.	314	60,114(6%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					228.0(2.4)			
GRADUATED COLLEGE	1416	285,166(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
						231.4(1.5)		
UNKNOWN	1432	261,658(3%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0
							212.5(1.0)	

* CV is the coefficient of variation for the sum of the weights.

Table 16-101

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 8, by Gender, for Main U.S. History Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	6460	731,581(1%)	49.5(0.8)	50.5(0.8)	0.0
			266.2(1.0)	261.6(0.8)	
GENDER					
MALE	3156	362,158(2%)	100.0(0.0)	0.0(0.0)	0.1
			266.2(1.0)		
FEMALE	3304	369,423(2%)	0.0(0.0)	100.0(0.0)	0.0
				261.6(0.8)	
RACE/ETHNICITY					
WHITE	4006	521,546(1%)	50.1(0.9)	49.9(0.9)	0.0
			272.7(1.2)	268.1(0.9)	
BLACK	1094	106,016(3%)	44.9(1.6)	55.1(1.6)	0.1
			248.7(2.1)	243.8(1.7)	
HISPANIC	1068	77,623(3%)	53.7(1.5)	46.3(1.5)	0.0
			245.6(2.4)	242.9(2.0)	
OTHER	292	26,396(7%)	44.2(4.0)	55.8(4.0)	0.4
			264.8(4.5)	263.1(3.0)	
PARENTAL EDUCATION					
LESS THAN H.S.	587	58,381(4%)	38.2(2.1)	61.8(2.1)	0.0
			245.0(2.7)	244.8(2.1)	
GRADUATED H.S.	1701	196,123(3%)	50.0(1.4)	50.0(1.4)	0.1
			257.4(1.9)	254.7(1.2)	
SOME EDUC AFTER H.S.	1195	135,031(3%)	50.1(1.6)	49.9(1.6)	0.0
			272.6(1.9)	265.6(1.5)	
GRADUATED COLLEGE	2386	284,300(3%)	50.8(1.1)	49.2(1.1)	0.0
			276.6(1.4)	273.1(1.1)	
UNKNOWN	574	55,504(5%)	51.1(1.9)	48.9(1.9)	0.0
			247.0(2.5)	240.6(2.3)	

* CV is the coefficient of variation for the sum of the weights.

492

Table 16-102

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 8, by Derived Race/ethnicity, for Main U.S. History Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	6462	731,766(1%)	71.3(0.6)	14.5(0.4)	10.6(0.3)	2.2(0.2)	1.3(0.2)	[illegible](0.0)	0.0
			270.4(0.8)	246.0(1.5)	244.3(1.9)	273.3(2.6)	248.6(3.8)	260.3(11.4)	
GENDER									
MALE	3158	362,343(2%)	72.1(0.8)	13.2(0.6)	11.5(0.5)	2.1(0.2)	1.0(0.3)	0.1(0.1)	0.0
			272.7(1.2)	248.8(2.1)	245.6(2.4)	272.8(5.0)	250.8(8.0)	255.3(16.9)	
FEMALE	3304	369,423(2%)	70.5(0.8)	15.8(0.6)	9.7(0.4)	2.3(0.3)	1.6(0.3)	0.1(0.0)	0.0
			268.1(0.9)	243.8(1.7)	242.9(2.0)	273.8(3.4)	247.2(5.0)	266.0(13.7)	
RACE/ETHNICITY									
WHITE	4006	521,546(1%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			270.4(0.8)						
BLACK	1095	106,085(3%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				246.0(1.5)					
HISPANIC	1068	77,623(3%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					244.3(1.9)				
OTHER	293	26,513(7%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	60.6(4.7)	36.1(4.6)	3.3(1.0)	0.0
						273.3(2.6)	248.6(3.8)	260.5(11.4)	
PARENTAL EDUCATION									
LESS THAN H.S.	587	58,381(4%)	53.5(2.5)	16.9(1.8)	24.7(2.3)	1.7(0.6)	2.9(0.8)	0.4(0.2)	0.0
			249.6(2.6)	238.7(2.9)	238.4(3.5)	270.5(18.0)	233.4(13.7)	255.6(11.0)	
GRADUATED H.S.	1703	196,308(3%)	70.6(1.2)	16.7(0.9)	9.9(0.7)	0.9(0.3)	1.7(0.4)	0.2(0.1)	0.0
			261.4(1.5)	240.6(2.3)	245.0(2.5)	258.9(5.6)	248.7(6.9)	275.3(12.8)	
SOME EDUC AFTER H.S.	1195	135,631(3%)	73.1(1.4)	15.3(1.1)	8.9(0.9)	1.8(0.3)	1.0(0.3)	0.1(0.0)	0.0
			274.6(1.3)	254.0(1.9)	249.6(2.9)	277.3(7.5)	260.1(10.7)	220.5(35.2)	
GRADUATED COLLEGE	2386	284,300(3%)	79.4(1.1)	10.8(0.8)	6.1(0.6)	2.9(0.3)	0.8(0.2)	0.1(0.0)	0.0
			279.0(0.9)	255.2(2.4)	253.8(2.9)	283.2(3.2)	259.5(7.8)	281.2(5.5)	
UNKNOWN	574	55,504(5%)	47.4(2.6)	20.4(2.2)	25.8(1.9)	4.8(1.1)	1.6(0.6)	0.0(0.0)	0.0
			254.0(2.6)	230.3(2.3)	233.8(2.7)	257.4(5.7)	236.6(8.4)		

* CV is the coefficient of variation for the sum of the weights.

493

494

Table 16-103

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 8, by Parental Education, for Main U.S. History Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	6445	730,125(1%)	8.0(0.3) 244.9(1.7)	26.9(0.8) 256.1(1.2)	18.6(0.6) 269.1(1.1)	38.9(0.9) 274.9(0.9)	7.6(0.4) 243.8(1.8)	0.2
GENDER								
MALE	3145	361,170(2%)	6.2(0.4) 245.0(2.7)	27.2(1.0) 257.5(1.9)	18.8(0.8) 272.6(1.9)	40.0(1.0) 276.6(1.4)	7.9(0.5) 247.0(2.5)	0.3
FEMALE	3300	368,954(2%)	9.8(0.6) 244.8(2.1)	26.6(0.9) 254.7(1.2)	18.3(0.7) 265.6(1.5)	37.9(1.2) 273.1(1.1)	7.4(0.4) 240.6(2.3)	0.1
RACE/ETHNICITY								
WHITE	4000	520,843(1%)	6.0(0.4) 249.6(2.6)	26.6(0.9) 261.4(1.5)	19.0(0.7) 274.6(1.3)	43.3(1.2) 279.0(0.9)	5.1(0.3) 254.0(2.6)	0.1
BLACK	1090	105,565(3%)	9.4(1.1) 238.7(2.9)	31.1(1.6) 240.6(2.3)	19.6(1.4) 254.0(1.9)	29.2(1.8) 255.2(2.4)	10.7(1.2) 230.3(2.3)	0.5
HISPANIC	1066	77,489(3%)	18.6(1.8) 238.4(3.5)	25.1(1.7) 245.0(2.5)	15.5(1.7) 249.6(2.9)	22.4(2.0) 253.8(2.9)	18.5(1.6) 233.8(2.7)	0.2
OTHER	289	26,228(7%)	11.0(1.9) 247.0(10.7)	20.8(2.7) 253.8(5.2)	14.6(2.4) 270.0(5.5)	40.0(2.7) 278.2(2.8)	13.6(2.5) 252.1(5.1)	1.1
PARENTAL EDUCATION								
LESS THAN H.S.	587	58,381(4%)	100.0(0.0) 244.9(1.7)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED H.S.	1703	196,308(3%)	0.0(0.0)	100.0(0.0) 256.1(1.2)	0.0([illegible]	0.0(0.0)	0.0(0.0)	0.0
SOME EDUC AFTER H.S.	1195	135,631(3%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 269.1(1.1)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED COLLEGE	2386	284,300(3%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 274.9(0.9)	0.0(0.0)	0.0
UNKNOWN	574	55,504(5%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 243.8(1.8)	0.0

* CV is the coefficient of variation for the sum of the weights.

495

Table 16-104

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 12, by Gender, for Main U.S. History Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	5507	677,105(1%)	47.7(1.3) 298.5(1.3)	52.3(1.3) 291.8(1.1)	0.0
GENDER					
MALE	2593	322,977(3%)	100.0(0.0) 298.5(1.3)	0.0(0.0)	0.0
FEMALE	2914	354,128(3%)	0.0(0.0)	100.0(0.0) 291.8(1.1)	0.0
RACE/ETHNICITY					
WHITE	3730	500,024(1%)	48.0(1.4) 305.3(1.4)	52.0(1.4) 297.2(1.3)	0.0
BLACK	975	92,352(3%)	47.5(1.8) 276.6(2.5)	52.5(1.8) 272.4(1.9)	0.0
HISPANIC	616	56,047(5%)	42.8(2.8) 274.7(2.3)	57.2(2.8) 273.2(2.3)	0.0
OTHER	186	28,681(8%)	52.0(3.2) 291.8(7.6)	48.0(3.2) 303.4(5.7)	0.0
PARENTAL EDUCATION					
LESS THAN H.S.	518	54,278(7%)	41.7(2.5) 275.8(2.8)	58.3(2.5) 273.1(3.0)	0.0
GRADUATED H.S.	1379	166,473(3%)	47.6(2.0) 287.8(1.6)	52.4(2.0) 283.0(1.6)	0.0
SOME EDUC AFTER H.S.	1344	162,888(3%)	45.6(2.0) 300.5(1.8)	54.4(2.0) 293.7(1.4)	0.0
GRADUATED COLLEGE	2119	277,223(4%)	49.4(1.6) 310.0(1.9)	50.6(1.6) 302.1(1.7)	0.0
UNKNOWN	123	13,361(12%)	62.7(5.1) 264.4(5.5)	37.3(5.1) 258.0(5.7)	0.0

* CV is the coefficient of variation for the sum of the weights.

Table 16-105

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 12, by Derived Race/ethnicity, for Main U.S. History Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	5507	677,105(1%)	73.8(0.6)	13.6(0.4)	8.3(0.4)	3.3(0.3)	0.9(0.2)	0.1(0.1)	0.0
			301.1(1.2)	274.4(1.7)	273.9(1.8)	299.4(7.3)	293.1(6.0)	264.6(36.8)	
GENDER									
MALE	2593	322,977(3%)	74.4(0.9)	13.6(0.7)	7.4(0.6)	3.5(0.4)	0.9(0.2)	0.2(0.1)	0.0
			305.3(1.4)	276.6(2.5)	274.7(2.3)	293.3(9.4)	291.4(6.0)	264.6(36.8)	
FEMALE	2914	354,128(3%)	73.4(0.7)	13.7(0.7)	9.1(0.4)	3.0(0.3)	0.9(0.3)	0.0(0.0)	0.0
			297.2(1.3)	272.4(1.9)	273.2(2.3)	305.9([illegible]6)	294.7(9.8)		
RACE/ETHNICITY									
WHITE	3730	500,024(1%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			301.1(1.2)						
BLACK	975	92,352(3%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				274.4(1.7)					
HISPANIC	616	56,047(5%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					273.9(1.8)				
OTHER	186	28,681(8%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	77.2(4.6)	20.8(4.3)	2.1(1.5)	0.0
						299.4(7.3)	293.1(6.0)	264.6(36.8)	
PARENTAL EDUCATION									
LESS THAN H.S.	518	54,278(7%)	50.8(3.3)	16.3(2.4)	26.9(2.6)	4.1(1.8)	1.3(0.8)	0.0(0.0)	0.0
			282.5(3.7)	261.9(5.3)	265.8(3.8)	278.8(17.3)	271.6(8.6)		
GRADUATED H.S.	1379	166,473(3%)	74.0(1.2)	16.4(1.0)	6.5(0.6)	1.9(0.4)	1.1(0.3)	0.1(0.1)	0.0
			289.8(1.3)	270.2(1.9)	272.5(4.4)	286.8(7.7)	277.6(7.0)	273.0(7.7)	
SOME EDUC AFTER H.S.	1344	162,888(3%)	75.4(1.2)	15.1(1.1)	6.5(0.6)	2.2(0.6)	0.9(0.3)	0.0(0.0)	0.0
			301.0(1.3)	283.2(2.8)	281.4(3.1)	293.1(11.0)	292.3(16.2)		
GRADUATED COLLEGE	2119	277,223(4%)	79.4(1.0)	10.1(0.6)	5.3(0.5)	4.5(0.7)	0.7(0.3)	0.0(0.0)	0.0
			310.5(1.6)	277.9(2.9)	285.9(3.1)	311.8(8.2)	316.2(9.1)	286.9(****)	
UNKNOWN	123	13,361(12%)	35.4(6.0)	24.5(5.3)	33.9(5.3)	6.1(2.7)	0.0(0.0)	0.0(0.0)	0.0
			283.9(8.0)	248.7(7.0)	252.0(5.0)	244.1(11.5)			

* CV is the coefficient of variation for the sum of the weights.
(****) Standard error is greater than 99.9

497

498

Table 16-106

Weighted Response Percentages and U.S. History Proficiency Means with Standard Errors
Grade 12, by Parental Education, for Main U.S. History Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	5483	674,223(1%)	8.1(0.6) 274.2(2.0)	24.7(0.9) 285.3(1.1)	24.2(0.8) 296.8(1.1)	41.1(1.4) 306.0(1.6)	2.0(0.2) 262.0(3.9)	0.4
GENDER								
MALE	2580	321,479(3%)	7.0(0.7) 275.8(2.8)	24.7(1.3) 28[illegible]8(1.6)	23.1(1.1) 300.5(1.8)	42.6(1.8) 310.0(1.9)	2.6(0.4) 264.4(5.5)	0.5
FEMALE	2903	352,744(3%)	9.0(0.6) 273.1(3.0)	24.7(0.9) 283.0(1.6)	25.1(1.0) 293.7(1.4)	39.8(1.5) 302.1(1.7)	1.4(0.2) 258.0(5.7)	0.4
RACE/ETHNICITY								
WHITE	3717	498,532(1%)	5.5(0.6) 282.5(3.7)	24.7(1.0) 289.8(1.3)	24.6(0.9) 301.0(1.3)	44.2(1.6) 310.5(1.6)	0.9(0.2) 283.9(8.0)	0.3
BLACK	973	92,195(3%)	9.9(1.4) 261.9(5.3)	29.7(1.5) 270.2(1.9)	26.6(1.7) 283.2(2.8)	30.2(2.1) 277.9(2.9)	3.6(1.0) 248.7(7.0)	0.2
HISPANIC	608	55,167(4%)	26.5(2.6) 265.8(3.8)	19.5(1.8) 272.5(4.4)	19.2(1.7) 281.4(3.1)	26.7(2.2) 285.9(3.1)	8.2(1.5) 252.0(5.0)	1.6
OTHER	185	28,326(8%)	10.3(3.6) 277.0(11.4)	18.0(2.8) 283.0(5.2)	17.9(3.8) 292.9(10.3)	50.9(6.2) 312.3(6.9)	2.9(1.3) 244.1(11.5)	1.2
PARENTAL EDUCATION								
LESS THAN H.S.	518	54,278(7%)	100.0(0.0) 274.2(2.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED H.S.	1379	166,473(3%)	0.0(0.0)	100.0(0.0) 285.3(1.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
SOME EDUC AFTER H.S.	1344	162,888(3%)	0.0(0.0)	0.0(0.0)	100.0(0.0) 296.8(1.1)	0.0(0.0)	0.0(0.0)	0.0
GRADUATED COLLEGE	2119	277,223(4%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 306.0(1.6)	0.0(0.0)	0.0
UNKNOWN	123	13,361(12%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0) 262.0(3.9)	0.[illegible]

* CV is the coefficient of variation for the sum of the weights.

Table 16-107

Weighted Response Percentages and Geography Proficiency Means with Standard Errors Grade 12, by Gender, for Main Geography Samples

GENDER OF SUBJECT (NAEP ID: SEX)

	N	WEIGHTED N (CV)*	MALE	FEMALE	MISSING
-- TOTAL --	3030	371,737(1%)	47.8(1.3) 301.2(1.6)	52.2(1.3) 285.7(1.2)	0.0
GENDER					
MALE	1425	177,542(4%)	100.0(0.0) 301.2(1.6)	0.0(0.0)	0.0
FEMALE	1605	194,195(2%)	0.0(0.0)	100.0(0.0) 285.7(1.2)	0.0
RACE/ETHNICITY					
WHITE	2093	280,495(2%)	47.4(1.5) 309.8(1.6)	52.6(1.5) 293.3(1.2)	0.0
BLACK	535	50,197(4%)	43.6(3.0) 262.3(2.6)	56.4(3.0) 255.4(2.5)	0.0
HISPANIC	300	25,966(7%)	50.4(3.1) 279.1(5.5)	49.6(3.1) 264.4(3.7)	0.0
OTHER	102	15,079(10%)	63.1(5.3) 301.6(4.2)	36.9(5.3) 288.1(6.8)	0.0
PARENTAL EDUCATION					
LESS THAN H.S.	268	28,609(7%)	39.2(3.4) 271.8(3.4)	60.8(3.4) 263.9(2.9)	0.0
GRADUATED H.S.	750	90,037(5%)	49.3(2.2) 290.5(2.4)	50.7(2.2) 276.7(2.3)	0.0
SOME EDUC AFTER H.S.	725	84,847(4%)	46.3(1.6) 303.3(2.6)	53.7(1.6) 286.4(1.7)	0.0
GRADUATED COLLEGE	1191	157,243(3%)	48.5(2.2) 314.0(2.3)	51.5(2.2) 297.1(2.0)	0.0
UNKNOWN	82	9,552(12%)	58.3(5.4) 258.4(7.1)	41.7(5.4) 250.9(9.7)	0.0

* CV is the coefficient of variation for the sum of the weights.

500

Table 16-108

Weighted Response Percentages and Geography Proficiency Means with Standard Errors Grade 12, by Derived Race/ethnicity, for Main Geography Samples

DERIVED RACE/ETHNICITY (NAEP ID: DRACE)

	N	WEIGHTED N (CV)*	WHITE	BLACK	HISPANIC	ASIAN	AMER IND	UNCLASSIFIED	MISSING
-- TOTAL --	3030	371,737(1%)	75.5(0.8)	13.5(0.5)	7.0(0.5)	2.7(0.3)	1.1(0.3)	0.3(0.1)	0.0
			301.1(1.1)	258.4(2.0)	271.8(3.9)	298.7(4.3)	292.5(4.8)	291.0(22.6)	
GENDER									
MALE	1425	177,542(4%)	74.9(1.3)	12.3(1.0)	7.4(0.7)	3.3(0.6)	1.6(0.4)	0.5(0.3)	0.0
			309.8(1.6)	262.3(2.6)	279.1(5.5)	304.2(4.4)	297.5(6.1)	295.7(30.0)	
FEMALE	1605	194,195(2%)	75.9(1.0)	14.6(0.6)	6.6(0.7)	2.2(0.4)	0.6(0.2)	0.1(0.1)	0.0
			293.3(1.2)	255.4(2.5)	264.4(3.7)	290.9(8.1)	280.5(8.2)	267.8(****)	
RACE/ETHNICITY									
WHITE	2093	280,495(2%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			301.1(1.1)						
BLACK	535	50,197(4%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				258.4(2.0)					
HISPANIC	300	25,966(7%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					271.8(3.9)				
OTHER	102	15,079(10%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	67.5(6.4)	26.0(5.9)	6.5(3.5)	0.0
						298.7(4.3)	292.5(4.8)	291.0(22.6)	
PARENTAL EDUCATION									
LESS THAN H.S.	268	28,609(7%)	51.0(3.3)	20.4(2.5)	22.0(2.4)	3.1(1.4)	2.9(1.[illegible])	0.0(0.0)	0.0
			276.1(3.1)	248.0(5.6)	258.8(6.2)	278.1(8.4)	289.2(11.9)		
GRADUATED H.S.	750	90,037(5%)	74.3(1.8)	17.3(1.3)	5.5(0.9)	1.7(0.4)	0.9(0.5)	0.2(0.2)	0.0
			291.1(1.7)	254.5(4.8)	270.7(3.7)	285.5(6.9)	290.6(13.7)	284.7(10.1)	
SOME EDUC AFTER H.S.	725	84,847(4%)	75.8(1.5)	14.9(1.2)	6.8(1.1)	1.9(0.8)	0.6(0.4)	0.0(0.0)	0.0
			301.2(1.6)	263.1(3.4)	279.5(7.3)	310.1(7.1)	314.4(17.7)		
GRADUATED COLLEGE	1181	157,243(3%)	82.2(1.2)	8.4(0.7)	4.5(0.5)	3.5(0.5)	1.1(0.4)	0.3(0.2)	0.0
			310.5(1.5)	267.0(4.1)	281.9(5.5)	309.0(5.4)	288.9(6.6)	324.0(41.7)	
UNKNOWN	82	9,552(12%)	46.6(6.5)	26.2(5.4)	16.0(3.7)	7.5(2.4)	0.0(0.0)	3.7(3.7)	0.0
			264.5(10.9)	241.2(8.0)	255.1(8.1)	248.6(20.5)		251.9(****)	

* CV is the coefficient of variation for the sum of the weights.
(****) Standard error is greater than 99.9.

501

502

Table 16-109

Weighted Response Percentages and Geography Proficiency Means with Standard Errors Grade 12, by Parental Education, for Main Geography Samples

PARENTS' EDUCATION (NAEP ID: PARED)

	N	WEIGHTED N (CV)*	NOT HS	GRAD HS	POST HS	GRAD COL	UNKNOWN	MISSING
-- TOTAL --	3016	370,289(1%)	7.7(0.6)	24.3(1.1)	22.9(0.9)	42.5(1.2)	2.6(0.3)	0.4
			267.0(2.4)	283.5(1.6)	294.2(1.6)	305.3(1.5)	255.2(5.6)	
GENDER								
MALE	1417	176,688(4%)	6.4(0.8)	25.1(1.4)	22.2(1.2)	43.1(1.6)	3.2(0.5)	0.5
			271.8(3.4)	290.5(2.4)	303.3(2.6)	314.0(2.3)	258.4(7.1)	
FEMALE	1599	193,601(2%)	9.0(0.7)	23.6(1.4)	23.5(1.1)	41.8(1.8)	2.1(0.3)	0.3
			263.9(2.9)	276.7(2.3)	286.4(1.7)	297.1(2.0)	250.9(9.7)	
RACE/ETHNICITY								
WHITE	2087	279,707(2%)	5.3(0.6)	23.9(1.4)	23.0(1.0)	46.2(1.5)	1.6(0.3)	0.3
			276.1(3.1)	291.1(1.7)	301.2(1.6)	310.5(1.5)	264.5(10.9)	
BLACK	531	49,882(4%)	11.7(1.5)	31.3(2.1)	25.4(2.2)	26.6(2.0)	5.0(1.2)	0.6
			248.0(5.6)	254.5(4.8)	263.1(3.4)	267.0(4.1)	241.2(8.0)	
HISPANIC	296	25,622(7%)	24.5(2.7)	19.3(2.5)	22.5(2.7)	27.6(2.7)	6.0(1.1)	1.3
			258.8(6.2)	270.7(3.7)	279.5(7.3)	281.9(5.5)	255.1(8.1)	
OTHER	102	15,079(10%)	11.2(3.2)	16.9(4.0)	13.7(4.0)	51.2(4.8)	7.1(2.7)	0.0
			283.5(7.3)	287.1(6.0)	311.1(5.8)	305.2(4.9)	249.7(15.5)	
PARENTAL EDUCATION								
LESS THAN H.S.	268	28,609(7%)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
			267.0(2.4)					
GRADUATED H.S.	750	90,037(5%)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
				283.5(1.8)				
SOME EDUC AFTER H.S.	725	84,847(4%)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0(0.0)	0.0
					294.2(1.6)			
GRADUATED COLLEGE	1191	157,243(3%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0(0.0)	0.0
						305.3(1.5)		
UNKNOWN	82	9,552(12%)	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	100.0(0.0)	0.0
							255.2(5.6)	

* CV is the coefficient of variation for the sum of the weights.

503

APPENDIX A

Consultants for the Development of 1988 NAEP Objectives and Items

Appendix A

CONSULTANTS FOR THE DEVELOPMENT OF 1988 NAEP OBJECTIVES AND ITEMS

READING

Learning Area Committee

Peter Mosenthal, Syracuse University, Syracuse, NY
Richard Venezky, University of Delaware, Newark, DE

Development Consultants and Reviewers

Jo Beth Allen, Kansas State University, Manhattan, KS
Arthur Applebee, National Council of Teachers of English, Urbana, IL
Michael Axline, University of Oregon, Eugene, OR
Fernie Baca, University of Colorado, Denver, CO
Richard Beach, University of Minnesota, Minneapolis, MN
Barbara Bianchi, Paideia School, Atlanta, GA
Susan Blank, Manpower Demonstration Research Corporation, Brooklyn, NY
Robin Butterfield, Northwest Regional Educational Laboratory, Portland, OR
Robert Calfee, Stanford University, Stanford, CA
Jeanne Chall, Harvard Graduate School of Education, Cambridge, MA
Carita Chapman, Chicago Public Schools, Chicago, IL
Ruth Coleman, North Side High School Mothers Alumni Club, Fort Wayne, IN
Christopher Connell, Associated Press, Washington, DC
Larry Coon, McDonald's Restaurant, Henderson, TX
Bernice Cullinan, New York University, New York, NY
Mary E. Curtis, Harvard Graduate School of Education, Cambridge, MA
Jacqueline Danzberger, Youthwork Inc., Arlington, VA
Martha Darling, Washington Roundtable Education Study, Bellevue, WA
Philip DiStefano, University of Colorado, Boulder, CO
Terry Dozier, South Carolina State Department of Education, Columbia, SC
Priscilla Drum. University of California at Santa Barbara, Santa Barbara, CA
William Eller, State University of New York at Buffalo, Amherst, NY
Leo Estrada, University of California, Los Angeles, CA
Claryce Evans, Boston Public Schools, Boston, MA
Marjorie Farmer, School District of Philadelphia, Philadelphia, PA
Roger Farr, University of Indiana, Bloomington, IN
Edmund Farrell, University of Texas, Austin, TX
Edward Fry, Rutgers University, New Brunswick, NJ
Carol Gibson, National Urban League, New York, NY
Kenneth Goodman, University of Arizona, Tucson, AZ
Naomi Gordon, Public Schools of Brookline, Lexington, MA
Donald Graves, University of New Hampshire, Durham, NH
Jean Greenlaw, North Texas State University, Denton, TX
Doris Hankins, Germantown High School, Germantown, TN
Jerome Harste, University of Indiana, Bloomington, IN

David Hayes, University of Georgia, Athens, GA
Paul Heffernan, Star Market, Newtonville, MA
Harold Herber, Syracuse University, Syracuse, NY
Jane Holt, Champlain Valley Union High School, Hinesburg, VT
Shu-in Huang, City of Thornton, Thornton, CO
Evaline Khayat Kruse, Audubon Junion High School, Los Angeles, CA
Judith Langer, University of California, Berkeley, CA
Diane Lapp, Boston University, Boston, MA
Herbert J. Lapp, Jr., GPU Nuclear Corporation, Parsippany, NJ
Ron Lessnau, Hamburger University, Oakbrook, IL
Ray Marshall, University of Texas, Austin, TX
Phyllis A. Miller, Reading Development Seminars, Minneapolis, MN
Charles Moody, University of Michigan, National Alliance of Black School Educators, Ann Arbor, MI
Edwin Newman, NBC News, New York, NY
Pedro Pedraza, Jr., Hunter College, New York, NY
Anthony Petrosky, University of Pittsburgh, Pittsburgh, PA
Carolyn Pollan, State of Arkansas, Fort Smith, AR
Walter L. Powers, School District #271, Coeur d'Alene, ID
John Readance, Louisiana State University, Baton Rouge, LA
Beverly Roller, Jefferson County Public Schools, Littleton, CO
Glenn E. Rotz, Highland Elementary School, Clarkson, WA
Sarah Saint-Onge, Godine Publishing Co., Boston, MA
Adan C. Salgado, Johnston High School, Austin, TX
S. Jay Samuels, University of Minnesota, Minneapolis, MN
Robert Schreiner, University of Minnesota, Minneapolis, MN
John Stewig, University of Wisconsin, Milwaukee, WI
Dorothy Strickland, Columbia University, New York, NY
Robert Tierney, University of Illinois, Champagne-Urbana, IL
Jaap Tuinman, Simon Fraser University, Burnaby, B.C., Canada
Janet Tully, Marriott Corporationnn, Washington, DC
Richard Vacca, Kent State University, Kent, OH
Rod Vahl, Central High School, Davenport, IA
Sheila Valencia, University of Colorado, Boulder, CO
Thomas Vallejos, University of Colorado, Boulder, CO
Maria Watkins, University of Pennsylvania, Graduate School of Education, Philadelphia, Pennsylvania
Rick Wetherell, North Bend High School, West Bend, OR
Susan M. Wolf, The Hastings Center, Hastings-on-Hudson, NY
Kathy Yen, San Francisco Public Schools, San Francisco, CA
Seymour Yesner, Brookline High School, Brookline, MA

WRITING

Learning Area Committee

Arthur Applebee, Stanford University, Stanford, CA
Mary Barr, California State Department of Education, San Diego, CA
David Bartholomae, University of Pittsburgh, Pittsburgh, PA
Miriam Chaplin,

Paul Diehl, University of Iowa, Iowa City, IA
Ed Folsom, University of Iowa, Iowa City, IA
Helen Kelley, State Department of Education, Commonwealth of Virginia, Richmond, VA

Development Consultants and Reviewers

Robert Alligood, Plymouth Public Schools, Plymouth, NC
Lee Antell, Champlin, MN
Elsa Bartlett, New York University Medical Center, New York, NY
Opaline Brice, California PTA, Inglewood, CA
Bill Burns, Boulder High School, Boulder, CO
Mary Busch, Indiana Central College, Indianapolis, IN
Robin Butterfield, N.W. Indian Reading Program, Portland, OR
Shirley Cain, Golden, CO
Rosa Casarez, Downey, CA
Patricia Candal, University of South Carolina, Columbia, SC
Courtney Cazden, Harvard University, Cambridge, MA
Gilbert Cho, Cambridge, MA
Jane Christensen, National Council of Teachers of English Urbana, IL
Nancy Ciarleglio, New Haven PTA, New Haven, CT
Edmond Cody, North Side Independent School District, San Antonio, TX
Charles Co[illegible], University of San Diego-La Jolla, La Jolla, CA
Florence Cox, Chicago Region PTA, Chicago, IL
Alonzo Crim, Atlanta City Schools, Atlanta, GA
John Daly, University of Texas at Austin, Austin, TX
Leonard Da[illegible]con, Washoe City School District, Reno, NV
Vivian Davis, Tri-Ethnic Committee, Dallas, TX
John Dawkins, Yardley, PA
Margaret Dinn, Dade County Elementary & Secondary Schools, Miami, FL
Joanne Eresh, West Liberty Training Center, Pittsburgh, PA
Marjorie Farmer, School District of Philadelphia, Philadelphia, PA
Marcia Farr, University of Illinois at Chicago, Chicago, IL
Paul Fawson, Weber County School District, Ogden, UT
Eleanor Francke, Lincoln Public Schools, Lincoln, NE
June Gabler, Woodhaven School District, Romulus, MI
Darrell Garber, New Orleans Public Schols, New Orleans, LA
Pete Garcia, Espanola School District, Espanola, NM
Perry Gilmore, University of Alaska at Fairbanks, Fairbanks, AK
Gene Goff Jr., West Virginia Congress of Parents and Teachers, Inc., Huntington, WV
Donald Graves, University of New Hampshire, Durham, NH
Karen Greenberg The City University of New York, New York, NY
Carol Ann Greenhalgh, Texas Education Agency, Austin, TX
Robert Gundlach, Northwestern University, Evanston, IL
Kris Gutierrez, University of Colorado, Boulder, CO
Carol Harner, University of Denver, Denver, CO
Barbara Hasek, Oregon State PTA, Springfield, OR
Mary Heim, Grant Elementary School, Casper, WY
Diane Hernandez, Lafayette Elementary School, Lafayette, CO
Wayne Hill, Board of Education, Cumberland, MD

Susan Hogsgard, Pittsburgh Board of Education, Pittsburgh, PA
Jack Holmquist, Nebraska PTA, York, NE
Shu-In Huang, City of Thornton, Thornton, CO
Ann Humes, Southwest Regional Laboratory, Las Alamitos, CA
Enid Humphrey, West Side Community Schools, Omaha, NE
Richard Johnson, Center for New Schools, Highland Park, IL
Donald Jones, Jefferson County Public Schools, Lakewood, CO
Kenneth Kantor, University of Georgia, Athens, GA
Jaima Killian, Wheat Ridge, CO
James Rex Kirk Sr., Utah PTA, Tooele, UT
Carl Klaus, University of Iowa, Iowa City, IA
Rita Klemm, Wheat Ridge Senior High School, Wheat Ridge, CO
Shari Lahr, Buckeye Central School District, New Washington, OH
Judith Langer, University of California at Berkeley, Berkeley, CA
Glenda Liston, Texas Congress of Parents and Teacheers, Marshall, TX
Wendy Littlefair, Measurement, Inc., Durham, NC
Richard Lloyd-Jones, University of Iowa, Iowa City, IA
Fay Loo, Seward Park High School, New York, NY
Dorothy Magett, Seattle Public Schools, Seattle, WA
Betty Mangum, State Department of Public Instruction, Raleigh, NC
Maureen McCarthy, Totten Intermediate School, Staten Island, NY
George McCulley, Michigan Tech University, Houghton, MI
Carol Mathews, Boulder High School, Boulder, CO
Frances McCormick, Manning Junior High School, Golden, CO
Sister Jarlath McManus, Archdiocese of Denver, Denver, CO
Mary Meier, Eugene School District, Eugene, OR
John Mellon, University of Illinois, Evanston, IL
Patti Mendes, University of Colorado, Boulder, CO
Vana Meredith, State Department of Education, Columbia SC
Donna Miller, Orange County Public Schools, Orlando,
Sandra Murphy, University of California at Berkeley, Berkeley, CA
Jeff Oliver, Lincoln Elementary School, Boulder, CO
Robert Palmatier, Birmingham Board of Education, Birmingham, AL
Jan Patton, Findlay City Schools, Findlay, OH
Jesse Perry, San Diego Public Schools, San Diego, CA
Anthony Petrosky, University of Pittsburgh, Pittsburgh, PA
Lorraine Plasse, Springfield Public Schools, Springfield, MA
Sanford Powell, Northwest Mississippi Teacher Center, Senatobia, NY
Edys Quellmaiz, University of California at Los Angeles, Los Angeles, CA
Edward Reidy, West Hartford Public Schools, West Hartford, CT
Edward Roeber, Michigan Department of Education, Lansing, MI
Marilyn Rosenblat, College Learning Laboratory, Buffalo, NY
Thomas Roy, Ithaca City Schools, Ithaca, NY
Charles Schuster, University of Wisconsin, Milwaukee, WI
Evelyn Scott, Englewood, CO
Sandra Seale, Cherry Creek High School, Aurora, CO
Sharif Shakrani, Michigan Department of Education, Lansing, MI
Mary Ann Shea, University of Colorado, Boulder, CO
Yvonne Siu-Runyan, Boulder Valley Schools, Pasadena, CA
Robert Smith, Sandusky City Schols, Sandusky, OH
Geneva Smitherman, Wayne State University, Detroit, MI
Susan Sowers, Cambridge, MA

Mary Stitt, Jefferson County Public Schools, Lakewood, CO
Barbara Thompson, Wisconsin Department of Public Instruction, Madison, WI
Donna Townsend, Texas Education Agency, Austin, TX
Lynn Troyka, CUNY, Queensborough Community College, Bayside, NY
Tomas Vallejos, University of Minnesota, Minneapolis, MN
Faith Waters, Bucks County School District, Doylestown, PA
Maria Watkins, Philadelphia, PA
Richard Weaver, Bentonville Schools, Bentonville, AR
Ben Williams, Chicago Board of Education, Chicago, IL
Darnell Williams, Bishop College, Dallas, TX
John Wood, Juchem Elementary School, Broomfield, CO
Seymour Yesner, Brookline Education Center, Brookline, MA

GEOGRAPHY

Learning Area Committee

Russ Allen, Wisconsin Department of Public Instruction, Madison, WI
Norman Bettis, Illinois State University, Normal, IL
Marianne Kenney, Cherry Creek Schools, Denver, CO
Dana Kurfman, Prince Georges County Schools, Landover, MD
Christopher Salter, National Geographic Society, Washington, DC

Development Consultants and Reviewers

Alan Backler, Indiana University, Bloomington, IN
Phillip Bacon, University of Houston, Houston, TX
Richard Boehm, Texas State University, San Marcos. TX
Nancy Brown, Mississippi Department of Education, Jackson, MS
Gary Elbow, Texas Tech University, Lubbock, TX
Charles Gritzner, South Dakota State University, Brookings, SD
Michael Hartoonian, Wisconsin Department of Public Instruction, Madison, WI
Richard Helbock, La Posta: Journal of American Postal History, West Linn, OR
Gail Hobbs, University of California, Los Angeles, CA
Briavel Holcomb, Rutgers University, New Brunswick, NJ
Oscar Horst, Western Michigan University, Kalamazoo, MI
Mary Jean Katz, Portland, OR
Gail Ludwig, University of Missouri, Columbia, MO
James Marran, New Trier Township High School, Winnetka, IL
Gary Miller, F.W. Cox High School, Virginia Beach, VA
William Miller, Louisiana Department of Education, Baton Rouge, LA
Janice Monk, University of Arizona, Tucson, AZ
Salvatore Natoli, National Council for the Social Studies, Washington, DC
Fleta Nockels, Dogulas County Schools, Castle Rock, CO
Douglas Phillips, Anchorage School District, Anchorage, AK
Sandra Pritchard, West Chester University, West Chester, PA
William R. Strong, University of North Alabama, Florence, AL
Doug Wilms, East Carolina University, Greenville, NC
Barbara Winston, Northeastern Illinois University, Chicago, IL

U.S. HISTORY

Learning Area Committee

Mary Graham, Metropolitan Public Schools, Nashville, TN
Joel Grossman, University of Wisconsin, Madison, WI
David Hammack, Case Western Reserve University, Cleveland, OH
Peter Kneedler, California Department of Education, Sacramento, CA
Sheila Mann, American Political Science Associates, Washington, DC
Sam Natoli, National Council for the Social Studies, Washington, DC
Harry Scheiber, Chair, University of California, Berkeley, CA
Betty Waugh, West Mesa High School, Albuquerque, NM

State Advisory Committee

Russ Allen, Wisconsin Department of Public Instruction, Madison, WI
Helen Brown, Bureau of Curriculum, In-Service & State Dept., Baton Rouge, LA
John Chapman, Michigan State Department of Education Lansing, MI
Roger Hammer, Wyoming Department of Education, Cheyenne, WY
Peter Kneedler, California Department of Education, Sacramento, CA
Douglas Robertson, State Department of Public Adminstration, Raleigh, NC
Lois Rubin, Virginia Department of Education, Richmond, VA

Development Consultants and Reviewers

Ron Adams, Kingswood High School, Wolfeboro, NH
James F. Adomanis, Annapolis, MD
Susan Austin, Bala Cynwyd, PA
Marvin Awbrey, Fresno Unified School District, Fresno, CA
Rosemary Bane, Robertson Academy, Nashville, TN
J. Sherman Barker, The Hotchkiss School, Lakeville, CT
Jim Bell, Poway High School, Poway, CA
Marjorie Bingham, St. Louis Park Schools, St. Louis Park, MN
Jeanine Blumberg, SUNY at Stony Brook, Stony Brook, NY
Mabel McKinney Bowning, YEFC, American Bar Assn., Chicago, IL
Judy Bristol, Houston Independent School District, Houston, TX
Nancy Brown, Mississippi Board of Education, Jackson, MS
George Burson, Aspen High School, Aspen, CO
Robert Campbell, Wesleyan Church, Lambertville, NJ
Mary Ann Cardia, The Philadelphia School, Philadelphia, PA
Carol Chaet, Cambridge Public Schools, Cambridge, MA
Phyllis Clarke, Boulder Valley Schools, Boulder, CO
Harriet Crane, The Baldwin School, Bryn Mawr, PA
Gerald Danzer, University of Chicago, Chicago, IL
Robert M. Dawson, Albuquerque, NM
Charlotte de Costa, The Baldwin School, Bryn Mawr, PA
Cooper Delk, McMurray Middle School, Nashville, TN

Deborah Drucker, Springbrook High School, Chevy Chase, MD
Terrie Epstein, Cherry Creek High School, Englewood, CO
William Fernekes, Flemington, NJ
George Flittie, Portland, OR
Mary Giunta (ex officio), NARA/NHPRC, Washington, DC
C. Dale Greewald, Learning Improvement Services, Nederland, CO
David Harris, Oakland County Public Schools, Pontiac, MI
JoAnn Heidenrich, Hillwood High School, Nashville, TN
Sylvia Hoffert, St. Louis Country Day School, St. Louis, MO
Reverend Giles P. Hayes, O.S.B., The Delbarton School, Morristown, NJ
Linda Glickstein, Philadelphia, PA
David Glickstein, Philadelphia, PA
Paul Dennis Hoffm[illegible] Canyon del Oro High School, Tucson, AZ
Donna Hudson, Ol[illegible]ter Elementary School, Antioch, TN
Alison Johnson, Glencliff High School, Nashville, TN
Savannah Jones, Birmingham Public Schools, Birmingham, AL
Sylvia Karnowsky, Overton High School, Nashville, TN
Joseph Kovacs, Edison School District, Edison, NJ
David Laudenschlager, Rapid City Central High School, Rapid City, SD
James Loguidice, Bucks County Intermediate Unit, Doylestown, PA
Thomas Lyons, Phillips Academy, Andover, MA
Anne McNair, Kirkpatrick Elementary School, Nashville, TN
Mary McFarland, Parkway School District, Chesterfield, MO
David Meisrow, Portland, OR
Fay Metcalf, National Council of the Social Studies, Washington, DC
S. Rex Morrow, Pennsylvania State University, University Park, PA
Carla Nankervis, Meigs Magnet School, Nashville, TN
Louise Osborne, Fall-Hamilton Elementary School, Nashville, TN
Arthur Pease, Lebanon High School, Lebanon, NH
Frederick W. Pfister, Bloomfield Hills, MI
Douglas Phillips, Anchorage School District, Anchorage, AK
John Phillips, Sacramento, CA
Catherine Pickle, Memphis City Schools, Memphis, TN
Gene Pickel, Oak Ridge School, Oak Ridge, TN
Helen Richardson, Fulton County Board of Education, Atlanta, GA
Susan Roberts, Albuquerque, NM
Lucille Robinson, Ontario-Montclair, Ontario, CA
Al Rocca, Sequoia Junior High School, Redding, CA
John Rossi, El Cerrito Senior High, El Cerrito, CA
Jennie Scott, Moore Middle School, Nashville, TN
Denny Shillings, Homewood-Flossmore Community High School, Flossmore, IL
Larry Strickland, Public Instruction Office, Tumwater, WA
Ruth Sykes, Haywood Elementary School, Nashville, TN
Elaine M. Takenaka, Department of Education, Honolulu, HI
Jan Talbot, Fair Oaks, CA
Steven Teal, Hercules, CA
Nan Teeter, Percy Priest Elementary School, Nashville, TN
Paul Tice, McMurray Middle School, Nashville, TN
Mary Jane Turner, Center for Civic Education, Calabasas, CA
Jane Vandercook, Hunters Lane Comprehensive High School, Hendersonville, TN
Deborah Welch, AHA, Washington, DC
Mary Lou Williams, Santa Fe, NM

Thomasine Wilson, Berkeley, CA
Virginia Wilson, North Carolina School of Science and Mathematics, Durham, NC
Celeste Woodley, Boulder Valley Schools, Boulder, CO
Brain R. Wright, Neel's Bend Middle School, Nashville, TN

CIVICS

Learning Area Committee

Russ Allen, Wisconsin Department of Public Instruction, Madison, WI
Mary Graham, Metropolitan Public Schools, Nashville, TN
Joel Grossman, University of Wisconsin, Madison, WI
David Hammack, Case Western Reserve University, Cleveland, OH
Peter Kneedler, California Department of Education, Sacramento, CA
Sheila Mann, American Political Science Associates, Washington, DC
Sam Natoli, National Council for the Social Studies, Washington, DC
Harry Scheiber, Chair, University of California, Berkeley, CA
Betty Waugh, West Mesa High School, Albuquerque, NM

State Advisory Committee

Russ Allen, Wisconsin Department of Public Instruction, Madison, WI
Helen Brown, Bureau of Curriculum, In-Service & State D ,c., Baton Rouge, LA
John Chapman, Michigan State Department of Education, Lansing, MI
Roger Hammer, Wyoming Department of Education, Cheyenne, WY
Peter Kneedler, California Department of Education, Sacramento, CA
Douglas Robertson, State Department of Public Admirstration, Raleigh, NC
Lois Rubin, Virginia Department of Education, Richmond, VA

Development Consultants and Reviewers

Ron Adams, Kingswood High School, Wolfeboro, NH
James F. Adomanis, Annapolis, MD
Susan Austin, Bala Cynwyd, PA
Marvin Awbrey, Fresno Unified School District, Fresno, CA
Rosemary Banc, Robertson Academy, Nashville, TN
J. Sherman Barker, The Hotchkiss School Lakeville, CT
Jim Bell, Poway High School, Poway, CA
Marjorie Bingham, St. Louis Park Schools, St. Louis Park, MN
Jeanine Blumberg, SUNY at Stony Brook, Stony Brook, NY
Mabel McKinney Bowning, YEFC, American Bar Assn., Chicago, IL
Judy Bristol, Houston Independent School District, Houston, TX
Nancy Brown, Mississippi Board of Education, Jackson, MS
George Burson, Aspen High School, Aspen, CO
Robert Campbell, Wesleyan Church, Lambertville, NJ
Mary Ann Cardia, The Philadelphia School, Philadelphia, PA
Carol Chaet, Cambridge Public Schools, Cambridge, MA
Phyllis Clarke, Boulder Valley Schools, Boulder, CO

Harriet Crane, The Baldwin School, Bryn Mawr, PA
Gerald Danzer, University of Chicago, Chicago, IL
Robert M. Dawson, Albuquerque, NM
Charlotte de Costa, The Baldwin School, Bryn Mawr, PA
Cooper Delk, McMurray Middle School, Nashville, TN
Deborah Drucker, Springbrook High School, Chevy Chase, MD
Terrie Epstein, Cherry Creek High School, Englewood, CO
William Fernekes, Flemington, NJ
George Flittie, Portland, OR
Mary Giunta (ex officio), NARA/NHPRC, Washington, DC
C. Dale Greenwald, Learning Improvement Services, Nederland, CO
David Harris, Oakland County Public Schools, Pontiac, MI
JoAnn Heidenreich, Hillwood High School, Nashville, TN
Sylvia Hoffert, St. Louis Country Day School, St. Louis, MO
Reverend Giles P. Hayes, O.S.B., The Delbarton School, Morristown, NJ
Linda Glickstein, Philadelphia, PA
David Glickstein, Philadelphia, PA
Paul Dennis Hoffman, Canyon del Oro High School, Tucson, AZ
Donna Hudson, Old Center Elementary School, Antioch, TN
Alison Johnson, Glencliff High School, Nashville, TN
Savannah Jones, Birmingham Public Schools, Birmingham, AL
Sylvia Karnowsky, Overton High School, Nashville, TN
Joseph Kovacs, Edison School District Edison, NJ
David Laudenschlager, Rapid City Central High School, Rapid City, SD
James Loguidice, Bucks County Intermediate Unit, Doylestown, PA
Thomas Lyons, Phillips Academy, Andover, MA
Ann McNair, Kirkpatrick Elementary School, Nashville, TN
Mary McFarland, Parkway School District, Chesterfield, MO
David Meisrow, Portland, OR
Fay Metcalf, National Council of the Social Studies, Washington, DC
S. Rex Morrow, Pennsylvania State University, University Park, PA
Carla Nankervis, Meigs Magnet School, Nashville, TN
Louis Osborne, Fall-Hamilton Elementary School Nashville, TN
Arthur Pease, Lebanon High School, Lebanon, NH
Frederick W. Pfister, Bloomfield Hills, MI
Douglas Phillips, Anchorage School District, Anchorage, AK
John Phillips, California Department of Education, Sacramento, CA
Catherine Pickle, Memphis City Schools, Memphis, TN
Gene Pickel, Oak Ridge School, Oak Ridge, TN
Helen Richardson, Fulton County Board of Education, Atlanta, GA
Susan Roberts, Albuquerque, NM
Lucille Robinson, Ontario-Montclair, Ontario, CA
Al Rocca, Sequioa Junior High School, Redding CA
John Rossi, El Cerrito Senior High, El Cerrito, CA
Jennie Scott, Moore Middle School, Nashville, TN
Denny Shillings, Homewood-Flossmore Community High School, Flossmore, IL
Larry Strickland, Public Instruction Office, Tumwater, WA
Ruth Sykes, Haywood Elementary School, Nashville, TN
Elaine M. Takenaka, Department of Education, Honolulu, HI
Jan Talbot, Fair Oaks, CA
Steven Teal, Hercules, CA
Nan Teeter, Percy Priest Elementary School, Nashville, TN

Paul Tice, McMurray Middle School, Nashville, TN
Mary Jane Turner, Center for Civic Education, Calabasas, CA
Jane Vandercook, Hunters Lane Comprehensive High School, Hendersonville, TN
Deborah Welch, AHA, Washington. DC
Joy Weldon, Bethlehem, PA
Mary Lou Williams, Santa Fe, NM
Thomasine Wilson, Berkeley, CA
Virginia Wilson, NC School of Science and Mathematics, Durham, NC
Celeste Woodley, Boul · Valley Schools, Bc er, CO
Brian R. Wright, Neely's Bend Middle School, Nashville, TN

BACKGROUND QUESTIONS CONSULTANTS

Valeria Ford, Office of Research and Student Assessment, DC Public Schools
Susan Fuhrman, Eagleton Institute for Politics, Rutgers University
Barbara Heyns, Center for Applied Social Science Research, . York University
Nancy Karweit, Center for the Social Organization of Schools, Hopkins University
Lorraine McDonnell, Rand Corporation
Vilma Ortiz, Manpower Demonstration Research Corporation
Michael Smith, School of Education, Stanford University
Herb Walberg, University of Illinois at Chicago Circle

APPENDIX B

Distribution of Weight Components for 1988 NAEP Samples

Appendix B

DISTRIBUTION OF WEIGHT COMPONENTS FOR 1988 NAEP SAMPLES

The following 11 tables, which are cited throughout Chapter 8, "Weighting Procedures and Estimation of Sampling Variance," show the distribution of weight components for the 1988 NAEP samples, including base weights, the various nonresponse adjustment factors, trimming factors, and poststratification factors, for the student weights, the excluded student weights and the teacher-student weights. A description of each table is given in Chapter 8, sections 8.1.1 through 8.1.6.

Table B-1

Distribution of Student Base Weights - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	20312	157.1	63.5	56.8	107.7	123.1	215.7	860.9
Grade 8/Age 13 Main	36699*	88.6	45.4	21.2	61.0	77.6	120.1	526.0
Grade 12/Age 17 Main	32710	72.9	35.9	20.7	44.9	56.1	100.4	402.7
Age 9 Bridge - Books 51-56	5188	728.5	428.7	213.1	396.2	617.6	980.3	4787.3
Age 9 Bridge - Book 91	1274	2159.2	1259.0	364.0	1121.5	1833.9	2971.5	8187.1
Age 9 Bridge - Book 92	1240	2145.0	1272.7	364.0	1145.1	1852.9	2940.9	8187.1
Age 9 Bridge - Book 93	1197	2277.7	1593.7	639.3	1149.3	1873.6	2958.9	13169.7
Age 13 Bridge - Books 51-56	5500	646.7	245.7	180.1	480.4	647.1	740.3	3867.7
Age 13 Bridge - Book 90	1938	1288.5	444.0	360.2	984.5	1281.2	1558.0	3970.0
Age 13 Bridge - Book 91	1405	1924.4	630.0	853.5	1450.3	1921.8	2220.8	3593.9
Age 13 Bridge - Book 92	1281	2057.7	1031.0	600.3	1483.5	1906.1	2496.2	11603.0
Age 13 Bridge - Books 93	1256	1904.7	810.9	540.3	1488.8	1921.8	2249.5	12149.2
Age 17 Bridge - Books 51-56	4622	492.0	248.7	166.1	306.5	405.0	609.7	1552.4
Age 17 Bridge - Books 61 - 67	7052	335.9	169.5	95.1	203.8	281.2	447.6	1075.3
Age 17 Bridge - Book 90	1786	1017.4	484.3	332.3	633.7	854.3	1300.9	2328.5

* Number includes 5 098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences. An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

Table B-2

Distribution of School Nonresponse Adjustments - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	20312	1.108	0.289	0.796	1.000	1.000	1.000	3.174
Grade 8/Age 13 Main	36699*	1.150	0.318	0.836	1.000	1.000	1.107	2.847
Grade 12/Age 17 Main	32710	1.222	0.450	0.872	1.000	1.000	1.260	3.045
Age 9 Bridge - Books 51-56	5188	1.055	0.172	1.000	1.000	1.000	1.000	2.000
Age 9 Bridge - Book 91	1274	1.047	0.152	1.000	1.000	1.000	1.000	2.007
Age 9 Bridge - Book 92	1240	1.055	0.167	1.000	1.000	1.000	1.000	2.007
Age 9 Bridge - Book 93	1197	1.057	0.174	1.000	1.000	1.000	1.000	2.007
Age 13 Bridge - Books 51-56	5500	1.002	0.097	0.760	1.000	1.000	1.000	1.890
Age 13 Bridge - Book 90	1938	1.008	0.117	0.775	1.000	1.000	1.000	1.913
Age 13 Bridge - Book 91	1405	1.006	0.106	0.775	1.000	1.000	1.000	1.913
Age 13 Bridge - Book 92	1281	1.000	0.065	0.775	1.000	1.000	1.000	1.913
Age 13 Bridge - Books 93	1256	1.006	0.107	0.775	1.000	1.000	1.000	1.913
Age 17 Bridge - Books 51-56	4622	1.154	0.331	0.990	1.000	1.000	1.005	2.000
Age 17 Bridge - Books 61 - 67	7052	1.148	0.325	0.990	1.000	1.000	1.005	2.000
Age 17 Bridge - Book 90	1786	1.159	0.331	0.990	1.000	1.000	1.930	2.012

* Number includes 5,098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences. An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

Table B-3

Distribution of Session Nonresponse Adjustments - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percent	Max.
Grade 4/Age 9 Main	N/A							
Grade 8/Age 13 Main	N/A							
Grade 12/Age 17 Main	N/A							
Age 9 Bridge - Books 51-56	5188	1.039	0 142	1.000	1 000	1.000	1.000	2.000
Age 9 Bridge - Book 91	1274	1.037	0.153	1.000	1.000	1.000	1.000	1.997
Age 9 Bridge - Book 92	1240	1.039	0.154	1.000	1.000	1.000	1.000	1.997
Age 9 Bridge - Book 93	1197	1.062	0.210	1.000	1.000	1.000	1.000	2.453
Age 13 Bridge - Books 51-56	55[illegible]0	1.006	0.050	1.000	1.000	1.000	1.000	1.434
Age 13 Bridge - Book 90	1938	1.018	0.161	1.000	1.000	1.000	1.000	2.961
Age 13 Bridge - Book 91	1405	0.982	0.099	0.411	1.000	1.000	1.000	1.000
Age 13 Bridge - Book 92	1281	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Age 13 Bridge - Books 93	1256	0.990	0.151	0.464	1.000	1.000	1.000	2.647
Age 17 Bridge - Books 51-56	4622	1.093	0.266	1.000	1.000	1.000	1.000	2.504
Age 17 Bridge - Books 61 - 67	7052	1.068	0.209	1.000	1.000	1.000	1.000	2.000
Age 17 Bridge - Book 90	1786	1.098	0.260	1.000	1.000	1.000	1.000	2.131

Table B-4

Distribution of Age-Only-Eligible Students Nonresponse Adjustments - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	23012	1.010	0.087	1.000	1.000	1.000	1.000	2.314
Grade 8/Age 13 Main	36699*	1.007	0.097	1.000	1.000	1.000	1.000	3.068
Grade 12/Age 17 Main	32710	1.011	0.160	1.000	1.000	1.000	1.000	4.656
Age 9 Bridge - Books 51-56	5188	1.014	0.110	1.000	1.000	1.000	1.000	2.273
Age 9 Bridge - Book 91	1274	1.008	0.087	1.000	1.000	1.000	1.000	2.248
Age 9 Bridge - Book 92	1240	1.010	0.081	1.000	1.000	1.000	1.000	1.780
Age 9 Bridge - Book 93	1197	1.022	0.120	1.000	1.000	1.000	1.000	2.273
Age 13 Bridge - Books 51-56	5500	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Age 13 Bridge - Book 90	1938	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Age 13 Bridge - Book 91	1405	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Age 13 Bridge - Book 92	1281	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Age 13 Bridge - Books 93	1256	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Age 17 Bridge - Books 51-56	4622	1.003	0.065	1.000	1.000	1.000	1.000	2.270
Age 17 Bridge - Books 61 - 67	7052	1.005	0.081	1.000	1.000	1.000	1.000	2.270
Age 17 Bridge - Book 90	1786	1.006	0.080	1.000	1.000	1.000	1.000	2.270

* Number includes 5,098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences. An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

Table B-5

Distribution of Student Nonresponse Adjustments - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	23012	1.080	0.062	1.000	1.044	1.063	1.101	2.121
Grade 8/Age 13 Main	36699*	1.133	0.095	1.000	1.073	1.112	1.159	1.857
Grade 12/Age 17 Main	32710	1.265	0.198	1.000	1.135	1.216	1.320	2.672
Age 9 Bridge - Books 51-56	5188	1.077	0.053	1.000	1.037	1.067	1.108	1.265
Age 9 Bridge - Book 91	1274	1.090	0.069	1.000	1.039	1.077	1.138	1.258
Age 9 Bridge - Book 92	1240	1.088	0.091	1.000	1.021	1.[illegible]5	1.112	1.373
Age 9 Bridge - Book 93	1197	1.081	0.092	1.000	1.019	1.073	1.102	1.667
Age 13 Bridge - Books 51-56	5500	1.113	0.1[illegible]	1.000	1.052	1.091	1.140	2.333
Age 13 Bridge - Book 90	1938	1.121	0.147	1.000	1.048	1.102	1.152	3.667
Age 13 Bridge - Book 91	1405	1.108	0.117	1.000	1.046	1.068	1.138	1.904
Age 13 Bridge - Book 92	1281	1.095	0.107	1.000	1.030	1.065	1.124	2.000
Age 13 Bridge - Books 93	1256	1.121	0.137	1.000	1.053	1.097	1.161	2.200
Age 17 Bridge - Books 51-56	4622	1.277	0.241	1.055	1.167	1.223	1.296	2.749
Age 17 Bridge - Books 61 - 67	7052	1.268	0.262	1.045	1.149	1.205	1.267	3.14[illegible]
Age 17 Bridge - Book 90	1786	1.271	0.210	1.026	1.166	1.232	1.311	2.078

* Number includes 5,098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences. An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

Table B-6

Distribution of Trimming Factors - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	23012	0.995	0.026	0.739	1.000	1.000	1.000	1.000
Grade 8/Age 13 Main	36699*	0.998	0.022	0.729	1.000	1.000	1.000	1.000
Grade 12/Age 17 Main	32710	0.998	0.021	0.760	1.000	1.000	1.000	1.000
Age 9 Bridge - Books 51-56	5188	0.995	0.046	0.560	1.000	1.000	1.000	1.000
Age 9 Bridge - Book 91	1274	0.993	0.032	0.809	1.000	1.000	1.000	1.000
Age 9 Bridge - Book 92	1240	0.996	0.026	0.759	1.000	1.000	1.000	1.000
Age 9 Bridge - Book 93	1197	0.991	0.054	0.556	1.000	1.000	1.000	1.000
Age 13 Bridge - Books 51-56	5500	1.000	0.004	0.970	1.000	1.000	1.000	1.000
Age 13 Bridge - Book 90	1938	0.998	0.012	0.901	1.000	1.000	1.000	1.000
Age 13 Bridge - Book 91	1405	0.999	0.007	0.953	1.000	1.000	1.000	1.000
Age 13 Bridge - Book 92	1281	1.000	0.000	1.000	1.000	1.000	1.000	1.000
Age 13 Bridge - Books 93	1256	0.996	0.024	0.847	1.000	1.000	1.000	1.000
Age 17 Bridge - Books 51-56	4622	0.991	0.053	0.[illegible]71	1.000	1.000	1.000	1.000
Age 17 Bridge - Books 61 - 67	7052	0.992	0.050	0.654	1.000	1.000	1.000	1.000
Age 17 Bridge - Book 90	1786	0.993	0.040	0.691	1.000	1.000	1.000	1.000

* Number includes 5,[illegible] 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences: An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

Table B-7

Distribution of Poststratification Factors - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	23012	1.037	0.228	0.618	0.862	1.146	1.146	1.435
Grade 4/Age 9 Main Winter	12293	0.960	0.244	0.529	0.774	0.995	1.123	1.857
Grade 4/Age 9 Main Spring	10719	1.127	0.253	0.667	0.915	1.174	1.290	1.680
Grade 8/Age 13 Main	36699*	1.046	0.165	0.715	0.939	1.051	1.147	1.381
Grade 8/Age 13 Main Winter	19131	1.007	0.271	0.664	0.833	0.924	1.146	1.680
Grade 8/Age 13 Main Spring	17568	1.083	0.224	0.752	0.958	1.027	1.173	1.556
Grade 12/Age 17 Main	32710	1.120	0.116	0.909	1.009	1.110	1.243	1.380
Grade 12/Age 17 Main Winter	18542	1.086	0.166	0.747	0.955	1.067	1.237	1.272
Grade 12/Age 17 Main Spring	14168	1.152	0.118	0.943	1.061	1.156	1.251	1.591
Age 9 Bridge - Books 51-56	5188	1.010	0.273	0.544	0.858	1.072	1.159	1.984
Age 9 Bridge - Book 91	1274	1.025	0.190	0.656	0.992	1.048	1.118	1.262
Age 9 Bridge - Book 92	1240	1.012	0.250	0.598	0.818	1.093	1.212	1.330
Age 9 Bridge - Book 93	1197	1.084	0.236	0.577	1.024	1.051	1.127	1.331
Age 13 Bridge - Books 51-56	5500	1.072	0.167	0.926	0.970	1.030	1.110	1.548
Age 13 Bridge - Book 90	1938	1.050	0.124	0.89[illegible]	0.939	1.061	1.227	1.252
Age 13 Bridge - Book 91	1405	1.078	0.183	0.785	0.942	1.054	1.236	1.340
Age 13 Bridge - Book 92	1281	1.170	0.177	0.842	0.926	1.043	1.267	1.310
Age 13 Bridge - Books 93	1256	1.281	0.176	0.974	1.044	1.069	1.329	1.596
Age 17 Bridge - Books 51-56	4622	1.259	0.376	0.939	0.997	1.098	1.327	2.210
Age 17 Bridge - Books 51 - 67	7052	1.211	0.376	0.871	1.077	1.108	1.316	2.985
Age 17 Bridge - Book 90	1786	0.993	0.186	0.985	1.087	1.195	1.282	1.625

---

* Number includes 5,098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences: An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

Table B-8

Distribution of Aggregate Adjustments to Base Weights - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	23012	1.245	0.451	0.510	1.012	1.189	1.393	4.702
Grade 8/Age 13 Main	36699*	1.367	0.467	0.542	1.109	1.240	1.510	4.757
Grade 12/Age 17 Main	32710	1.744	0.768	0.970	1.322	1.447	1.793	8.106
Age 9 Bridge - Books 51-56	5188	1.216	0.401	320	0.957	1.184	1.347	3.172
Age 9 Bridge - Book 91	1274	1.191	0.330	84	1.049	1.167	1.320	2.459
Age 9 Bridge - Book 92	1240	1.222	0.393	0.476	1.026	1.213	1.389	2.947
Age 9 Bridge - Book 93	1197	1.230	0.437	0.377	1.052	1.161	1.403	3.645
Age 13 Bridge - Books 51-56	5500	1.220	0.292	0.728	1.033	1.127	1.300	3.772
Age 13 Bridge Book 90	1938	1.230	0.315	0.802	1.069	1.186	1.315	4.178
Age 13 Bridge - Book 91	1405	1.145	0.275	0.414	0.969	1.112	1.293	2.354
Age 13 Bridge - Book 92	1281	1.178	0.224	0.830	0.980	1.189	1.332	2.229
Age 13 Bridge - Books 93	1256	1.305	0.404	0.496	1.089	1.257	1.450	3.685
Age 17 Bridge - Books 51-56	4622	2.027	0.956	1.021	1.266	1.639	2.554	6.708
Age 17 Bridge - Books 61 - 67	7052	1.951	0.993	0.930	1.274	1.608	2.334	9.134
Age 17 Bridge - Book 90	1786	1.937	0.778	1.110	1.394	1.592	2.286	5.495

---

* Number includes 5,098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences: An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

531

532

Table B-9

Distribution of Final Student Weight - Assessed Students

Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Grade 4/Age 9 Main	23012	194.5	100.6	37.6	122.7	170.0	261.2	1517.4
Grade 8/Age 13 Main	36699*	118.2	65.1	17.7	74.8	104.9	149.7	1449.6
Grade 12/Age 17 Main	32710	123.0	71.4	31.6	68.0	114.2	152.5	840.0
Age 9 Bridge - Books 51-56	5188	863.0	518.6	134.7	462.8	727.8	1155.3	6584.1
Age 9 Bridge - Book 91	1274	2520.6	1492.7	242.6	1369.7	2222.9	3355.1	9311.3
Age 9 Bridge - Book 92	1240	2586.7	1589.8	227.8	1381.4	2163.3	3497.3	7898.9
Age 9 Bridge - Book 93	1197	2681.8	1654.8	470.7	1444.3	2256.7	3788.1	15842.6
Age 13 Bridge - Books 51-56	5500	784.0	349.3	174.6	549.7	740.0	955.1	4352.6
Age 13 Bridge - Book 90	1938	1569.6	610.4	326.4	1176.3	1484.6	1961.7	6029.8
Age 13 Bridge - Book 91	1405	2170.4	767.0	685.2	1612.6	2163.4	2688.5	4436.1
Age 13 Bridge - Book 92	1281	2373.0	1017.8	587.5	1859.2	2265.3	2797.9	12106.2
Age 13 Bridge - Books 93	1256	2443.2	1205.5	630.3	1821.3	2328.9	2994.6	15380.3
Age 17 Bridge - Books 51-56	4622	971.0	674.2	268.5	536.4	802.8	1152.6	5870.1
Age 17 Bridge - Books 61 - 67	7052	636.4	463.8	125.6	349.2	531.0	744.1	6121.6
Age 17 Bridge - Book 90	1786	1925.8	1226.2	612.4	1106.6	1566.0	2251.3	8733.1

* Number includes 5,098 13-year-old students who were assessed as part of the International Assessment of Mathematics and Science (see *A World of Differences. An International Assessment of Mathematics and Science. Technical Report* [King, Bertrand, & Dupuis, 1989]).

Table B-10

Distributions of Weight Components for Excluded Students

Component/Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Base Weight								
Grade 4/Age 9	2254	83.8	45.8	25.6	58.3	66.2	108.3	292.7
Grade 8/Age 13	3064	65.5	36.0	12.8	43.0	59.7	82.0	494.5
Grade 11/Age 17	538	174.7	83.8	78.7	107.2	140.6	228.7	388.1
Grade 12/Age 17	1527	70.5	51.3	25.0	41.8	50.3	98.1	330.9
School Nonresponse Adjustment								
Grade 4/Age 9	2254	1.070	0.216	0.796	1.000	1.000	1.000	3.174
Grade 8/Age 13	3064	1.119	0.295	0.760	1.000	1.000	1.058	2.847
Grade 11/Age 17	538	1.158	0.338	0.990	1.000	1.000	1.078	2.000
Grade 12/Age 17	1527	1.212	0.393	0.872	1.000	1.000	1.260	3.045
Session Nonresponse Adjustment								
Grade 4/Age 9	2254	1.048	0.153	1.000	1.000	1.000	1.000	2.000
Grade 8/Age 13	3064	1.003	0.064	1.000	1.000	1.000	1.000	1.513
Grade 11/Age 17	538	1.046	0.158	1.000	1.000	1.000	1.000	2.000
Grade 12/Age 17	N/A							
Age-only Eligible Nonresponse Adjustment								
Grade 4/Age 9	2254	1.007	0.066	1.000	1.000	1.000	1.000	2.039
Grade 8/Age 13	3064	1.006	0.089	1.000	1.000	1.000	1.000	3.068
Grade 11/Age 17	538	1.009	0.109	1.000	1.000	1.000	1.000	2.270
Grade 12/Age 17	1527	1.026	0.179	1.000	1.000	1.000	1.000	4.656
Student Nonresponse Adjustment								
Grade 4/Age 9	2254	1.076	0.168	1.000	1.000	1.004	1.040	1.882
Grade 8/Age 13	3064	1.064	0.188	1.000	1.000	1.000	1.041	3.000
Grade 11/Age 17	538	1.06	0.114	1.000	1.000	1.000	1.056	1.444
Grade 12/Age 17	1527	1.038	0.117	1.000	1.000	1.000	1.032	1.807

Table B-10 (continued)

Distributions of Weight Components for Excluded Students

Component/Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Trimming Factor								
Grade 4/Age 9	2253	0.936	0.179	0.301	1.000	1.000	1.000	1.000
Grade 8/Age 13	3064	0.988	0.051	0.622	1.000	1.000	1.000	1.000
Grade 11/Age 17	538	0.961	0.143	0.363	1.000	1.000	1.000	1.000
Grade 12/Age 17	1527	0.972	0.131	0.180	1.000	1.000	1.000	1.000
Poststratification Factor								
Grade 4/Age 9	2253	0.981	0.271	0.618	0.696	1.069	1.164	1.447
Grade 8/Age 13	3064	1.062	0.186	0.713	0.939	1.082	1.222	1.377
Grade 11/Age 17	538	1.509	0.401	0.912	1.104	1.548	1.805	2.465
Grade 12/Age 17	1527	1.109	0.130	0.909	0.991	1.110	1.209	1.380
Combined Weighted Adjustments								
Grade 4/Age 9	2253	1.117	0.528	0.314	0.696	1.107	1.349	4.282
Grade 8/Age 13	3064	1.256	0.461	0.596	0.958	1.184	1.317	3.933
Grade 11/Age 17	538	1.814	0.766	0.912	1.280	1.692	2.019	6.547
Grade 12/Age 17	1527	1.365	0.526	0.223	1.091	1.235	1.509	5.921
Final Student Weight								
Grade 4/Age 9	2253	102.9	78.4	18.3	45.1	81.3	139.1	566.4
Grade 8/Age 13	3064	80.3	49.5	9.1	52.0	70.4	96.4	641.2
Grade 11/Age 17	538	318.9	231.3	86.9	183.2	246.5	384.8	1889[illegible]
Grade 12/Age 17	1527	88.0	47.8	23.7	51.6	71.6	115.3	333.[illegible]

Table B-11

Distributions of Weight Components for Teacher Data Files

Component/Sample	No. of Cases	Mean	Standard Deviation	Min.	25th Percentile	Median	75th Percentile	Max.
Student Nonresponse Adjusted Student Weight								
Grade 4	3901	790.0	403.2	251.7	473.9	784.3	944.6	3004.2
Grade 8	3570	869.8	522.1	157.0	540.7	762.4	1039.1	5344.7
Teacher Nonresponse Adjustment								
Grade 4	3901	1.138	0.227	0.946	1.008	1.042	1.158	2.319
Grade 8	3570	1.174	0.255	0.986	1.026	1.095	1.219	3.359
Trimming Factor								
Grade 4	3901	0.994	0.036	0.715	1.000	1.000	1.000	1.000
Grade 8	3570	0.997	0.023	0.757	1.000	1.000	1.000	1.000
Poststratification Factor								
Grade 4	3901	1.050	0.194	0.731	1.056	1.067	1.212	1.415
Grade 8	3570	1.021	0.111	0.844	0.967	0.981	1.133	1.170
Final Teacher Data Student Weight								
Grade 4	3901	835.5	467.0	184.1	499.0	752.1	015.4	3283.0
Grade 8	3570	873.1	488.3	155.5	557.3	772.3	1089.6	5241.0

539 540

APPENDIX C

Contrast Codings and Estimated Effects for 1988 NAEP Conditioning Variables

Appendix C

CONTRAST CODINGS AND ESTIMATED EFFECTS FOR 1988 NAEP CONDITIONING VARIABLES

This appendix contains information about the conditioning variables used in the construction of plausible values for the 1988 assessments of reading, writing, civics, U.S. history, geography, mathematics, and science.

The first part of the appendix gives the contrast codings for each set of conditioning variables used in 1988. Codings for the common conditioning variables, which were used for cross-sectional studies in reading, writing, civics, U.S. history, and geography, are given in Table C-1. In addition, subject-specific conditioning variables used in each cross-sectional study are given for reading (Table C-2), writing (C-4), civics (C-6), U.S. history (C-8), and geography (C-9). The complete set of conditioning variables for each subject area in which plausible values were constructed for trend studies are given for reading (Table C-3), writing (C-5), civics (C-7), mathematics (C-10), and science (C-11).

The second part of the appendix shows the estimated effects, by subject area, sample(s), and age class, for the conditioning variables used in cross-sectional studies for reading (Tables C-12 to C-18), civics (C-22 to C-27), U.S. history (C-34 to C-36), and geography (C-37) and for the conditioning variables used in trend studies for reading (C-19 to C-21), civics (C-28 to C-33), mathematics (C-38 to C-40), and science (C-41 to C-43).

Note that all effect estimates are in the metrics used in the original calibration of the scale. The transformations needed to represent these effects in terms of the metric of the final reporting scales appear in the chapters that describe the scaling of each subject area. Note also that certain conditioning variables do not have effect estimates. This is because those variables are approximate linear combinations of the other conditioning variables.

Some conditioning variables were constructed by recoding the values of a data variable or by combining and recoding data from two or more variables. A description of how these conditioning variables were derived for each subject area is provided in Appendix D.

Tables C-1 through C-11

CONTRAST CODINGS FOR 1988 NAEP CONDITIONING VARIABLES

Table C-1

Contrast Codings for 1988 Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Overall	All	---	1
Gender	All	1 Male	0
		2 Female	1
Ethnicity	All	1 White	000
		2 Black	100
		3 Hispanic	010
		4 Asian American	001
		5 American Indian	000
		6 Unclassified	000
		BLK Missing	000
STOC	All	2 Low Metro	00
		3 High Metro	10
		1,4-7 All Others and Missing	01
Region	All	1 Northeast	000
		2 Southeast	100
		3 Central	010
		4 West	001
Parents' Education	All	1 < High School	0000
		2 High School Grad	1000
		3 Post-High School	0100
		4 College Grad	0010
		BLK Missing and I Don't Know	0001
Items in the Home (Items asked about are newspaper, >25 books, enclopedia, and magazines. Two or more missing - Missing.)	All	1 0 to 2 of the four items	00
		2 Three of the four items	10
		3 Four of the four items	01
		BLK Missing	00

---

* Multicolumn entries without overbars indicate multiple contrasts. "Items in the home," for example, induces two contrasts. A response of 2 vs. all other responses, and a response of 3 vs. all other responses. Barred columns treated as one contrast.

Table C-1 (continued)

Contrast Codings for 1988 Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
TV Watching	All	1 None	0 $\overline{00}$
		2 One hour or less per day	1 01
		3 Two hours	2 04
		4 Three hours	3 09
		5 Four hours	4 16
		6 Five hours	5 25
		7 Six or more hours per day	6 36
		BLK Missing	3 09
Home Language Minority (How often do people in your home speak a language other than English)	All	1 Never	0
		2,3 Sometimes, Always	1
		BLK Missing	0
Homework	9	1 Don't have any	100
		2 Don't do any	010
		3 1/2 hour	011
		4 One hour	012
		5 > One hour	013
		BLK Missing	000
	13, 17	1 Don't have any	100
		2 Don't do any	010
		3 1/2 hour	011
		4 One hour	012
		5 Two hours	013
		6 > Two hours	014
		BLK Missing	000
Percent in Lunch Program	All	0	$\overline{000}$ 0
		1	001 0
		2	002 0
		.	.
		.	.
		99	099 0
		100	100 0
		BLK	000 1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-1 (continued)

Contrast Codings for 1988 Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Percent White in School	All	0 - 49 White Minority	10
		50 - 79 Integrated	01
		80 - 100 Predominantly White	00
		BLK Missing	00
Age by Grade	All	1 < Modal Age, Modal Grade	0000
		2 Modal Age, < Modal Grade	1000
		3 Modal Age, Modal Grade, and Missing	0100
		4 Modal Age, > Modal Grade**	0010
		5 > Modal Age, Modal Grade	0001
Public v. Private Schools	All	1 Public	0
		2 Private	1
		3 Catholic	1
		4 Bureau of Indian Affairs	1
		5 Dept. of Defense	1
		BLK Missing	1
Someone at Home Helps with Homework	All	1 Almost Every Day	1
		2 Once or Twice a Week	1
		3 Once or Twice a Month	0
		4 Never	0
		5 Don't Have Homework	0
		BLK Missing	0
Went to Preschool	9	1 Yes	1
		2 No	0
		3 I Don't Know	0
		BLK Missing	0
Single/Multiple Parent at Home	All	1 Yes to Father and Mother at Home	1
		2 Any Other Responses	0
		BLK Missing	0

---

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

** Category not applicable for age class 17.

Table C-1 (continued)

Contrast Codings for 1988 Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Mother at Home (Does your mother live at home)	All	1 Yes	1
		2 No	0
		BLK Missing	0
Mother Works Outside of Home	9	1 Yes	1
		2 No	0
		3 Mother Not at Home	0
		BLK Missing	0
	13, 17	1 Yes, Full-Time	1
		2 Yes, Part-Time	1
		3 No	0
		4 Mother Not at Home	0
		BLK Missing	0
Grownup at Home Right After School	9	1 Yes	1
		0 No	0
		BLK Missing	0
Pages a Day Read for School and Homework	All	1 More Than 20	11
		2 16 - 20	11
		3 11 - 15	11
		4 6 - 10	10
		5 5 or Fewer	00
		BLK Missing	00
Do You Expect to Graduate from High School	13	1 Yes	1
		2 No	0
		BLK Missing	0
Days of School Missed Last Month	13,17	1 None	1
		2 1 or 2 Days	1
		3 3 or 4 Days	0
		4 5 to 10 Days	0
		5 More than 10 Days	0
		BLK Missing	0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-1 (continued)

Contrast Codings for 1988 Common Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Grades in School	13, 17	1 A	4.0
		2 A-B	3.5
		3 B	3.0
		4 B-C	2.5
		5 C	2.0
		6 C-D	1.5
		7 D	1.0
		8 < D	0.5
		BLK Missing	2.0
High School Program	17	1 General	00
		2 College Preparatory	10
		3 Vocational, Technical	01
		BLK Missing	00
Post-secondary Plans	17	1 Work Full Time	00
		2 Two-year College	10
		3 Four-year College	01
		4 Other	00
		BLK Missing	00
Hours of Outside Work	17	1 None	0
		2 < 6 Hours	4
		3 6 to 10 Hours	8
		4 11 to 15 Hours	13
		5 16 to 20 Hours	18
		6 21 to 25 Hours	23
		7 26 to 30 Hours	28
		8 > 30 Hours	40
		BLK Missing	0
Type of English Class In	17	1 Not Taking	00
		2 Advanced Placement	10
		3 College Preparatory	10
		4 General	00
		5 Remedial	01
		BLK Missing	00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Home environment and support	All	RHOME	Low	1
			Medium	2
			High	3
Students' independent reading	All	RINDRDG	Low	1
			Medium	2
			High	3
Students' use of materials	13, 17	RUSEMAT	Low	1
			Medium	2
			High	3
Teacher instructional behaviors	13, 17	RTEACH	Low	1
			Medium	2
			High	3
Students' behaviors and support	17	RBEHAV	Low	1
			Medium	2
			High	3
Reading strategies	17	RSTRATG	Low	1
			Medium	2
			High	3
Students' school and coursework	17	RSCHWRK	Low	1
			Medium	2
			High	3
Magazines at home	9	R800101	Yes	1
			No	2
			Missing	2
Books at home	All	R800301	10 or fewer	1
			11 to 20	2
			21 to 30	3
			More than 30	4
			Missing	1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Read stories	All	R800401	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			Never or hardly ever	4
			I don't know	4
			Missing	4
Read aloud	9	R800501	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			Never or hardly ever	4
			Missing	4
Read for fun	All	S003501	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Tell about book	All	S003502	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Books from library	All	S003503	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Buy own books	All	S003504	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Kind of reader	All	S003301	A poor reader	1
			A good reader	2
			A very good reader	3
			I don't know	1
			Missing	1
Teacher asks	9	R800701	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			Never or hardly ever	4
			Missing	4
Magazines at home	13, 17	R800201	Yes	1
			No	2
			Missing	2
Talk about reading	13, 17	R800601	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			Never or hardly ever	4
			Missing	4
Read after TV show	13, 17	S003505	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Author you like	13, 17	S003506	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Read in spare time	13, 17	L801101	Never	1
			Fiction	2
			Nonfiction	2
			Fiction and nonfiction	2
			Missing	1
Use dictionary	13, 17	R800801	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Use encyclopedia	13, 17	S007309	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Analyze reading	13, 17	R801201	A lot	1
			Some	2
			Not at all	3
			Missing	3
Write on reading	13, 17	S008501	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Teacher-main idea	13, 17	S004602	Almost every time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			Never or hardly ever	5
			Missing	5
Teacher-vocabulary	13, 17	S004601	Almost every time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			Never or hardly ever	5
			Missing	
Teacher-questions	13, 17	S004701	Almost every time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			Never or hardly ever	5
			Missing	5
Can't understand	13, 17	S008601	A lot	1
			Some	2
			None	3
			Missing	1
Read story, novel	17	S004301	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Read newspaper	17	S004304	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	[illegible]

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Read magazine	17	S004305	Almost every day	1
			Once or twice a week	2
			Once or twice a month	3
			A few times a year	4
			Never or hardly ever	5
			Missing	5
Read material over	17	S005101	Almost every time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			Never or hardly ever	5
			Missing	5
Take notes	17	S005102	Almost every time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			Never or hardly ever	5
			Missing	5
Make outlines	17	S005103	Almost every time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			Never or hardly ever	5
			Missing	5
Answer questions	17	S005104	Almost every time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			Never or hardly ever	5
			Missing	5
AP English	17	S006403	Yes	1
			No	2
			Missing	2

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Remedial English	17	S006401	Yes	1
			No	2
			Missing	1
Time on literature	17	R800901	Most of the time	1
			More than 1/2 the time	2
			About 1/2 the time	3
			Less than 1/2 the time	4
			None or almost none	5
			Missing	5
Time on homework	17	R801001	None assigned	0
			Don't do it	0
			Less than 1 hour	1
			1 hour	2
			2 hours	3
			3 hours	4
			4 hours	5
			5 or more hours	6
			Missing	0
English grades	17	R801101	Mostly A	1
			A and B	2
			Mostly B	3
			B and C	4
			Mostly C	5
			C and D	6
			Mostly D	7
			Mostly below D	8
			Missing	8
Reading level of student	Grd. 4	T019701	1 above grade level	1
			2 at grade level	2
			3 below grade level	3
			Missing	3

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Level of student's reading group	Grd. 4	T020001	1 high	11
			2 average	12
			3 low	13
			4 do not have groups	00
			Missing	00
Help student with reading with reading aloud aloud	Grd. 4	T020111	1 almost every day	1
			2 several times a week	2
			3 once a week or less	3
			Missing	1
Help student with comprehensive skills	Grd. 4	T020121	1 almost every day	1
			2 several times a week	2
			3 once a week or less	3
			Missing	1
Help student with word-attack skills	Grd. 4	T020131	1 almost every day	1
			2 several times a week	2
			3 once a week or less	3
			Missing	1
Help student with vocabulary	Grd. 4	T020141	1 almost every day	1
			2 several times a week	2
			3 once a week or less	3
			Missing	1
Does student receive remedial reading prog	Grd. 4	T020701	1 yes	1
			2 no	2
			Missing	1
Are students assigned to class by ability	Grd. 4	T012601	1 yes	1
			2 no	2
			Missing	2
Reading ability level of student in class	Grd. 4	T021701	1 mostly high	1
			2 mostly average	2
			3 mostly low	3
			4 mixed ability levels	2
			Missing	2

---

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Time spent on typical day day for direct instruction	Grd. 4	T021801	1 30 minutes or less	1
			2 60 minutes	2
			3 90 minutes	3
			4 120 minutes or more	3
			Missing	3
Percent of reading instruction-individual	Grd. 4	T021901	< 10	11
			10-14	12
			15-24	13
			25-100	14
			Missing	00
Percent of reading instruction-small group	Grd. 4	T021902	0-24	11
			25-49	12
			50-74	13
			75-100	14
			Missing	00
Percent of reading instruction-whole class	Grd. 4	T021903	0-24	11
			25-49	12
			50-74	13
			75-100	14
			Missing	00
Are students assigned to groups by ability	Grd. 4	T022101	1 yes	1
			2 no	2
			3 Do not form reading groups	3
			Missing	3
Use same reading book for all students in class	Grd. 4	T022201	1 yes	100
			2 No, diff levels within same basal	010
			3 No, use different basal series	001
			4 No, use other books or magazines	000

---

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Number of resources used (of 5--book collection, children's paper, reading kits, computer, instruction software	Grd. 4	T022301- T022305	0	0
			1	1
			2	2
			3	3
			4	4
			5	5
How often - read aloud to students	Grd. 4	T022421	1 almost every day	1
			2 1-2 times a week	2
			3 1-2 times a month	3
			4 a few times a year	4
			5 never	5
			Missing	5
How often - have students meet in small groups	Grd. 4	T022431	1 almost every day	1
			2 1-2 times a week	2
			3 1-2 times a month	3
			4 a few times a year	4
			5 never	5
			Missing	5
How often - have students write about something they read	Grd. 4	T022441	1 almost every day	1
			2 1-2 times a week	2
			3 1-2 times a month	3
			4 a few times a year	4
			5 never	5
			Missing	5
How often - have students complete workbooks	Grd. 4	T022451	1 almost every day	1
			2 1-2 times a week	2
			3 1-2 times a month	3
			4 a few times a year	4
			5 never	5
			Missing	5

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
How often - have students read books they choose	Grd. 4	T022461	1 almost every day	1
			2 1-2 times a week	2
			3 1-2 times a month	3
			4 a few times a year	4
			5 never	5
			Missing	5
How often - have students read information material	Grd. 4	T022471	1 almost every day	1
			2 1-2 times a week	2
			3 1-2 times a month	3
			4 a few times a year	4
			5 never	5
			Missing	5
How often - take class to the library	Grd. 4	T022481	1 almost every day	1
			2 1-2 times a week	2
			3 1-2 times a month	3
			4 a few times a year	4
			5 never	5
			Missing	5
Highest academic degree held	Grd. 4	T023201	1 High school diploma	1
			2 Associate degree/ voc certificate	2
			3 Bachelor's degree	2
			4 Master's degree	3
			5 Educatnl specialist	4
			6 Doctorate	4
			7 First professional degree	4
			Missing	2
Number of special trainings in teaching reading (of 4--inservice, undergrad, graduate, continuing educ.)	Grd. 4	T023502-T023505	0	0
			1	1
			2	2
			3	3
			4	4

---

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-2 (continued)

Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Ability to get desired instructional materials and resources	Grd. 4	T024001	1 get all needed	1
			2 get most needed	2
			3 get some needed	3
			4 don't get any needed	3
			Missing	3
If could start over, would become a teacher	Grd. 4	T024101	1 certainly	1
			2 probably	2
			3 chances about even	3
			4 probably not	4
			5 certainly not	5
			Missing	3

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-3

Contrast Codings for 1988 Reading Trend Conditioning Variables

Conditioning Variable	Ages	Variable Coding	Contrast Coding*
Overall	All	---	1
Gender	All	1 Male	0
		2 Female	1
Ethnicity	All	1 White	000
		2 Black	100
		3 Hispanic	010
		4 Asian American	001
		5 American Indian	000
		6 Unclassified	000
STOC	All	2 Low Metro	00
		3 High Metro	10
		1,4-7 All Others and Missing	01
Region	All	1 Northeast	000
		2 Southeast	100
		3 Central	010
		4 West	001
Parents' Education	All	1 < High School	0000
		2 High School Grad	1000
		3 Post-High School	0100
		4 College Grad	0010
		BLK Missing and I Don't Know	0001
Items in the Home (the items asked are: Newspaper, Dictionary, > 25 books, Encyclopedia, Video games, Magazines)	All	0 None of the six items	00
		1 One of the six items	10
		2 Two of the six items	20
		3 Three of the six items	30
		4 Four of the six items	40
		5 Five of the six items	50
		6 Six of the six items	60
		BLK Missing	01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-3 (continued)

Contrast Codings for 1988 Reading Trend Conditioning Variables

Conditioning Variable	Ages	Variable Coding	Contrast Coding*
TV Watching	All	1 None	0 00
		2 One hour or less per day	1 01
		3 Two hours	2 04
		4 Three hours	3 09
		5 Four hours	4 16
		6 Five hours	5 25
		7 Six or more hours per day	6 36
		BLK Missing	3 09
Homework	All	1 Don't have any	00
		2 Don't do any	00
		3 < one hour	10
		4 One to two hours	20
		5 > two hours	30
		BLK Missing	01
Home Language Minority	All	1 English	00
		2 Spanish	10
		3 Other	10
		BLK Missing	01
Pages Read	All	1 More than 20	10
		2 16-20	10
		3 11-15	10
		4 6-10	10
		5 5 or fewer	00
		BLK Missing	01
Percent in Lunch Program	All	0	0000
		1	0010
		2	0020
		.	.
		.	.
		99	0990
		100	1000
		BLK Missing	0001

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-3 (continued)

Contrast Codings for 1988 Reading Trend Conditioning Variables

Conditioning Variable	Ages	Variable Coding	Contrast Coding*
Percent White in School	All	0-49 White Minority	100
		50-79 Integrated	010
		80-100 Predominantly White	001
		BLK Missing	000
Courses Taken (The courses asked about are Gym, Art, Music, Foreign Lang, Computer, Drama, Science. 3 or more missing = missing)	9,13	0 None of the seven	00.00
		1 One of the seven	01.00
		2 Two of the seven	02.00
		3 Three of the seven	03.00
		4 Four of the seven	04.00
		5 Five of the seven	05.00
		6 Six of the seven	06.00
		7 All seven	07.00
		BLK Missing	00.01
Age by Grade	All	1 < modal age, At modal grade	0000
		2 At modal age, < modal grade	1000
		3 At modal age, At modal grade	0100
		4 At modal age, > modal grade	0010
		5 > modal age, At modal grade	0001
		BLK Missing	0100

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-4

Contrast Codings for 1988 Writing Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Coding	Contrast Coding*
Enjoy Writing	All	1. Almost Always	5
		2. More than half the time	4
		3. Half the time	3
		4. Less than half the time	2
		5. never	1
		BLK missing	3
Positive Teacher Feedback	All	0 No positive feedback	0
		1 Little positive feedback	1
		2 Some positive feedback	2
		3 A Lot of positive feedback	3
		BLK missing	0
Negative Teacher Feedback	All	0 no negative feedback	0
		1 Little negative feedback	1
		2 Lots of negative feedback	2
		BLK missing	0
Amount Written for English Class	8,12	alot of short, never long	100000
		a lot of short, a lot of long	010000
		a lot of short, some long	001000
		some short, some long	000100
		some short, never long	000010
		some short, a lot of long	000001
		never short, a lot of long	000001
		never short, some long	000001
		never short, never long	000000
		either missing	000100
Amount Written for History Class	8,12	a lot of short, never long	100000
		a lot of short, a lot of long	010000
		a lot of short, some long	001000
		some short, some long	000100
		some short, never long	000010
		some short, a lot of long	000001
		never short, a lot of long	000001
		never short, some long	000001
		never short, never long	000000
		either missing	000100

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-4 (continued)

Contrast Codings for 1988 Writing Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Coding	Contrast Coding*
Instruction on Writing Process II	8,12	1. Almost Always	5
		2. More than half the time	4
		3. Half time	3
		4. Less than half the time	2
		5. Never	1
		BLK missing	3
Success in English Class II	8,12	1. a lot	1
		2. some	2
		3. none	3
		BLK missing	2
Success in English Class III	12	1 Mostly A	8
		2 Between A and B	7
		3 Mostly B	6
		4 Between B and C	5
		5 Mostly C	4
		6 Between C and D	3
		7 Mostly D	2
		8 Mostly < D	1
		BLK missing	5
Time Spent on Writing III	12	1 lots of time	4
		2 some time	3
		3 little time	2
		4 no time	1
		BLK missing	2
Revisions	12	1 Almost Always	5
		2 More than half the time	4
		3 Half the time	3
		4 Less than half the time	2
		5 Never	1
		BLK Missing	3

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-4 (continued)

Contrast Codings for 1988 Writing Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Coding	Contrast Coding*
Planning	12	1 Almost Always	5
		2 More than half the time	4
		3 Half the time	3
		4 Less than half the time	2
		5 Never	1
		BLK missing	3
Time Spent on Writing II	8	1 Weekly	4
		2 Monthly	3
		3 Yearly	2
		4 Never	1
		BLK Missing	3
Time Spent on Writing I	4	1 Weekly	100
		2 Monthly	010
		3 Yearly	001
		4 never	000
		BLK missing	010
Success in English Class I	4	1 Daily	1000
		2 Weekly	0100
		3 Monthly	0010
		4 Yearly	0001
		5 Never	0000
		BLK Missing	0010
Instruction on Writing Process I	4	1 Almost Always	5
		2 More than half the time	4
		3 Half time	3
		2 Less than half the time	2
		1 Never	1
		BLK Missing	3

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-5

Contrast Codings for 1988 Writing Trend Conditioning Variables

Conditioning Variable	Ages	Variable Coding	Contrast Coding*
Overall	All	--	1
Sex	All	1 Male	0
		2 Female	1
Ethnicity	All	1 White	10
		2 Black	00
		3 Hispanic	01
		4 Asian American	10
		5 American Indian	10
		6 Unclassified	10
		BLK Missing	10
STOC	All	2 Low Metro	00
		3 High Metro	10
		1,4-7 All others	01
Region	All	1 Northeast	000
		2 Southeast	100
		3 Central	010
		4 West	001
Parents' Education	All	1 < High School	0000
		2 High School Grad	1000
		3 Post-High School	0100
		4 College Grad	0010
		5 I Don't Know, Missing	0001
Grade	All	4	00
		8	10
		11	01
Items in the Home	All	1 0 to 3 of five items	00
		2 4 of five items	10
		3 all five items	01
		BLK missing	00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-5 (continued)

Contrast Codings for 1988 Writing Trend Conditioning Variables

Conditioning Variable	Ages	Variable Coding	Contrast Coding*
Modal age	All	1 less than modal age	10
		2 modal age	00
		3 greater than modal age	01
Homework	All	1 Don't have any	00000
		2 0 done	10000
		3 less than 1 hour	01000
		4 1 to 2 hours	00100
		5 2+ hours	00010
		BLK missing	00001
School Type	All	1 public	0
		2 private	1
		3 Catholic	1
		4 Bureau of Indian Affairs	1
		5 Dept. of Defense	1
		missing	1
TV Watching	All	1 less than 1 hour	0 00
		2 1 hour	1 01
		3 2 hours	2 04
		4 3 hours	3 09
		5 4 hours	4 16
		6 5 hours	5 25
		7 more than 6 hours	6 36
		BLK missing	3 09
Mother's working Outside of home	All	1 yes	1
		2 no	0
		BLK missing	0
Primary language Spoken in the home	All	1 English	00
		2 Spanish	10
		3 other	01
		4 missing	01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-5 (continued)

Contrast Codings for 1988 Writing Trend Conditioning Variables

Conditioning Variable	Ages	Variable Coding	Contrast Coding*
Percent Lunch Program	All	0	000 0
		1	001 0
		2	002 0
		.	.
		99	099 0
		100	100 0
		BLK missing	000 1
Percent of White Students in School	All	0-49.9 Minority school	00
		50-79.9 Integrated school	10
		80-100 Predominantly White	01
		BLK Missing	01
Grades in School	All	1. Mostly A	4.0
		2. Mostly B	3.0
		3. Mostly C	2.0
		4. Mostly D	1.0
		5. Less than D	0.0
		6. between A and B	3.5
		7. between B and C	2.5
		8. between C and D	1.5
		9. between D and F	0.5
		BLK missing	2.0
Pages a Day Read for School and Homework	All	1. more than 20	100
		2. 16 - 20	100
		3. 11 - 15	100
		4. 6 - 10	010
		5. less than 5	001
		BLK missing	000
Number of Reports Written	All	0 no reports, or missing	0
		1 one report	1
		2 two reports	2
		3 three reports	3
		4 four reports	4
		5 five reports	5
		6 six reports	6
		7 seven reports	7

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Frequency of Social Studies Class	4 (Main and Intercorr.)	S008701	1 everyday	1000
			2 3-4/wk	0100
			3 1-2/wk	0010
			4 less than 1/wk	0001
			5 never or hardly ever	0000
			missing	0000
Studied Government	4 (Main and Intercorr.)	P800501	1 yes	10
			2 no	00
			missing	01
Number of subjects studied a lot (out of 6)	4 (Main)	ST-ALOT	0	0
			1	1
			2	2
			3	3
			4	4
			5	5
			6	6
Studied Laws	4 (Main and Intercorr.)	P800601	1 a lot	1
			2 some	1
			3 not at all	0
			missing	0
Studied Judges and Courts	4 (Main)	P800701	1 a lot	1
			2 some	1
			3 not at all	0
			missing	0
Studied President and Leaders	4 (Main and Intercorr.)	P800801	1 a lot	1
			2 some	1
			3 not at all	0
			missing	0
Studied Elections and Voting	4 (Main)	P800901	1 a lot	1
			2 some	1
			3 not at all	0
			missing	0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6 (continued)

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Studied Community	4 (Main and Intercorr.)	P801001	1 a lot	1
			2 some	1
			3 not at all	0
			missing	0
Studied Rights and Responsibilities	4 (Main)	P801101	1 a lot	1
			2 some	1
			3 not at all	0
			missing	0
Frequency of Current Events	4 (Main)	P802101	1 every day	1000
			2 3-4/wk	0100
			3 1-2/month	0010
			4 few/year	0001
			5 never	0000
			missing	0000
Studied Civics in Fifth Grade	8 (Main and Intercorr.)	P800101	1 yes	1
			2 no	0
			3 don't know	0
			missing	0
Studied Civics in Sixth Grade	8 (Main and Intercorr.)	P800102	1 yes	1
			2 no	0
			3 don't know	0
			missing	0
Studied Civics in Seventh Grade	8 (Main and Intercorr.)	P800103	1 yes	1
			2 no	0
			3 don't know	0
			missing	0
Studied Civics in Eighth Grade	8 (Main and Intercorr.)	P800104	1 yes	1
			2 no	0
			3 don't know	0
			missing	0

---

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6 (continued)

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Studied Civics Incomplete Set of Answers	8 (Main and Intercorr.)	STCV-INC	1	1
			0	0
Number of Subjects Studied a lot (out of 10)	8 (Main)	ST-ALOT	0	0
			1	1
			2	2
			3	3
			4	4
			5	5
			6	6
			7	7
			8	8
			9	9
			10	10
Studied Constitution	8 (Main and Intercorr.)	P801201	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied Congress	8 (Main and Intercorr.)	P801202	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied Laws	8 (Main and Intercorr.)	P801203	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied Courts	8 (Main and Intercorr.)	P801204	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6 (continued)

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Studied President	8 (Main and Intercorr.)	P801205	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied Political Parties	8 (Main and Intercorr.)	P801206	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied State and Local Government	8 (Main and Intercorr.)	P801207	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied Democracy	8 (Main and Intercorr.)	P801208	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied Other Governments	8 (Main)	P801209	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Studied Rights & Responsibilities	8 (Main)	P801210	1 a lot	2
			2 some	2
			3 not at all	1
			missing	0
Active in 4 Areas of Civics (1, 2, 5, 8)	8 (Main)	CVA-ACT	1	1
			0	0
Active in 2 Areas of Civics (3, 10)	8 (Main)	CVB-ACT	1	1
			0	0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6 (continued)

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Active in 2 Areas of Civics (4, 9)	8 (Main)	CVC-ACT	1 0	1 0
Active in 2 Areas of Civics (6, 7)	8 (Main)	CVD-ACT	1 0	1 0
Not Active in 2 Areas of Civics (6, 7)	8 (Main)	CVD-NOT	1 0	1 0
Difficulty Reading Text	8 (Main)	S008801	1 a lot 2 some 3 none missing	1 1 0 1
Studied Civics in Ninth Grade	12 (Main and Intercorr.)	P800201	1 yes 2 no 3 don't know 0	1 0 0 0
Studied Civics in Tenth Grade	12 (Main and Intercorr.)	P800202	1 yes 2 no 3 don't know 0	1 0 0 0
Studied Civics in Eleventh Grade	12 (Main and Intercorr.)	P800203	1 yes 2 no 3 don't know 0	1 0 0 0
Studied Civics in Twelfth Grade	12 (Main and Intercorr.)	P800204	1 yes 2 no 3 don't know missing	1 0 0 0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6 (continued)

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Studied Civics Incomplete Set of Answers	12 (Main and Intercorr.)	STCV-INC	1 0	1 0
Years of Government Study You Have Had	12 (Main and Intercorr.)	GOVT	0 .25 .5 1.0 2.0	0 .25 .5 1.0 2.0
Advanced Placement	12 (Main and Intercorr.)	P800401	1 yes 2 no missing	1 0 0
Number of Subjects Studied a Lot (out of 10)	12 (Main and Intercorr.)	ST-ALOT	1 0	1 0
Average Level of Study for 10 Subjects	12 (Main and Intercorr.)	ST-AVE	1 0	1 0
Level of Interest in Civics	12 (Main)	CIV-INT	3 2 1 1 0	3 2 1 1 0
Mock Election Trials	12 (Main)	P801501	1 several times 2 once or twice 3 never missing	1 1 0 0
Degree of Attention Government Pays	12 (Main)	GOVATTN	2 1 0	2 1 0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6 (continued)

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Took Civics Class	12 (Main)	HADCIV (P801801)	1 yes	1
			2 no	0
			missing	0
Amount of Civics Homework	12 (Main)	CVHWORK	0	0
			.5	.5
			1	1
			2	2
			3	3
			4	4
			5	5
Civics Grades in School	12 (Main)	CVGRADES	4.0	4.0
			3.5	3.5
			3.0	3.0
			2.5	2.5
			2.0	2.0
			1.5	1.5
			1.0	1.0
			.5	.5
Active in 4 Areas of Civics (1, 2, 5, 8)	12 (Main)	CVA-ACT	1	1
			0	0
Active in 2 Areas of Civics (3, 10)	12 (Main)	CVB-ACT	1	1
			0	0
Active in 2 Areas of Civics (4, 9)	12 (Main)	CVC-ACT	1	1
			0	0
Active in 2 Areas of Civics (6, 7)	12 (Main)	CVD-ACT	1	1
			0	0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-6 (continued)

Contrast Codings for 1988 Civics Cross-sectional Conditioning Variables

Conditioning Variable	Grades	Variable Name	Variable Coding	Contrast Coding*
Not Very Active in These 2 Areas of Civics	12 (Main)	CVD-NOT	1 0	1 0
At Least Some Difficulty in Reading or Missing	12 (Main)	RDDIF	1 0	1 0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
Gender	13, 17	1976	SEX	1 Male	0
				2 Female	1
		1982	SEX	see 1976-SEX	
		1988	DSEX	see 1976-SEX	
Ethnicity	13, 17	1976	COLLRACE	1 White	000
				2 Black	100
				3 Hispanic	010
				4 Other	001
				BLK missing	001
		1982	COLLRACE	see 1976-COLLRACE	
		1988	RACE	1 White	000
				2 Black	100
				3 Hispanic	010
				4 Asian Am.	001
				5 Am. Ind.	001
				6 Unclass.	001
				Blk,miss	001
STOC	13, 17	1976	STOC	2 Low Metro	00
				3 High Metro	10
				1,4-7,miss	01
		1982	STOC	see 1976-STOC	
		1988	STOC	see 1976-STOC	
Region	13, 17	1976	REGOBE	1 Northeast	000
				2 Southeast	100
				3 Central	010
				4 West	001
		1982	REGOBE	see 1976-REGOBE	
		1988	REGION	see 1976-REGOBE	

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7 (continued)

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
Parents' Ed.	13, 17	1976	PARED	1 < hs	000
				2 hs grad	100
				3 post-hs	010
				Blk,miss	001
Parents' Ed. 2	13, 17	1982	PPARED	1 < hs	0000
			(derived)	2 hs grad	1000
				3 post-hs	0100
				4 col grad	0010
				Blk,miss,don't know	0001
		1988	PARED	1 < hs	0000
				2 hs grad	1000
				3 post-hs	0100
				4 col grad	0010
				Blk,miss,don't know	0001
Items in the Home	13, 17	1976	HOMEENV	1 0-2 items	00
				2 3 items	10
				3 4 items	01
				Blk,miss,don't know	00
		1982	HOMEENV	see 1976-HOMEENV	
		1988	HOMEEN2	see 1976-HOMEENV	

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7 (continued)

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
TV Watching	13, 17	1976	not avail.		
		1982	E075301	1 none	0 $\overline{00}$
				2 1 hr,less	1 01
				3 1 hr	1 01
				4 2 hrs	2 04
				5 3 hrs	3 09
				6 4 hrs	4 16
				7 5 hrs	5 25
				8 6 or more	6 36
				Blk,miss	3 09
		1988	B001801	1 none	0 00
				2 1 hr,less	1 01
				3 2 hrs	2 04
				4 3 hrs	3 09
				5 4 hrs	4 16
				6 5 hrs	5 25
				7 6 or more	6 36
				Blk,miss	3 09
Home Language Minority	13	1976	not avail.		
		1982	E075602	1 never	0
				2 sometimes	1
				3 often	1
				Blk,miss	0
		1988	B003201	1 never	0
				2 sometimes	1
				3 always	1
				Blk,miss	0
	17	1976	OTHLANG	1 often	1
				2 sometimes	1
				3 never	0
				Blk,miss	0
		1982	E075602	see 1976-OTHLANG	
		1988	B003201	1 never	0
				2 sometimes	1
				3 always	1
				Blk,miss	0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7 (continued)

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
Homework	13, 17	1976	not avail.		
		1982	E075801	1 don't have	100
				2 don't do	010
				3 < 1 hr	011
				4 1-2 hrs	012
				5 > 2 hrs	013
				Blk,miss	000
		1988	B003901	1 don't have	100
				2 don't do	010
				3 1/2 hr	011
				4 1 hr	012
				5 2 hrs	012
				6 > 2 hrs	013
				Blk,miss	000
% White in School	13, 17	1976	PCTWHITE	1 none	10
				2 0-9%	10
				3 10-19%	10
				4 20-29%	10
				5 30-39%	10
				6 40-49%	10
				7 50-59%	01
				8 60-69%	01
				9 70-79%	01
				10 80-89%	00
				11 90-99%	00
				12 100%	00
				Blk,miss	00
		1982	PCTWHITE	0-49 white minor.	10
				50-79 integ.	01
				80-100 white	00
				Blk,miss	00
		1988	PCTWHT	see 1982-PCTWHITE	

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7 (continued)

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
Grade	13	1976	COLLGRD	1 < modal gr	00
				2 7th gr	00
				3 8th gr	10
				4 > modal gr	01
		1982	COLLGRD	see 1976-COLLGRD	
		1988	MODGRD	1 < modal gr	00
				2 8th gr	10
				3 > modal gr	01
	17	1976	COLLGRD	1 < modal gr	00
				2 10th gr	00
				3 11th gr	10
				4 > modal gr	01
		1982	COLLGRD	see 1976-COLLGRD	
		1988	MODGRD	1 < modal gr	00
				2 11th gr	10
				3 > modal gr	01
Public v.	13, 17	1976	not avail.		
Private School		1982	PUBPRVSC	1 public	0
				2 other	1
				Blk,miss	1
		1988	SCHTYPE	1 public	0
				2 private	1
				3 Catholic	1
				4 B. Ind.	1
				5 D. Defense	1
				Blk,miss	1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7 (continued)

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
Grades in School	13	1976	not avail.		
		1982	E075901	1 A	4.0
				2 A-B	3.5
				3 B	3.0
				4 B-C	2.5
				5 C	2.0
				6 C-D	1.5
				7 D	1.0
				8 < D	0.5
				Blk,miss	2.0
		1988	B005401	see 1982-E075901	
	17	1976	GRADES	1 A	4.0
				2 A-B	3.5
				3 B	3.0
				4 B-C	2.5
				5 C	2.0
				6 C-D	1.5
				7 D	1.0
				8 < D	0.5
				Blk,miss	2.0
		1982	E075901	see 1976-GRADES	
		1988	B005401	see 1976-GRADES	
H.S. Program	17	1976	HSPROG	1 general	00
				2 col prep	10
				3 agriculture	01
				4 business	01
				5 distributive	01
				6 health	01
				7 home ec	01
				8 industrial	01
				Blk,miss	00
		1982	E076001	1 general	00
				2 col prep	10
				3 vocational	01
				Blk,miss	00
		1988	B005001	see 1982-E076001	

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7 (continued)

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
Issues in courses	17	1976	E141801	1 true	10
				2 somewhat	01
				3 not true	00
				Blk,miss	01
		1982	not avail.		
		1988	P802501	see 1976-E141801	
Politics in courses	17	1976	E141802	1 true	10
				2 somewhat	01
				3 not true	00
				Blk,miss	01
		1982	not avail.		
		1988	P802601	see 1976-E141802	
Nothing new	17	1976	E141803	1 true	10
				2 somewhat	01
				3 not true	00
				Blk,miss	01
		1982	not avail.		
		1988	P802701	see 1976-E141803	
Knowledge to Participate	17	1976	E141804	1 true	10
				2 somewhat	01
				3 not true	00
				Blk,miss	01
		1982	not avail.		
		1988	P802801	see 1976-E141804	
National Discussions	17	1976	E141401	1 every day	100
				2 1-2/wk.	010
				3 3-4/month	001
				4 hardly ever	000
				Blk,miss	010
		1982	not avail.		
		1988	P802901	see 1976-E141401	

---

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-7 (continued)

Contrast Codings for Civics Trend Conditioning Variables

Conditioning Variable	Ages	Year	Variable Name	Variable Coding	Contrast Coding*
International Discussions	17	1976	E141402	1 every day	100
				2 1-2/wk.	010
				3 3-4/month	001
				4 hardly ever	000
				Blk,miss	010
		1982	not avail.		
		1988	P803001	see 1976-E141402	

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-8

Contrast Codings for 1988 U.S. History Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
How often have social studies class	9	HIS TAKE	Never, hardly ever or missing	0
			Less than once a week	1
			Once or twice a week	2
			Three or four times a week	3
			Every day	4
Ever studied history of our country	9	STUD.HIS	No or missing	0
			Yes	1
Studied early explorers	9	STUD.EXP	Not at all or missing	0
			Some	1
			A lot	2
Studied Indians	9	STUD.IND	Not at all or missing	0
			Some	1
			A lot	2
Studied pilgrims	9	STUD.PIL	Not at all or missing	0
			Some	1
			A lot	2
Studied George Washington	9	STUD.GW	Not at all or missing	0
			Some	1
			A lot	2
Studied pioneers	9	STUD.PIO	Not at all or missing	0
			Some	1
			A lot	2
Studied slavery	9	STUD.SLA	Not at all or missing	0
			Some	1
			A lot	2
Studied people who invent things and make new discoveries	9	STUD.INV	Not at all or missing	0
			Some	1
			A lot	2

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-8 (continued)

Contrast Codings for 1988 U.S. History Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Study U.S. history in grades 5, 6, 7, 8	13	HIS TAKE	Not studied	0
			Studied in 1 grade	1
			Studied in 2 grades	2
			Studied in 3 grades	3
			Studied in 4 grades	4
Studied early periods of U.S. history	13	EARLY H	Not at all	0
			Some	1
			A lot	2
Studied later periods of U.S. history	13	POST CH	Not at all	0
			Some	1
			A lot	2
Social studies teacher asks you to do things	13	TEACH H	Never	0
			Few times a year	1
			Once or twice a month	2
			Once or twice a week	3
			Almost every day	4
Difficulty reading social studies textbook	13	TEXTBOOK	None	0
			Some	1
			A lot	2
Study U.S. history in grade 9, 10, 11, 12	17	HIS TAKE	Not studied	0
			Studied in 1 grade	1
			Studied in 2 grades	2
			Studied in 3 grades	3
			Studied in 4 grades	4
U.S. history coursework completed	17	HIS COUR	None	$\overline{00}$
			< half a year	25
			Half a year	50
			Half a year to 1 year	75
			One year	100
			More than one year	200

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-8 (continued)

Contrast Codings for 1988 U.S. History Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Studied early periods of U.S. history	17	AM HIST	Not at all	0
			Some	1
			A lot	2
Studied later periods of U.S. history	17	WORLD H	Not at all	0
			Some	1
			A lot	2
Studied minorities and women	17	MIN/WOM	Not at all	0
			Some	1
			A lot	2
Like studying U.S. history	17	LIKE H	Never studied	0
			Like other subjects better	1
			Interesting	2
			One of favorites	3
Ever had U.S. history	17	HAD HIST	No	0
			Yes	1
Time spent each week on U.S. history homework	17	HOMEWK H	None assigned	0
			Haven't done it	0
			Less than 1 hour	.5
			1 hour	1
			2 hours	2
			3 hours	3
			4 hours	4
			5 hours or more	5
Grades in U.S. history	17	GRADES H	Mostly A	4.0
			Half A and half B	3.5
			Mostly B	3.0
			Half B and half C	2.5
			Mostly C	2.0
			Half C and half D	1.5
			Mostly D	1.0
			Mostly below D	0.0

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-8 (continued)

Contrast Codings for 1988 U.S. History Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Teacher asks you to do things in class	17	TEACH/H	Never	0
			Few times a year	1
			Once or twice a month	2
			Once or twice a week	3
			Almost every day	4
Difficult reading U.S. history textbook	17	TEXTBOOK	None	0
			Some	1
			A lot	2

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-9

Contrast Codings for 1988 Geography Cross-sectional Conditioning Variables

Conditioning Variable	Age Classes	Variable Name	Variable Coding	Contrast Coding*
Take geography course in grades 9, 10, 11, 12	17	GEO TAKE	Not take	0
			Take in 1 grade	1
			Take in 2 grades	2
			Take in 3 grades	3
			Take in 4 grades	4
Studied physical geography	17	PHYSICAL	Not at all	0
			Very little	1
			Some	2
			A lot	3
Studied social geography	17	SOCIAL	Not at all	0
			Very little	1
			Some	2
			A lot	3
World history and geography coursework	17	W HIS/G	No, haven't taken	0
			Yes	1
State/regional geography coursework	17	STAT/REG	No, haven't taken	0
			Yes	1
U.S. geography coursework	17	US GEO	No, haven't taken	0
			Yes	1
U.S. history and geography coursework	17	US H/G	No, haven't taken	0
				Yes1
Physical geography coursework	17	PHY/GEO	No, haven't taken	0
			Yes	1
Economic and political geography coursework	17	ECON/P	No, haven't taken	0
				Yes1
Human and cultural geography coursework	17	HUM/CULT	No, haven't taken	0
			Yes	1
Urban geography coursework	17	URBAN G	No, haven't taken	0
			Yes	1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-10

Contrast Codings for 1988 Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Overall	All	---	1
Gender	All	1 Male	0
		2 Female	1
Observed Race	All	1 White	000
		2 Black	100
		3 Hispanic	010
		4 Asian American	001
		5 American Indian	000
		6 Other	000
		BLK Missing	000
STOC	All	1,4-7 all except 2&3	00
		2 low metro	10
		3 high metro	01
Region	All	1 Northeast	000
		2 Southeast	100
		3 Central	010
		4 West	001
Parents' Education	All	1 < High school	0000
		2 = High school	1000
		3 > High school	0100
		4 Graduated College	0010
		5 Unknown, missing	0001
Modal grade	All	1 < modal grade	10
		2 = modal grade/missing	00
		3 > modal grade	01
Items in the Home	All	1 0-2 items	00
		2 3 items	10
		3 4 items	01

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-10 (continued)

Contrast Codings for 1988 Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Observed Race x Gender ("White" includes American Indian, Other, and Unclassified.)	All	1 White,Male	000
		2 Black,Male	000
		3 Hispanic,Male	000
		4 Asian American,Male	000
		5 White,Female	000
		6 Black,Female	100
		7 Hispanic,Female	010
		8 Asian American,Female	001
Observed Race x Parents' Education ("White" includes American Indian, Other, and Unclassified)	All	1 White, <HS	0000 0000 0000
		2 White, HS Grad	0000 0000 0000
		3 White, HS +	0000 0000 0000
		4 White, Col Grad	0000 0000 0000
		5 White, missing	0000 0000 0000
		6 Black, <HS	0000 0000 0000
		7 Black, HS Grad	1000 0000 0000
		8 Black, HS +	0100 0000 0000
		9 Black, Col Grad	0010 0000 0000
		10 Black, missing	0001 0000 0000
		11 Hispanic, <HS	0000 0000 0000
		12 Hispanic, HS Grad	0000 1000 0000
		13 Hispanic, HS +	0000 0100 0000
		14 Hispanic, Col Grad	0000 0010 0000
		15 Hispanic, missing	0000 0001 0000
		16 Asian Amer., <HS	0000 0000 0000
		17 Asian Amer., HS Grad	0000 0000 1000
		18 Asian Amer., HS +	0000 0000 0100
		19 Asian Amer., Col Grd	0000 0000 0010
		20 Asian Amer., missing	0000 0000 0001
School Type	All	1 Public	0
		2 Private	1
		3 Catholic	1
		4 Bureau Indian Affairs	1
		5 Department of Defense	1

* Mult column entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-10 (continued)

Contrast Codings for 1988 Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
TV Watching	All	1-3 0-2 hours	100
		4-6 3-5 hours	010
		7 6 + hours	001
		missing	000
Homework	13.17	1 None assigned	100
		2 Didn't do	010
		3 1/2 hour or less	012
		4 1 hour	013
		5 2 hours	014
		6 More than 2 hours	000
		missing	000
Language in the Home	9	1 Never	00
		2 Sometimes	10
		3 Always	01
	13,17	1 Never	00
		2 Occasionally	01
		3 About half the time	01
		4 Most of the time	10
		5 Always	10
Observed Race x Language in the Home	All	1 White,Often	00 00 00
		2 White,Sometimes	00 00 00
		3 White,Never	00 00 00
		4 Black,Often	10 00 00
		5 Black,Sometimes	01 00 00
		6 Black,Never	00 00 00
		7 Hispanic,Often	00 10 00
		8 Hispanic,Some	00 01 00
		9 Hispanic,Never	00 00 00
		10 Asian Am.,Often	00 00 10
		11 Asian Am.,Some	00 00 01
		12 Asian Am.,Never	00 00 00

---

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-10 (continued)

Contrast Codings for 1988 Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
NMATH	17	1 Pre-Algebra	10000
		2 Algebra	01000
		3 Geometry	00100
		4 Algebra 2	00010
		5 Calculus	00001
		6 Something Else	00000
Computer Prog. Class	17	1 Taken	1
		2 Not Taken	0
Grades	13,17	1 A	$\overline{4.0}$
		2 A-B	3.5
		3 B	3.0
		4 B-C	2.5
		5 C	2.0
		6 C-D	1.5
		7 D	1.0
		8 < D	0.5
		Missing	2.0
Type of Math Class	13	1 None	000
		2 Regular Math	100
		3 Pre-Algebra	010
		4 Algebra	001
		5 Other	001
		Missing	000
Are You Studying Computers	9,13	1 Yes	1
		2 No	0
High School Program	17	1 General	00
		2 College Preparatory	10
		3 Vocational,Technical	01
		missing	00

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-10 (continued)

Contrast Codings for 1988 Mathematics Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Derived Race	All	1 White	000
		2 Black	100
		3 Hispanic	010
		4 Asian American	001
		5 Other	000
		missing	000

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Table C-11

Contrast Codings for 1988 Science Trend Conditioning Variables

Conditioning Variable	Age Classes	Variable Coding	Contrast Coding*
Overall	All	---	1
Gender	All	1 Male	0
		2 Female	1
Observed Race	All	1 White	000
		2 Black	100
		3 Hispanic	010
		4 Asian American	001
		5 American Indian	000
		6 Other	000
		BLK Missing	000
STOC	All	1,4-7 all except 2&3	00
		2 low metro	10
		3 high metro	01
Region	All	1 Northeast	000
		2 Southeast	100
		3 Central	010
		4 West	001
Parents' Education	All	1 < High school	0000
		2 = High school	1000
		3 > High school	0100
		4 Graduated College	0010
		5 Unknown, missing	0001
Modal grade	All	1 < modal grade	10
		2 = modal grade/missing	00
		3 > modal grade	01
Items in the Home	All	1 0-2 items	00
		2 3 items	10
		3 4 items	01
School Type	All	1 Public	0
		2 Private	1
		3 Catholic	1
		4 Bureau Indian Affairs	1
		5 Department of Defense	1

* Multicolumn entries without overbars indicate multiple contrasts. Barred columns treated as one contrast.

Tables C-12 through C-43

ESTIMATED EFFECTS FOR 1988 NAEP CONDITIONING VARIABLES

Table C-12

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 4/Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-3.493459	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.088505	SEX (FEMALE)
3	ETHNIC2	-0.293405	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.316179	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.010243	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.293781	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.096133	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.075012	REGION (SOUTHEAST)
9	REGION3	0.009300	REGION (CENTRAL)
10	REGION4	0.027525	REGION (WEST)
11	PARED2	0.344202	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.509880	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.438445	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.314739	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.111460	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.161661	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.209353	TV WATCHING
18	TV. 2	-0.034422	TV WATCHING
19	HOMELNG1	-0.020347	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	0.220639	HOMEWORK (NONE ASSIGNED)
21	HW-YES	0.222398	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	0.006786	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.001664	PERCENT IN LUNCH PROGRAM
24	%LUNCH 2	-0.123354	MISSING LUNCH PROGRAM
25	%WHITE 1	-0.054569	WHITE MINORITY
26	%WHITE 2	-0.031799	INTEGRATED
27	AGE/GRD1	-0.604877	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	0.010017	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD3	0.544105	MODAL AGE, > MODAL GRADE
30	AGE/GRD4	-0.193332	> MODAL AGE, MODAL GRADE
31	SCH.TYPE	0.182076	PUBLIC V. PRIVATE SCHOOLS
32	H/HOMEWK	-0.106614	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
33	PRESCH	0.173482	WENT TO PRESCHOOL
34	PAR/HOME	0.018582	SINGLE/MULTIPLE PARENT (MULTIPLE)
35	MOTH/HOM	0.268381	MOTHER AT HOME (YES)
36	MOTH/WRK	-0.065688	MOTHER WORK OUTSIDE HOME (YES)
37	GRWN HOM	0.098832	GROWNUP AT HOME RIGHT AFTER SCHOOL
38	PAGE/RD1	0.129134	PAGES A DAY READ (6 AND UP)
39	PAGE/RD2	-0.000352	PAGES A DAY READ (11 AND UP)
40	RDGCOMP1	0.095330	READING COMPOSITE # 1
41	RDGCOMP2	0.060951	READING COMPOSITE # 1

(continued)

Table C-12 (continued)

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 4/Age 9

	Variable	Estimated Effect	Description
42	RDGBKGD1	0.110859	READING BACKGROUND # 1
43	RDGBKGD2	0.005445	READING BACKGROUND # 2
44	RDGBKGD3	0.016781	READING BACKGROUND # 3
45	RDGBKGD4	0.097556	READING BACKGROUND # 4
46	RDGBKGD5	-0.072348	READING BACKGROUND # 5
47	RDGBKGD6	0.009887	READING BACKGROUND # 6
48	RDGBKGD7	-0.033549	READING BACKGROUND # 7
49	RDGBKGD8	0.065660	READING BACKGROUND # 8
50	RDGBKGD9	0.272768	READING BACKGROUND # 9
51	RDGBKG10	0.026865	READING BACKGROUND # 10

Table C-13

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 4/Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-1.703461	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.296941	SEX (FEMALE)
3	ETHNIC2	-0.349940	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.361704	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.232422	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.519909	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.168985	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.176519	REGION (SOUTHEAST)
9	REGION3	0.096705	REGION (CENTRAL)
10	REGION4	0.053068	REGION (WEST)
11	PARED2	0.108159	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.440773	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.407230	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.081511	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.090329	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.164800	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.306408	TV WATCHING
18	TV. 2	-0.045694	TV WATCHING
19	HOMELNG1	-0.006770	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	-0.179979	HOMEWORK (NONE ASSIGNED)
21	HW-YES	-0.187228	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	-0.004841	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.003672	PERCENT IN LUNCH PROGRAM
24	%LUNCH 2	-0.112552	MISSING LUNCH PROGRAM
25	%WHITE 1	-0.096829	WHITE MINORITY
26	%WHITE 2	-0.104274	INTEGRATED
27	AGE/GRD1	-0.636146	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	0.015431	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD3	0.190052	MODAL AGE, > MODAL GRADE
30	AGE/GRD4	-0.188559	> MODAL AGE, MODAL GRADE
31	SCH.TYPE	0.041692	PUBLIC V. PRIVATE SCHOOLS
32	H/HOMEWK	-0.260516	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
33	PRESCH	0.208281	WENT TO PRESCHOOL
34	PAR/HOME	0.040124	SINGLE/MULTIPLE PARENT (MULTIPLE)
35	MOTH/HOM	0.303563	MOTHER AT HOME (YES)
36	MOTH/WRK	0.028529	MOTHER WORK OUTSIDE HOME (YES)
37	GRWN HOM	0.088326	GROWNUP AT HOME RIGHT AFTER SCHOOL
38	PAGE/RD1	0.119129	PAGES A DAY READ (6 AND UP)
39	PAGE/RD2	0.010113	PAGES A DAY READ (11 AND UP)

Table C-14

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 8/Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-2.878516	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.095517	SEX (FEMALE)
3	ETHNIC2	-0.200724	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.205606	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.130494	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.072850	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	-0.016063	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	-0.020760	REGION (SOUTHEAST)
9	REGION3	-0.017065	REGION (CENTRAL)
10	REGION4	-0.031544	REGION (WEST)
11	PARED2	0.055548	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.208053	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.191380	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.024154	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.045925	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.073831	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.039022	TV WATCHING
18	TV. 2	-0.006148	TV WATCHING
19	HOMELNG1	-0.024874	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	0.423507	HOMEWORK (NONE ASSIGNED)
21	HW-YES	0.519192	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	-0.017066	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.000706	PERCENT IN LUNCH PROGRAM
24	%LUNCH 2	-0.088738	MISSING LUNCH PROGRAM
25	%WHITE 1	-0.047834	WHITE MINORITY
26	%WHITE 2	-0.090714	INTEGRATED
27	AGE/GRD1	-0.494470	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	-0.170739	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD3	0.067865	MODAL AGE, > MODAL GRADE
30	AGE/GRD4	-0.290966	> MODAL AGE, MODAL GRADE
31	SCH.TYPE	0.179425	PUBLIC V. PRIVATE SCHOOLS
32	H/HOMEWK	-0.109597	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
33	PAR/HOME	0.005510	SINGLE/MULTIPLE PARENT (MULTIPLE)
34	MOTH/HOM	0.134401	MOTHER AT HOME (YES)
35	MOTH/WRK	-0.054176	MOTHER WORK OUTSIDE HOME (YES)
36	PAGE/RD1	0.126464	PAGES A DAY READ (6 AND UP)
37	PAGE/RD2	0.027989	PAGES A DAY READ (11 AND UP)
38	GRADRATE	0.269446	DO YOU EXPECT TO GRADUATE FROM HIGH SCHOOL
39	DAYSMISS	0.010742	DAYS OF SCHOOL MISSED LAST MONTH
40	GRADES	0.243183	GRADES IN SCHOOL
41	RDGCOMP1	0.016368	READING COMPOSITE # 1

(continued)

Table C-14 (continued)

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 8/Age 13

	Variable	Estimated Effect	Description
42	RDGCOMP2	-0.031322	READING COMPOSITE # 2
43	RDGCOMP3	-0.026681	READING COMPOSITE # 3
44	RDGCOMP4	0.062159	READING COMPOSITE # 4
45	RDGBKGD1	0.032512	READING BACKGROUND # 1
46	RDGBKGD2	0.017270	READING BACKGROUND # 2
47	RDGBKGD3	-0.002779	READING BACKGROUND # 3
48	RDGBKGD4	-0.041347	READING BACKGROUND # 4
49	RDGBKGD5	-0.062938	READING BACKGROUND # 5
50	RDGBKGD6	0.013374	READING BACKGROUND # 6
51	RDGBKGD7	-0.012108	READING BACKGROUND # 7
52	RDGBKGD8	0.036475	READING BACKGROUND # 8
53	RDGBKGD9	0.067985	READING BACKGROUND # 9
54	RDGBKG10	0.005173	READING BACKGROUND # 10
55	RDGBKG11	0.271575	READING BACKGROUND # 11
56	RDGBKG12	0.172633	READING BACKGROUND # 12
57	RDGBKG13	-0.037839	READING BACKGROUND # 13
58	RDGBKG14	-0.006027	READING BACKGROUND # 14
59	RDGBKG15	0.017113	READING BACKGROUND # 15
60	RDGBKG16	0.049804	READING BACKGROUND # 16
61	RDGBKG17	-0.007475	READING BACKGROUND # 17
62	RDGBKG18	-0.014529	READING BACKGROUND # 18
63	RDGBKG19	0.028883	READING BACKGROUND # 19
64	RDGBKG20	0.101803	READING BACKGROUND # 20

Table C-15

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 8/Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-2.030369	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.166260	SEX (FEMALE)
3	ETHNIC2	-0.198242	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.195727	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.066718	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.159133	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.131332	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.004577	REGION (SOUTHEAST)
9	REGION3	0.070446	REGION (CENTRAL)
10	REGION4	-0.012584	REGION (WEST)
11	PARED2	-0.030636	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.060334	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.059461	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.214610	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.059626	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.158479	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.027980	TV WATCHING
18	TV. 2	-0.010843	TV WATCHING
19	HOMELNG1	0.011643	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	0.199523	HOMEWORK (NONE ASSIGNED)
21	HW-YES	0.443435	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	0.003335	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.000250	PERCENT IN LUNCH PROGRAM
24	%LUNCH 2	0.024631	MISSING LUNCH PROGRAM
25	%WHITE 1	0.018448	WHITE MINORITY
26	%WHITE 2	-0.056145	INTEGRATED
27	AGE/GRD1	-0.155339	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	0.140843	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD3	0.085695	MODAL AGE, > MODAL GRADE
30	AGE/GRD4	0.016198	> MODAL AGE, MODAL GRADE
31	SCH.TYPE	0.219716	PUBLIC V. PRIVATE SCHOOLS
32	H/HOMEWK	-0.110421	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
33	PAR/HOME	-0.073100	SINGLE/MULTIPLE PARENT (MULTIPLE)
34	MOTH/HOM	0.123323	MOTHER AT HOME (YES)
35	MOTH/WRK	-0.046198	MOTHER WORK OUTSIDE HOME (YES)
36	PAGE/RD1	0.101266	PAGES A DAY READ (6 AND UP)
37	PAGE/RD2	-0.020843	PAGES A DAY READ (11 AND UP)
38	GRADRATE	0.351427	DO YOU EXPECT TO GRADUATE FROM HIGH SCHOOL
39	DAYSMISS	0.054705	DAYS OF SCHOOL MISSED LAST MONTH
40	GRADES	0.330826	GRADES IN SCHOOL

Table C-16

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-1.588959	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.062483	SEX (FEMALE)
3	ETHNIC2	-0.265308	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.163774	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.043188	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.125134	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.090530	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.074697	REGION (SOUTHEAST)
9	REGION3	0.067739	REGION (CENTRAL)
10	REGION4	0.090882	REGION (WEST)
11	PARED2	-0.057941	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.018846	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.005209	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.133514	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.119471	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.109873	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.069582	TV WATCHING
18	TV. 2	-0.013791	TV WATCHING
19	HOMELNG1	-0.063237	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	-0.180651	HOMEWORK (NONE ASSIGNED)
21	HW-YES	-0.035056	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	-0.010146	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	0.002396	PERCENT IN LUNCH PROGRAM
24	%LUNCH 2	-0.018666	MISSING LUNCH PROGRAM
25	%WHITE 1	0 [illegible]01429	WHITE MINORITY
26	%WHITE 2	0.014251	INTEGRATED
27	AGE/GRD1	-0.329255	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	-0.207815	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD4	-0.337198	> MODAL AGE, MODAL GRADE
30	SCH.TYPE	0.019471	PUBLIC V. PRIVATE SCHOOLS
31	H/HOMEWK	-0.180211	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
32	PAR/HOME	-0.018235	SINGLE/MULTIPLE PARENT (MULTIPLE)
33	MOTH/HOM	0.035154	MOTHER AT HOME (YES)
34	MOTH/WRK	0.000284	MOTHER WORK OUTSIDE HOME (YES)
35	PAGE/RD1	0.082686	PAGES A DAY READ (6 AND UP)
36	PAGE/RD2	0.019541	PAGES A DAY READ (11 AND UP)
37	DAYSMISS	0.041734	DAYS OF SCHOOL MISSED LAST MONTH
38	GRADES	0.233978	GRADES IN SCHOOL
39	HS PGM2	0.090736	HIGH SCHOOL PROGRAM (COLLEGE PREPARATORY)
40	HS PGM3	-0.094364	HIGH SCHOOL PROGRAM (VOCATIONAL, TECHNICAL)
41	POSTSEC2	-0.00797[illegible]	POST-SECONDARY PLANS (TWO-YEAR COLLEGE)

(continued)

Table C-16 (continued)

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
42	POSTSEC3	0.136958	POST-SECONDARY PLANS (FOUR-YEAR COLLEGE)
43	WORKHOUR	-0.001018	HOURS OF OUTSIDE WORK
44	ENG.23	0.053846	TYPE OF ENGLISH CLASS(ADV. PLACEMENT&COLLEGE)
45	ENGLISH5	-0.224120	TYPE OF ENGLISH CLASS(REMEDIAL)
46	RDGCOMP1	-0.001806	READING COMPOSITE # 1
47	RDGCOMP2	-0.017579	READING COMPOSITE # 2
48	RDGCOMP3	-0.039289	READING COMPOSITE # 3
49	RDGCOMP4	0.071240	READING COMPOSITE # 4
50	RDGCOMP5	0.001928	READING COMPOSITE # 5
51	RDGCOMP6	-0.047195	READING COMPOSITE # 6
52	RDGCOMP7	-0.032501	READING COMPOSITE # 7
53	RDGBKGD1	0.007427	READING BACKGROUND # 1
54	RDGBKGD2	0.020779	READING BACKGROUND # 2
55	RDGBKGD3	-0.002569	READING BACKGROUND # 3
56	RDGBKGD4	-0.031797	READING BACKGROUND # 4
57	RDGBKGD5	-0.021215	READING BACKGROUND # 5
58	RDGBKGD6	-0.006863	READING BACKGROUND # 6
59	RDGBKGD7	-0.002056	READING BACKGROUND # 7
60	RDGBKGD8	0.002351	READING BACKGROUND # 8
61	RDGBKGD9	0.088956	READING BACKGROUND # 9
62	RDGBKG10	0.043146	READING BACKGROUND # 10
63	RDGBKG11	-0.064605	READING BACKGROUND # 11
64	RDGBKG12	-0.040126	READING BACKGROUND # 12
65	RDGBKG13	0.003461	READING BACKGROUND # 13
66	RDGBKG14	0.195865	READING BACKGROUND # 14
67	RDGBKG15	0.142466	READING BACKGROUND # 15
68	RDGBKG16	-0.006831	READING BACKGROUND # 16
69	RDGBKG17	0.031799	READING BACKGROUND # 17
70	RDGBKG18	-0.050291	READING BACKGROUND # 18
71	RDGBKG19	-0.022255	READING BACKGROUND # 19
72	RDGBKG20	0.058058	READING BACKGROUND # 20
73	RDGBKG11	0.011322	READING BACKGROUND # 21
74	RDGBKG12	-0.027553	READING BACKGROUND # 22
75	RDGBKG13	0.084234	READING BACKGROUND # 23
76	RDGBKG14	-0.024406	READING BACKGROUND # 24
77	RDGBKG15	0.009928	READING BACKGROUND # 25
78	RDGBKG16	0.008416	READING BACKGROUND # 26
79	RDGBKG17	-0.050294	READING BACKGROUND # 27
80	RDGBKG18	0.046429	READING BACKGROUND # 28
81	RDGBKG19	0.000222	READING BACKGROUND # 29
82	RDGBKG20	0.005741	READING BACKGROUND # 30
83	RDGBKG20	0.051695	READING BACKGROUND # 31
84	RDGBKG20	0.095686	READING BACKGROUND # 32

Table C-17

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-0.285641	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.162409	SEX (FEMALE)
3	ETHNIC2	-0.251546	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.329229	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.261376	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.074295	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.041690	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.064445	REGION (SOUTHEAST)
9	REGION3	0.165534	REGION (CENTRAL)
10	REGION4	0.082071	REGION (WEST)
11	PARED2	-0.046125	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	-0.076116	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	-0.022725	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.279192	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.132520	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.215761	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.056234	TV WATCHING
18	TV. 2	-0.011638	TV WATCHING
19	HOMELNG1	-0.004110	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	-0.377706	HOMEWORK (NONE ASSIGNED)
21	HW-YES	-0.186437	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	0.045292	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	0.000698	PERCENT IN LUNCH PROGRAM
24	%LUNCH 2	-0.054536	MISSING LUNCH PROGRAM
25	%WHITE 1	-0.102933	WHITE MINORITY
26	%WHITE 2	0.040726	INTEGRATED
27	AGE/GRD1	-0.457217	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	-0.249898	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD4	-0.502886	> MODAL AGE, MODAL GRADE
30	SCH.TYPE	-0.044550	PUBLIC V. PRIVATE SCHOOLS
31	H/HOMEWK	-0.220694	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
32	PAR/HOME	-0.041477	SINGLE/MULTIPLE PARENT (MULTIPLE)
33	MOTH/HOM	0.010448	MOTHER AT HOME (YES)
34	MOTH/WRK	0.001454	MOTHER WORK OUTSIDE HOME (YES)
35	PAGE/RD1	0.217577	PAGES A DAY READ (6 AND UP)
36	PAGE/RD2	-0.051548	PAGES A DAY READ (11 AND UP)
37	DAYSMISS	-0.027542	DAYS OF SCHOOL MISSED LAST MONTH
38	GRADES	0.283600	GRADES IN SCHOOL
39	HS PGM2	0.133174	HIGH SCHOOL PROGRAM (COLLEGE PREPARATORY)
40	HS PGM3	0.015758	HIGH SCHOOL PROGRAM (VOCATIONAL, TECHNICAL)
41	POSTSEC2	0.012074	POST-SECONDARY PLANS (TWO-YEAR COLLEGE)

(continued)

Table C-17 (continued)

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
42	POSTSEC3	0.181729	POST-SECONDARY PLANS (FOUR-YEAR COLLEGE)
43	WORKHOUR	-0.001030	HOURS OF OUTSIDE WORK
44	ENG.23	0.103382	TYPE OF ENGLISH CLASS(ADV. PLACEMENT&COLLEGE)
45	ENGLISH5	-0.112947	TYPE OF ENGLISH CLASS(REMEDIAL)

Table C-18

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Grade 4 Students with Surveyed Teachers

	Variable	Estimated Effect	Description
1	OVERALL	-1.643733	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.054165	SEX (FEMALE)
3	ETHNIC2	-0.170767	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.164307	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.065474	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.254828	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.097182	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.113317	REGION (SOUTHEAST)
9	REGION3	0.030617	REGION (CENTRAL)
10	REGION4	0.030780	REGION (WEST)
11	PARED2	0.229261	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.295141	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.266411	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.210889	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.056898	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.043555	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.122396	TV WATCHING
18	TV. 2	-0.019522	TV WATCHING
19	HOMELNG1	-0.039185	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	0.356544	HOMEWORK (NONE ASSIGNED)
21	HW-YES	0.330480	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	0.012380	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.001171	PERCENT IN LUNCH PROGRAM
24	%LUNCH 2	-0.170076	MISSING LUNCH PROGRAM
25	%WHITE 1	-0.013090	WHITE MINORITY
26	%WHITE 2	-0.038655	INTEGRATED
27	MA,MG	-0.149529	MODAL AGE, MODAL GRADE, MISSING
28	>MA,MG	-0.221906	> MODAL AGE, MODAL GRADE
29	SCH.TYPE	0.108642	PUBLIC V. PRIVATE SCHOOLS
30	H/HOMEWK	-0.033086	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
31	PRESCH	0.078802	WENT TO PRESCHOOL
32	PAR/HOME	-0.000269	SINGLE/MULTIPLE PARENT (MULTIPLE)
33	MOTH/HOM	0.152510	MOTHER AT HOME (YES)
34	MOTH/WRK	-0.028558	MOTHER WORK OUTSIDE HOME (YES)
35	GRWN HOM	0.062530	GROWNUP AT HOME RIGHT AFTER SCHOOL
36	PAGE/RD1	0.089515	PAGES A DAY READ (6 AND UP)
37	PAGE/RD2	-0.018468	PAGES A DAY READ (11 AND UP)
38	RDGCOMP1	0.066655	READING COMPOSITE # 1
39	RDGCOMP2	-0.003565	READING COMPOSITE # 1
40	RDGBKGD1	0.056316	READING BACKGROUND # 1
41	RDGBKGD2	-0.010363	READING BACKGROUND # 2

(continued)

Table C-18 (continued)

Estimated Effects for Reading Cross-sectional Conditioning Variables
1988 Grade 4 Students with Surveyed Teachers

	Variable	Estimated Effect	Description
42	RDGBKGD3	0.011394	READING BACKGROUND # 3
43	RDGBKGD4	0.033141	READING BACKGROUND # 4
44	RDGBKGD5	-0.048777	READING BACKGROUND # 5
45	RDGBKGD6	0.006944	READING BACKGROUND # 6
46	RDGBKGD7	-0.021022	READING BACKGROUND # 7
47	RDGBKGD8	0.021402	READING BACKGROUND # 8
48	RDGBKGD9	0.138259	READING BACKGROUND # 9
49	RDGBKG10	0.019127	READING BACKGROUND # 10
50	TQS1	-0.366611	READING LEVEL OF STUDENT
51	TQS3	0.183620	HIGH,AVERAGE,LOW READING LEVELS
52	TQS3_	-0.113201	HIGH,AVERAGE,LOW READING LEVELS(1,2,3)
53	TQS4	0.045659	HELP STUDENT WITH READING ALOUD
54	TQS5	-0.040390	HELP STUDENT WITH COMPREHENSIVE SKILLS
55	TQS6	0.024487	HELP STUDENT WITH WORD-ATTACK SKILLS
56	TQS7	0.004351	HELP STUDENT WITH VOCABULARY
57	TQS10	0.256632	DOES STUDENT RECEIVE REMEDIAL READING PROG
58	TQC1	-0.016712	ARE STUDENTS ASSIGNED TO CLASS BY ABILITY
59	TQC2	-0.042102	READING ABILITY LEVEL OF STUDENT IN CLASS
60	TQC3	-0.011451	TIME SPENT ON TYPICAL DAY FOR DIRECT INSTRUCT.
61	TQC4AYES	-0.069730	HAVE READING INSTRUCTION - INDIVIDUAL
62	TQC4AY2	-0.007559	HAVE READING INSTRUCTION - INDIVIDUAL(1,2,3,4)
63	TQC4BYES	-0.042395	HAVE READING INSTRUCTION - SMALL GROUP
64	TQC4BY2	0.021101	HAVE READING INSTRUCTION - SM. GROUP(1,2,3,4)
65	TQC4CYES	0.022854	HAVE READING INSTRUCTION - WHOLE CLASS
66	TQC4CY2	0.015939	HAVE READING INSTRUCTION - WH. CLASS(1,2,3,4)
67	TQC6	-0.023527	ARE STUDENTS ASSIGNED TO GROUPS BY ABILITY
68	TQC7-YES	-0.048101	USE SAME READING BOOK FOR ALL IN CLASS
69	TQC7-2	0.022808	USE DIFFERENT LEVELS WITHIN SAME BASAL.
70	TQC7-3	0.089372	USE DIFFERENT BASAL SERIES
71	TQC8A-8E	0.003570	NUMBER OF RESOURCES
72	TQC10	0.014164	HOW OFTEN - READ ALOUD TO STUDENTS
73	TQC11	0.000255	HOW OFTEN - HAVE STUDENTS MEET IN SMALL GROUPS
74	TQC12	-0.004808	HOW OFTEN - HAVE STUD. WRITE ABOUT SOMETHING
75	TQC13	0.029973	HOW OFTEN - HAVE STUDENTS COMPLETE WORKBOOKS
76	TQC14	0.010835	HOW OFTEN - HAVE STUD. READ BOOKS THEY CHOOSE
77	TQC15	0.018556	HOW OFTEN - HAVE STUD. READ INFORM. MATERIAL
78	TQC16	-0.030612	HOW OFTEN - TAKE CLASS TO LIBRARY
79	TQTB9	0.009389	HIGHEST ACADEMIC DEGREE HELD
80	TQTB12BE	0.035126	NUMBER OF SPECIAL TRAININGS
81	TQTB23	-0.001884	ABILITY TO GET DESIRED INSTRUCTIONAL MATERIALS
82	TQTB24	-0.008206	IF COULD START OVER, WOULD BECOME A TEACHER

Table C-19

Estimated Effects for Reading Trend Conditioning Variables
1988 Bridge to 1984, Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-1.105701	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.155957	SEX (FEMALE& MISSING)
3	ETHNIC2	-0.410887	ETHNICITY (BLACK)
4	ETHNIC3	-0.297601	ETHNICITY (HISPANIC)
5	ETHNIC4	0.289799	ETHNICITY (ASIAN)
6	STOC2	0.175011	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.063150	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.018628	REGION (SOUTHEAST)
9	REGION3	0.100021	REGION (CENTRAL)
10	REGION4	-0.098282	REGION (WEST)
11	PARED2	0.202365	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.194730	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.277028	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.114303	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS	0.075757	ITEMS IN HOME
16	ITEMS2	0.221712	ITEMS IN HOME (BLANK MISSING)
17	TV	-0.051474	HOURS TV WATCHING (LINEAR)
18	TV2	-0.244725	HOURS TV WATCHING (MISSING)
19	HW-YES	-0.022698	HOMEWORK (DON'T HAVE ANY & SOME AMOUNT)
20	HW	-0.086636	HOMEWORK (MISSING)
21	LM BY E3	-0.092054	LANGUAGE MINORITY BY ETHNICITY (SPAN. & OTHERS)
22	LM	-0.342136	LANGUAGE MINORITY BY ETHNICITY (MISSING)
23	PAGE1234	0.124241	PAGES READ(MORE THAN 5)
24	PAGE	-0.305198	PAGES READ(MISSING)
25	LUNCH%	-0.189146	PERCENT IN LUNCH PROGRAM (F3.2)
26	LUNCH_	0.041344	LUNCH PROGRAM (MISSING)
27	%WHITE49	-0.009154	PERCENT WHITE IN SCHOOL (0-49% WHITE MINORITY)
28	%WHITE79	0.058382	PERCENT WHITE IN SCHOOL (50-79% INTEGRATED)
29	%WHITE00	0.026256	PERCENT WHITE IN SCHOOL (80 100% PREDOMINANTLY)
30	COURSES7	0.028015	COURSES TAKEN(1-7) F4.1
31	COURSES	-0.017418	COURSES TAKEN(MISSING)
32	<MA,<MG	-0.542511	MODAL AGE, LESS THAN MODAL GRADE
33	MA,MG	-0.025866	MODAL AGE, MODAL GRADE, MISSING
34	MA,>MG	0.571595	MODAL AGE, GREATER THAN MODAL GRADE
35	>MA,MG	-0.284144	GREATER THAN MODAL ACE, MODAL GRADE

Table C-20

Estimated Effects for Reading Trend Conditioning Variables
1988 Bridge to 1984, Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-0.514684	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	0.152055	SEX (FEMALE& MISSING)
3	ETHNIC2	-0.194921	ETHNICITY (BLACK)
4	ETHNIC3	-0.171738	ETHNICITY (HISPANIC)
5	ETHNIC4	0.212954	ETHNICITY (ASIAN)
6	STOC2	0.184238	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.123749	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.069277	REGION (SOUTHEAST)
9	REGION3	0.001563	REGION (CENTRAL)
10	REGION4	0.020983	REGION (WEST)
11	PARED2	0.048948	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.214287	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.205732	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.016082	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS	0.057393	ITEMS IN HOME
16	ITEMS2	0.209762	ITEMS IN HOME (BLANK MISSING)
17	TV	-0.038498	HOURS TV WATCHING (LINEAR)
18	TV2	-0.365324	HOURS TV WATCHING (MISSING)
19	HW-YES	0.030688	HOMEWORK (DON'T HAVE ANY & SOME AMOUNT)
20	HW	-0.177569	HOMEWORK (MISSING)
21	LM BY E3	-0.087368	LANGUAGE MINORITY BY ETHNICITY (SPAN. & OTHERS)
22	LM	0.034672	LANGUAGE MINORITY BY ETHNICITY (MISSING)
23	PAGE1234	0.125767	PAGES READ(MORE THAN 5)
24	PAGE	-0.056604	PAGES READ(MISSING)
25	LUNCH%	0.114286	PERCENT IN LUNCH PROGRAM (F3.2)
26	LUNCH_	0.123994	LUNCH PROGRAM (MISSING)
27	%WHITE49	0.089640	PERCENT WHITE IN SCHOOL (0-49% WHITE MINORITY)
28	%WHITE79	0.224434	PERCENT WHITE IN SCHOOL (50-79% INTEGRATED)
29	%WHITE00	0.138460	PERCENT WHITE IN SCHOOL (80-100% PREDOMINANTLY)
30	COURSES7	0.018064	COURSES TAKEN(1-7) F4.1
31	COURSES	-0.001403	COURSES TAKEN(MISSING)
32	<MA,<MG	-0.368692	MODAL AGE, LESS THAN MODAL GRADE
33	MA,MG	-0.020205	MODAL AGE, MODAL GRADE, MISSING
34	MA,>MG	0.083748	MODAL AGE, GREATER THAN MODAL GRADE
35	>MA,MG	-0.211022	GREATER THAN MODAL AGE, MODAL GRADE

Table C-21

Estimated Effects for Reading Trend Conditioning Variables
1988 Bridge to 1984, Age 17

	Variable	Estimated Effect	Description
1	OVERALL	0.127898	OVERALL CONSTANT '1' FOR EVERYONE
2	FEMALE	0.078578	SEX (FEMALE& MISSING)
3	STOC2	-0.309676	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
4	STOC3	-0.144500	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
5	REGION2	0.008773	REGION (SOUTHEAST)
6	REGION3	0.291371	REGION (CENTRAL)
7	REGION4	0.194298	REGION (WEST)
8	PARED2	-0.020116	PARENTS EDUCATION (HIGH SCHOOL GRAD)
9	PARED3	0.011057	PARENTS EDUCATION (POST HIGH SCHOOL)
10	PARED4	-0.023586	PARENTS EDUCATION (COLLEGE GRAD)
11	PARED_	0.063957	PARENTS EDUCATION (MISSING, I DON'T KNOW)
12	ITEMS-YES	0.253708	ITEMS IN HOME
13	ITEMS_	0.262798	ITEMS IN HOME (BLANK MISSING)
14	TV	-0.086053	HOURS TV WATCHING (LINEAR)
15	TV2	0.083831	HOURS TV WATCHING (MISSING)
16	HW-YES	0.075871	HOMEWORK (DON'T HAVE ANY & SOME AMOUNT)
17	HW	-0.044334	HOMEWORK (MISSING)
18	LM23	-0.683796	LANGUAGE MINORITY BY ETHNICITY (SPAN. & OTHERS)
19	LM_	0.077812	LANGUAGE MINORITY BY ETHNICITY (MISSING)
20	PAGE1234	-0.279886	PAGES READ(MORE THAN 5)
21	PAGERD_	-0.301479	PAGES READ(MISSING)
22	LUNCH%	-0.550248	PERCENT IN LUNCH PROGRAM (F3.2)
23	LUNCH_	0.229743	LUNCH PROGRAM (MISSING)
24	%WHITE49	0.518987	PERCENT WHITE IN SCHOOL (0-49% WHITE MINORITY)
25	%WHITE79	-0.003205	PERCENT WHITE IN SCHOOL (50-79% INTEGRATED)
26	%WHITE00	-0.029302	PERCENT WHITE IN SCHOOL (80-100% PREDOMINANTLY)
27	ETHNIC2	-0.052682	ETHNICITY (BLACK)
28	ETHNIC3	-0.033477	ETHNICITY (HISPANIC)
29	ETHNIC4	-0.069511	ETHNICITY (ASIAN)
30	<MA,<MG	-0.538537	MODAL AGE, LESS THAN MODAL GRADE
31	MA,MG	-0.113926	MODAL AGE, MODAL GRADE, MISSING
32	MA,>MG	0.065530	MODAL AGE, GREATER THAN MODAL GRADE
33	>MA,MG	-0.458032	GREATER THAN MODAL AGE, MODAL GRADE

Table C-22

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 4/Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-0.862717	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.023548	SEX (FEMALE)
3	ETHNIC-B	-0.212695	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.250943	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-A	-0.151950	OBSERVED ETHNICITY (ASIAN AMERICAN)
6	STOC-H	0.223446	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.097596	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.045772	REGION (SOUTHEAST)
9	REG-C	0.024092	REGION (CENTRAL)
10	REG-W	-0.040014	REGION (WEST)
11	PAREDHS	0.010546	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.090901	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.108382	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.089029	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.110087	3 ITEMS IN THE HOME
16	ITEMS-4	0.180746	4 ITEMS IN THE HOME
17	TVWATCH	0.118004	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.018143	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	0.044591	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	-0.348922	DON'T HAVE HOMEWORK
21	HOMEWRKO	-0.367229	DON'T DO HOMEWORK
22	HOMEWORK	-0.010127	AMOUNT OF HOMEWORK DONE
23	PCTLUNCH	-0.001857	PERCENT IN LUNCH PROGRAM
24	NOLUNCH	-0.049798	DON'T HAVE LUNCH PROGRAM OR MISSING
25	PCTWHLOW	-0.062140	PCT WHITE IN SCHOOL 0-49
26	PCTWHMED	0.019981	PCT WHITE IN SCHOOL 50-79
27	=MA:<MG	-0.657937	AT MODAL AGE; BELOW MODAL GRADE
28	=MA:=MG	-0.206925	AT MODAL AGE; AT MODAL GRADE
29	=MA:>MG	0.389979	AT MODAL AGE; ABOVE MODAL GRADE
30	>MA:=MG	-0.305617	ABOVE MODAL AGE; AT MODAL GRADE
31	SCH-PRIV	0.089136	SCHOOL OTHER THAN PUBLIC
32	HOMEHELP	-0.110131	USUALLY GET HELP AT HOME WITH HOMEWORK
33	PRESCH	0.065690	WENT TO PRESCHOOL
34	HOMEPARS	0.066022	FATHER AND MOTHER BOTH AT HOME
35	HOMEMOM	0.219030	MOTHER LIVES AT HOME
36	MOMWORK	0.036292	MOTHER WORKS OUTSIDE OF HOME
37	HOMEGRO	0.037314	GROWNUP AT HOME RIGHT AFTER SCHOOL
38	RD06+PP	0.034761	READ 6+ PAGES PER DAY FOR SCHOOL
39	RD11+PP	0.065288	READ 11+ PAGES PER DAY FOR SCHOOL
40	SSEVERY	0.101641	SOCIAL STUDIES CLASS EVERY DAY
41	SS3-4	0.163122	SOCIAL STUDIES CLASS 3-4 TIMES PER WEEK

(continued)

Table C-22 (continued)

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 4/Age 9

	Variable	Estimated Effect	Description
42	SS1-2	0.093823	SOCIAL STUDIES CLASS 1-2 TIMES PER WEEK
43	SS<1	0.038207	SOCIAL STUDIES CLASS < 1 TIME PER WEEK
44	STGOV	0.012140	STUDIED GOVERNMENT
45	STGOV-X	-0.192139	STUDIED GOVERNMENT RESP. MISSING
46	ST-ALOT	-0.019732	NUMBER OF SUBJECTS STUDIED A LOT (OUT OF 6)
47	STLAW-L	-0.044790	STUDY LAWS A LOT OR SOME
48	STCRT-L	-0.026685	STUDY JUDGES & COURTS A LOT OR SOME
49	STPRES-L	-0.008562	STUDY PRES. & LEADERS A LOT OR SOME
50	STVOT-L	0.059727	STUDY VOTING & ELECTIONS A LOT OR SOME
51	STCOMM-L	0 081548	STUDY COMMUNITY A LOT OR SOME
52	STRGHT-L	0.092128	STUDY RIGHTS & RESPONSIBILITIES A LOT OR SOME
53	CE-EVERY	-0.050833	DISCUSS CURRENT EVENTS ALMOST EVERY DAY
54	CE1-2	0.049644	DISCUSS CURRENT EVENTS 1-2 TIMES A WEEK
55	CE<1	0.073888	DISCUSS CURRENT EVENTS 1-2 TIMES A MONTH
56	CE<<1	-0.022439	DISCUSS CURRENT EVENTS FEW TIMES A YEAR

Table C-23

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 4/Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-2.423182	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.029573	SEX (FEMALE)
3	ETHNIC-B	-0.231977	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.251819	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-A	0.068154	OBSERVED ETHNICITY (ASIAN AMERICAN)
6	STOC-H	0.265959	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.110538	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.110236	REGION (SOUTHEAST)
9	REG-C	0.090622	REGION (CENTRAL)
10	REG-W	-0.003717	REGION (WEST)
11	PAREDHS	0.156360	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.298932	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.272740	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	0.049526	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.052185	3 ITEMS IN THE HOME
16	ITEMS-4	0.141589	4 ITEMS IN THE HOME
17	TVWATCH	0.277411	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.037488	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	-0.021578	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	0.312557	DON'T HAVE HOMEWORK
21	HOMEWRK0	0.308734	DON'T DO HOMEWORK
22	HOMEWORK	-0.010786	AMOUNT OF HOMEWORK DONE
23	PCTLUNCH	-0.002942	PERCENT IN LUNCH PROGRAM
24	NOLUNCH	-0.108688	DON'T HAVE LUNCH PROGRAM OR MISSING
25	PCTWHLOW	-0.094621	PCT WHITE IN SCHOOL 0-49
26	PCTWHMED	-0.015439	PCT WHITE IN SCHOOL 50-79
27	-MA:<MG	-0.017183	AT MODAL AGE; BELOW MODAL GRADE
28	-MA:-MG	0.421162	AT MODAL AGE; AT MODAL GRADE
29	-MA:>MG	0.458405	AT MODAL AGE; ABOVE MODAL GRADE
30	>MA:-MG	0.351086	ABOVE MODAL AGE; AT MODAL GRADE
31	SCH-PRIV	0.001673	SCHOOL OTHER THAN PUBLIC
32	HOMEHELP	-0.104533	USUALLY GET HELP AT HOME WITH HOMEWORK
33	PRESCH	0.104314	WENT TO PRESCHOOL
34	HOMEPARS	0.033862	FATHER AND MOTHER BOTH AT HOME
35	HOMEMOM	0.110458	MOTHER LIVES AT HOME
36	MOMWORK	0.011762	MOTHER WORKS OUTSIDE OF HOME
37	HOMEGRO	0.058718	GROWNUP AT HOME RIGHT AFTER SCHOOL
38	RD06+PP	0.070232	READ 6+ PAGES PER DAY FOR SCHOOL
39	RD11+PP	0.030869	READ 11+ PAGES PER DAY FOR SCHOOL
40	SSEVERY	0.146238	SOCIAL STUDIES CLASS EVERY DAY
41	SS3-4	0.201378	SOCIAL STUDIES CLASS 3-4 TIMES PER WEEK

(continued)

Table C-23 (continued)

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 4/Age 9

	Variable	Estimated Effect	Description
42	SS1-2	0.232414	SOCIAL STUDIES CLASS 1-2 TIMES PER WEEK
43	SS<1	0.060544	SOCIAL STUDIES CLASS < 1 TIME PER WEEK
44	STGOV	0.033795	STUDIED GOVERNMENT
45	STGOV-X	-0.500008	STUDIED GOVERNMENT RESP. MISSING
46	STLAW-L	0.068436	STUDY LAWS A LOT OR SOME
47	STPRES-L	0.007628	STUDY PRES. & LEADERS A LOT OR SOME
48	STCOMM-L	-0.074725	STUDY COMMUNITY A LOT OR SOME

Table C-24

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 8/Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-1.613136	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.077074	SEX (FEMALE)
3	ETHNIC-B	-0.138794	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.183800	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-A	-0.101094	OBSERVED ETHNICITY (ASIAN AMERICAN)
6	STOC-H	0.114793	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.061629	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.025581	REGION (SOUTHEAST)
9	REG-C	0.071906	REGION (CENTRAL)
10	REG-W	0.014446	REGION (WEST)
11	PAREDHS	0.015981	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.125404	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.154213	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.081084	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.032833	3 ITEMS IN THE HOME
16	ITEMS-4	0.101106	4 ITEMS IN THE HOME
17	TVWATCH	0.046246	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.008183	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	0.007020	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	-0.061921	DON'T HAVE HOMEWORK
21	HOMEWRK0	0.026918	DON'T DO HOMEWORK
22	HOMEWORK	0.007356	AMOUNT OF HOMEWORK DONE
23	PCTLUNCH	-0.001397	PERCENT IN LUNCH PROGRAM
24	NOLUNCH	-0.090389	DON'T HAVE LUNCH PROGRAM OR MISSING
25	PCTWHLOW	-0.022594	PCT WHITE IN SCHOOL 0-49
26	PCTWHMED	-0.041061	PCT WHITE IN SCHOOL 50-79
27	=MA:<MG	-0.179729	AT MODAL AGE; BELOW MODAL GRADE
28	=MA:=MG	0.032173	AT MODAL AGE; AT MODAL GRADE
29	=MA:>MG	0.407612	AT MODAL AGE; ABOVE MODAL GRADE
30	>MA:=MG	-0.124894	ABOVE MODAL AGE; AT MODAL GRADE
31	SCH-PRIV	0.061854	SCHOOL OTHER THAN PUBLIC
32	HOMEHELP	-0.088763	USUALLY GET HELP AT HOME WITH HOMEWORK
33	HOMEPARS	0.014917	FATHER AND MOTHER BOTH AT HOME
34	HOMEMOM	0.158676	MOTHER LIVES AT HOME
35	MOMWORK	0.00681[illegible]	MOTHER WORKS OUTSIDE OF HOME
36	RD06+PP	0.105422	READ 6+ PAGES PER DAY FOR SCHOOL
37	RD11+PP	-0.011569	READ 11+ PAGES PER DAY FOR SCHOOL
38	GRADEXP	0.207459	EXPECT TO GRADUATE FROM HIGH SCHOOL
39	NOMISS	0.007670	DID NOT MISS MUCH SCHOOL LAST MONTH
40	GRADES	0.196147	GRADES IN SCHOOL
41	STCV-INC	-0.049973	STUDIED CIVICS - INCOMPLETE SET OF ANSWERS

(continued)

Table C-24 (continued)

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Focused-BIB Sample Grade 8/Age 13

	Variable	Estimated Effect	Description
42	STCIV5	-0.022543	STUDIED CIVICS IN 5TH GRADE
43	STCIV6	-0.022328	STUDIED CIVICS IN 6TH GRADE
44	STCIV7	-0.034185	STUDIED CIVICS IN 7TH GRADE
45	STCIV8	0.140325	STUDIED CIVICS IN 8TH GRADE
46	ST-ALOT	0.005388	NUMBER OF SUBJECTS STUDIED A LOT (OUT OF 10)
47	STCON-L	0.154097	STUDIED U.S. CONSTITUTION A LOT OR SOME
48	STCONG-L	0.029228	STUDIED CONGRESS A LOT OR SOME
49	STLAW-L	0.072744	STUDIED LAWS A LOT OR SOME
50	STCRT-L	0.111552	STUDIED COURTS A LOT OR SOME
51	STPRES-L	0.050512	STUDIED PRESIDENT A LOT OR SOME
52	STPOL-L	-0.011655	STUDIED POLITICAL PARTIES A LOT OR SOME
53	STSLG-L	0.000487	STUDIED STATE/LOCAL GOVS. A LOT OR SOME
54	STDG-L	-0.021626	STUDIED DEMOCRATIC GOVS. A LOT OR SOME
55	STOG-L	-0.006720	STUDIED OTHER GOVS. A LOT OR SOME
56	STRGHT-L	-0.013451	STUDIED RIGHTS & RESPONSIB. A LOT OR SOME
57	CVA-ACT	-0.027045	ACTIVE IN 4 AREAS OF CIVICS (1,2,5,8)
58	CVB-ACT	-0.0[illegible]976	ACTIVE IN 2 AREAS OF CIVICS (3, 10)
59	CVC-ACT	0.[illegible]165	ACTIVE IN 2 AREAS OF CIVICS (4, 9)
60	CVD-ACT	-0.383355	ACTIVE IN 2 AREAS OF CIVICS (6, 7)
61	CVD-NOT	0.147750	NOT VERY ACTIVE IN THESE 2 AREAS OF CIVICS
62	RDDIF	-0.227011	AT LEAST SOME DIFFICULTY IN READING OR MISSING

Table C-25

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 8/Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-1.224095	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.118153	SEX (FEMALE)
3	ETHNIC-B	-0.135472	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.193761	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-A	0.027554	OBSERVED ETHNICITY (ASIAN AMERICAN)
6	STOC-H	0.080959	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.046864	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	-0.001454	REGION (SOUTHEAST)
9	REG-C	-0.001889	REGION (CENTRAL)
10	REG-W	-0.055975	REGION (WEST)
11	PAREDHS	0.030488	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.128680	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.151727	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.068832	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.122664	3 ITEMS IN THE HOME
16	ITEMS-4	0.173642	4 ITEMS IN THE HOME
17	TVWATCH	0.020052	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.006978	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	-0.046921	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	0.094715	DON'T HAVE HOMEWORK
21	HOMEWRKO	0.205456	DON'T DO HOMEWORK
22	HOMEWORK	0.040729	AMOUNT OF HOMEWORK DONE
23	PCTLUNCH	-0.001694	PERCENT IN LUNCH PROGRAM
24	NOLUNCH	-0.048478	DON'T HAVE LUNCH PROGRAM OR MISSING
25	PCTWHLOW	0.019132	PCT WHITE IN SCHOOL 0-49
26	PCTWHMED	-0.017506	PCT WHITE IN SCHOOL 50-79
27	=MA:<MG	-0.307013	AT MODAL AGE; BELOW MODAL GRADE
28	=MA:=MG	-0.008815	AT MODAL AGE; AT MODAL GRADE
29	=MA:>MG	0.332372	AT MODAL AGE; ABOVE MODAL GRADE
30	>MA:=MG	-0.127282	ABOVE MODAL AGE; AT MODAL GRADE
31	SCH-PRIV	0.160774	SCHOOL OTHER THAN PUBLIC
32	HOMEHELP	-0.134069	USUALLY GET HELP AT HOME WITH HOMEWORK
33	HOMEPARS	-0.009036	FATHER AND MOTHER BOTH AT HOME
34	HOMEMOM	0.123144	MOTHER LIVES AT HOME
35	MOMWORK	-0.022786	MOTHER WORKS OUTSIDE OF HOME
36	RD06+PP	0.096868	READ 6+ PAGES PER DAY FOR SCHOOL
37	RD11+PP	0.028811	READ 11+ PAGES PER DAY FOR SCHOOL
38	GRADEXP	0.034207	EXPECT TO GRADUATE FROM HIGH SCHOOL
39	NOMISS	0.018810	DID NOT MISS MUCH SCHOOL LAST MONTH
40	GRADES	0.283918	GRADES IN SCHOOL
41	STCV-INC	-0.061693	STUDIED CIVICS - INCOMPLETE SET OF ANSWERS

(continued)

Table C-25 (continued)

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 8/Age 13

	Variable	Estimated Effect	Description
42	STCIV5	-0.068930	STUDIED CIVICS IN 5TH GRADE
43	STCIV6	0.011925	STUDIED CIVICS IN 6TH GRADE
44	STCIV7	-0.031176	STUDIED CIVICS IN 7TH GRADE
45	STCIV8	0.039059	STUDIED CIVICS IN 8TH GRADE
46	STCON-L	0.002768	STUDIED U.S. CONSTITUTION A LOT OR SOME
47	STCONG-L	0.094652	STUDIED CONGRESS A LOT OR SOME
48	STLAW-L	-0.051939	STUDIED LAWS A LOT OR SOME
49	STCRT-L	0.100526	STUDIED COURTS A LOT OR SOME
50	STPRES-L	-0.039286	STUDIED PRESIDENT A LOT OR SOME
51	STPOL-L	0.054987	STUDIED POLITICAL PARTIES A LOT OR SOME
52	STSLG-L	0.000008	STUDIED STATE/LOCAL GOVS. A LOT OR SOME
53	STDG-L	-0.014294	STUDIED DEMOCRATIC GOVS. A LOT OR SOME

Table C-26

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-1.329683	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.161951	SEX (FEMALE)
3	ETHNIC-B	-0.225427	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.123948	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-A	0.037675	OBSERVED ETHNICITY (ASIAN AMERICAN)
6	STOC-H	0.175198	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.061377	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	-0.004221	REGION (SOUTHEAST)
9	REG-C	0.061998	REGION (CENTRAL)
10	REG-W	-0.018295	REGION (WEST)
11	PAREDHS	0.024432	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.150551	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.129662	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.029806	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.144291	3 ITEMS IN THE HOME
16	ITEMS-4	0.150336	4 ITEMS IN THE HOME
17	TVWATCH	0.062647	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.012239	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	-0.078055	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	-0.007298	DON'T HAVE HOMEWORK
21	HOMEWRKO	0.089084	DON'T DO HOMEWORK
22	HOMEWORK	-0.001492	AMOUNT OF HOMEWORK DONE
23	PCTLUNCH	0.000302	PERCENT IN LUNCH PROGRAM
24	NOLUNCH	-0.004135	DON'T HAVE LUNCH PROGRAM OR MISSING
25	PCTWHLOW	-0.011627	PCT WHITE IN SCHOOL 0-49
26	PCTWHMED	0.012093	PCT WHITE IN SCHOOL 50-79
27	=MA:<MG	-0.089241	AT MODAL AGE; BELOW MODAL GRADE
28	=MA:=MG	-0.017791	AT MODAL AGE; AT MODAL GRADE
29	>MA:=MG	-0.066792	ABOVE MODAL AGE; AT MODAL GRADE
30	SCH-PRIV	-0.027147	SCHOOL OTHER THAN PUBLIC
31	HOMEHELP	-0.185448	USUALLY GET HELP AT HOME WITH HOMEWORK
32	HOMEPARS	-0.002427	FATHER AND MOTHER BOTH AT HOME
33	HOMEMOM	0.132333	MOTHER LIVES AT HOME
34	MOMWORK	0.021772	MOTHER WORKS OUTSIDE OF HOME
35	RD06+PP	0.059618	READ 6+ PAGES PER DAY FOR SCHOOL
36	RD11+PP	0.008460	READ 11+ PAGES PER DAY FOR SCHOOL
37	NOMISS	0.022331	DID NOT MISS MUCH SCHOOL LAST MONTH
38	GRADES	0.120805	GRADES IN SCHOOL
39	COLLPREP	0.117544	COLLEGE PREP. PROGRAM
40	VOCTECH	-0.057613	VOCATIONAL/TECH. PROGRAM
41	PLAN2YR	0.016887	PLAN TO GO TO 2-YEAR COLLEGE

(continued)

Table C-26 (continued)

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Focused-BIB Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
42	PLAN4YR	0.151381	PLAN TO GO TO 4-YEAR COLLEGE
43	WORKHRS	-0.001538	NO. HOURS OF OUTSIDE WORK
44	ENG-AP	0.097595	TAKING A.P./COLL. PREP. ENGLISH CLASS
45	ENG-REM	-0.195803	TAKING REMEDIAL ENGLISH CLASS
46	STCV-INC	-0.148636	STUDIED CIVICS - INCOMPLETE SET OF ANSWERS
47	STCV9	-0.063248	STUDIED CIVICS IN 9TH GRADE
48	STCV10	-0.089337	STUDIED CIVICS IN 10TH GRADE
49	STCV11	-0.071126	STUDIED CIVICS IN 11TH GRADE
50	STCV12	0.037421	STUDIED CIVICS IN 12TH GRADE
51	GOVT	0.044203	HOW MANY YEARS OF GOV'T STUDY HAVE YOU HAD
52	AP	-0.109859	ADVANCED PLACEMENT YES
53	ST-ALOT	0.000510	NUMBER OF SUBJECTS STUDIED A LOT (OUT OF 10)
54	ST-AVE	0.364440	AVERAGE LEVEL OF STUDY FOR THESE TEN
55	CIV-INT	0.067754	LEVEL OF INTEREST IN CIVICS
56	MOCK-SOM	0.083116	MOCK ELECTIONS, TRIALS, ONCE TO SEVERAL TIMES
57	GOVATTN	0.110814	DEGREE OF ATTENTN GOVT. PAYS (AVE. OF 2 ITEMS)
58	HADCIV	-0.025435	YES, HAVE HAD A CIVICS CLASS
59	CVHWORK	0.032976	AMOUNT OF CIVICS HOMEWORK
60	CVGRADES	0.121410	CIVICS GRADES IN SCHOOL
61	CVA-ACT	-0.045429	ACTIVE IN 4 AREAS OF CIVICS (1,2,5,8)
62	CVB-ACT	-0.069698	ACTIVE IN 2 AREAS OF CIVICS (3, 10)
63	CVC-ACT	0.089997	ACTIVE IN 2 AREAS OF CIVICS (4, 9)
64	CVD-ACT	-0.193582	ACTIVE IN 2 AREAS OF CIVICS (6, 7)
65	CVD-NOT	0.071695	NOT VERY ACTIVE IN THESE 2 AREAS OF CIVICS
66	RDDIF	-0.170134	AT LEAST SOME DIFFICULTY IN READING OR MISSING

Table C-27

Estimated Effects for Civics Cross sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-0.911554	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.155584	SEX (FEMALE)
3	ETHNIC-B	-0.265358	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.301223	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-A	-0.5[illegible]3830	OBSERVED ETHNICITY (ASIAN AMERICAN)
6	STOC-H	0.097614	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.067059	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.091114	REGION (SOUTHEAST)
9	REG-C	0.021948	REGION (CENTRAL)
10	REG-W	0.006807	REGION (WEST)
11	PAREDHS	-0.134283	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	-0.069501	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	-0.015859	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.136091	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.0724[illegible]1	3 ITEMS IN THE HOME
16	ITEMS 4	0.104143	4 ITEMS IN THE HOME
17	TVWATCH	0.140501	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.021632	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	0.097374	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	0.148167	DON'T HAVE HOMEWORK
21	HOMEWRKO	0.185228	DON'T DO HOMEWORK
22	HOMEWORK	0.066495	AMOUNT OF HOMEWORK DONE
23	PCTLUNCH	0.003626	PERCENT [illegible] LUNCH PROGRAM
24	NOLUNCH	0.021832	DON'T HA[illegible] LUNCH PROGRAM OR MISSING
25	PCTWHLOW	-0.145977	PCT WHITE IN SCHOOL 0-49
26	PCTWHMED	-0.011403	PCT WHITE IN SCHOOL 50-79
27	=MA:<MG	-0.038936	AT MODAL AGE; BELOW MODAL GRADE
28	=MA:=MG	0.138[illegible]53	AT MODAL AGE; AT MODAL GRADE
29	>MA:=MG	0.098470	ABOVE MODAL AGE; AT MODAL GRADE
30	SCH-PRIV	0.081950	SCHOOL OTHER THAN PUBLIC
31	HOMEHELP	-0.183083	USUALLY GET HELP AT HOME WITH HOMEWORK
32	HOMEPARS	0.071978	FATHER AND MOTHER BOTH AT HOME
33	HOMEMOM	0.081156	MOTHER LIVES AT HOME
34	MOMWORK	0.026910	MOTHER WORKS OUTSIDE OF HOME
35	RD05+PP	0.145215	READ 6+ PAGES PER DAY FOR SCHOOL
36	RD11+PP	-0.066205	READ 11+ PAGES PER DAY FOR SCHOOL
37	NOMISS	0.040696	DID NOT MISS MUCH SCHOOL LAST MONTH
38	GRADES	0.234408	GRADES IN SCHOOL
39	COLLPREP	0.118036	COLLEGE PREP. PROGRAM
40	VOCTECH	-0.064741	VOCATIONAL/TECH. PROGRAM
41	PLAN2YR	0.059541	PLAN TO GO TO 2-YEAR COLLEGE

(continued)

Table C-27 (continued)

Estimated Effects for Civics Cross-sectional Conditioning Variables
1988 Intercorrelation Sample, Grade 12/Age 17

	Variable	Estimated Effect	Description
42	PLAN4YR	0.165027	PLAN TO GO TO 4-YEAR COLLEGE
43	WORKHRS	-0.002365	NO. HOURS OF OUTSIDE WORK
44	ENG-AP	0.147936	TAKING A.P./COLL. PREP. ENGLISH CLASS
45	ENG-REM	-0.085238	TAKING REMEDIAL ENGLISH CLASS
46	STCV-INC	-0.292528	STUDIED CIVICS - INCOMPLETE SET OF ANSWERS
47	STCV9	0.012856	STUDIED CIVICS IN 9TH GRADE
48	STCV10	-0.268831	STUDIED CIVICS IN 10TH GRADE
49	STCV11	-0.083368	STUDIED CIVICS IN 11TH GRADE
50	STCV12	0.102932	STUDIED CIVICS IN 12TH GRADE
51	GOVT	0.064298	HOW MANY YEARS OF GOV'T STUDY HAVE YOU HAD
52	AP	-0.070794	ADVANCED PLACEMENT YES
53	ST-ALOT	0.018307	NUMBER OF SUBJECTS STUDIED A LOT (OUT OF 10)
54	ST-AVG	-0.000833	AVERAGE LEVEL OF STUDY FOR THESE TEN

Table C-28

Estimated Effects for Civics Trend Conditioning Variables
1988 Trend Sample, Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-1.833217	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.274014	SEX (FEMALE)
3	ETHNIC-B	-0.245560	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.193642	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-O	0.022923	OBSERVED ETHNICITY (OTHER)
6	STOC-H	0.251747	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-N	0.228720	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.024840	REGION (SOUTHEAST)
9	REG-C	0.065093	REGION (CENTRAL)
10	REG-W	-0.209815	REGION (WEST)
11	PAREDHS	-0.016727	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.163282	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.287076	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.279247	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.093471	3 ITEMS IN THE HOME
16	ITEMS-4	0.197622	4 ITEMS IN THE HOME
17	TVWATCH	0.081961	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.016805	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	-0.054940	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	-0.290030	DON'T HAVE HOMEWORK
21	HOMEWRKD	-0.013951	DON'T DO HOMEWORK
22	HOMEWORK	0.049850	AMOUNT OF HOMEWORK DONE
23	PCTWHLOW	-0.192148	PCT WHITE IN SCHOOL 0-49
24	PCTWHMED	0.157684	PCT WHITE IN SCHOOL 50-79
25	= MODALG	0.575586	GRADE IN SCHOOL IS MODAL GRADE
26	> MODALG	1.748053	GRADE IN SCHOOL IS GREATER THAN MODAL GRADE
27	SCH-PRIV	0.213468	SCHOOL OTHER THAN PUBLIC
28	GRADES	0.426167	GRADES IN SCHOOL (4-POINT SCALE)

Table C-29

Estimated Effects for Civics Trend Conditioning Variables
1982 Trend Sample, Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-1.264436	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.321234	SEX (FEMALE)
3	ETHNIC-B	-0.533501	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.215298	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-O	-0.108223	OBSERVED ETHNICITY (OTHER)
6	STOC-H	0.247733	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	-0.027729	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	-0.065470	REGION (SOUTHEAST)
9	REG-C	-0.012846	REGION (CENTRAL)
10	REG-W	-0.045416	REGION (WEST)
11	PAREDHS	0.225191	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.409932	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.466953	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.052883	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.138441	3 ITEMS IN THE HOME
16	ITEMS-4	0.274918	4 ITEMS IN THE HOME
17	TVWATCH	0.004248	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.002819	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	0.097368	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	-0.311352	DON'T HAVE HOMEWORK
21	HOMEWRKO	-0.207615	DON'T DO HOMEWORK
22	HOMEWORK	0.005725	AMOUNT OF HOMEWORK DONE
23	PCTWHLOW	-0.299501	PCT WHITE IN SCHOOL 0-49
24	PCTWHMED	-0.036163	PCT WHITE IN SCHOOL 50-79
25	= MODALG	0.463258	GRADE IN SCHOOL IS MODAL GRADE
26	> MODALG	1.302472	GRADE IN SCHOOL IS GREATER THAN MODAL GRADE
27	SCH-PRIV	0.183278	SCHOOL OTHER THAN PUBLIC
28	GRADES	0.339169	GRADES IN SCHOOL (4-POINT SCALE)

Table C-30

Estimated Effects for Civics Trend Conditioning Variables
1976 Trend Sample, Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-0.865586	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.180073	SEX (FEMALE)
3	ETHNIC-B	-0.612380	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.553508	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-O	-0.394742	OBSERVED ETHNICITY (OTHER)
6	STOC-H	0.426680	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.019128	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.010113	REGION (SOUTHEAST)
9	REG-C	0.139607	REGION (CENTRAL)
10	REG-W	-0.032758	REGION (WEST)
11	PAREDHS	0.235653	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.560744	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDM	-0.003951	PARENTS EDUCATION (MISSING, I DON'T KNOW)
14	ITEMS-3	0.105100	3 ITEMS IN THE HOME
15	ITEMS-4	0.413272	4 ITEMS IN THE HOME
16	PCTWHLOW	-0.040969	PCT WHITE IN SCHOOL 0-49
17	PCTWHMED	0.023708	PCT WHITE IN SCHOOL 50-79
18	= MODALG	0.584018	GRADE IN SCHOOL IS MODAL GRADE
19	> MODALG	0.918076	GRADE IN SCHOOL IS GREATER THAN MODAL GRADE

Table C-31

Estimated Effects for Civics Trend onditioning Variables
1988 Trend Sample, Age 17

	Variable	Estimated Effect	Description
1	OVERALL	2.537850	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.354324	SEX (FEMALE)
3	ETHNIC-B	-0.371432	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.072776	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-O	-0.156620	OBSERVED ETHNICITY (OTHER)
6	STOC-H	0.217207	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.187236	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.083641	REGION (SOUTHEAST)
9	REG-C	-0.002988	REGION (CENTRAL)
10	REG-W	-0.226426	REGION (WEST)
11	PAREDHS	0.092559	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.315752	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.377825	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.029314	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.093015	3 ITEMS IN THE HOME
16	ITEMS-4	0.108895	4 ITEMS IN THE HOME
17	TVWATCH	0.015740	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.004894	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	-0.103856	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	-0.189856	DON'T HAVE HOMEWORK
21	HOMEWRKO	0.229131	DON'T DO HOMEWORK
22	HOMEWORK	-0.094754	AMOUNT OF HOMEWORK DONE
23	PCTWHLOW	0.035583	PCT WHITE IN SCHOOL 0-49
24	PCTWHMED	-0.013220	PCT WHITE IN SCHOOL 50-79
25	= MODALG	0.406556	GRADE IN SCHOOL IS MODAL GRADE
26	> MODALG	0.550452	GRADE IN SCHOOL IS GREATER THAN MODAL GRADE
27	SCH-PRIV	-0.002692	SCHOOL OTHER THAN PUBLIC
28	GRADES	0.407650	GRADES IN SCHOOL (4-POINT SCALE)
29	HSPGM-C	0.355545	HIGH SCH PROGRAM COLLEGE PREP
30	HSPGM-V	-0.148816	HIGH SCH PROGRAM VOCAT'L OR OTHER NONGENERAL
31	ISS-YES	0.150710	DISCUSS ISSUES IN COURSES
32	ISS-SOMX	0.141650	DISCUSS ISSUES IN COURSES SOMEWHAT OR MISSING
33	POL-YES	0.394967	DISCUSS POLITICS IN COURSES
34	POL-SOMX	0.259171	DISCUSS POL. IN COURSES SOMEWHAT OR MISSING
35	NONEW-Y	0.017690	NOTHING NEW DISCUSSED IN COURSES - TRUE
36	NONEW-SX	-0.005086	NOTH. NEW DISC. IN CRSES - SOMEWHAT OR MISSING
37	KNOW-YES	-0.015209	KNOWLEDGE TO PARTICIPATE - TRUE
38	KNOW-SX	-0.009888	KNOWLEDGE TO PARTICIPATE - SOMEWHAT OR MISSING
39	NATL-D	0.134013	NATIONAL DISCUSSIONS ALMOST EVERY DAY
40	NATL-C	-0.036763	NAT'L DISCUSSIONS 1-2 TIMES A WEEK OR MISSING
41	NATL-B	-0.050476	NATIONAL DISCUSSIONS 3-4 TIMES A MONTH

(continued)

Table C-31 (continued)

Estimated Effects for Civics Trend Conditioning Variables
1988 Trend Sample, Age 17

	Variable	Estimated Effect	Description
42	INTL-D	0.221026	INTERNATIONAL DISCUSSIONS ALMOST EVERY DAY
43	INTL-C	0.152830	INT'L DISCUSSIONS 1-2 TIMES A WEEK OR MISSING
44	INTL-B	0.181321	INTERNATIONAL DISCUSSIONS 3-4 TIMES A MONTH

Table C-32

Estimated Effects for Civics Trend Conditioning Variables
1982 Trend Sample, Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-1.361751	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.425031	SEX (FEMALE)
3	ETHNIC-B	-0.470033	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.296556	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-O	-0.421034	OBSERVED ETHNICITY (OTHER)
6	STOC-H	0.446878	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.203226	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	0.034580	REGION (SOUTHEAST)
9	REG-C	0.050915	REGION (CENTRAL)
10	REG-W	-0.013089	REGION (WEST)
11	PAREDHS	0.116176	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.267456	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDC	0.283765	PARENTS EDUCATION (COLLEGE GRADUATE)
14	PAREDM	-0.334944	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	ITEMS-3	0.165990	3 ITEMS IN THE HOME
16	ITEMS-4	0.275573	4 ITEMS IN THE HOME
17	TVWATCH	0.018188	NO. HOURS OF TV WATCHED PER DAY
18	TVWATCH2	-0.008675	NO. HOURS SQUARED OF TV WATCHED PER DAY
19	HOMELANG	-0.135970	MINORITY HOME LANGUAGE SPOKEN
20	HOMEWRK-	-0.302385	DON'T HAVE HOMEWORK
21	HOMEWRKO	-0.090445	DON'T DO HOMEWORK
22	HOMEWORK	0.004198	AMOUNT OF HOMEWORK DONE
23	PCTWHLOW	-0.090147	PCT WHITE IN SCHOOL 0-49
24	PCTWHMED	-0.044514	PCT WHITE IN SCHOOL 50-79
25	= MODALG	0.497047	GRADE IN SCHOOL IS MODAL GRADE
26	> MODALG	0.662263	GRADE IN SCHOOL IS GREATER THAN MODAL GRADE
27	SCH-PRIV	-0.140064	SCHOOL OTHER THAN PUBLIC
28	GRADES	0.283818	GRADES IN SCHOOL (4-POINT SCALE)
29	HSPGM-C	0.403098	HIGH SCH PROGRAM COLLEGE PREP
30	HSPGM-V	-0.009976	HIGH SCH PROGRAM VOCAT'L OR OTHER NONGENERAL

Table C-33

Estimated Effects for Civics Trend Conditioning Variables
1976 Trend Sample, Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-1.564018	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER-F	-0.439323	SEX (FEMALE)
3	ETHNIC-B	-0.375957	OBSERVED ETHNICITY (BLACK)
4	ETHNIC-H	-0.411097	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC-O	-0.351636	OBSERVED ETHNICITY (OTHER)
6	STOC-H	0.193189	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC-X	0.033124	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REG-SE	-0.062455	REGION (SOUTHEAST)
9	REG-C	0.059840	REGION (CENTRAL)
10	REG-W	-0.024240	REGION (WEST)
11	PAREDHS	-0.005646	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PAREDHS+	0.215860	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PAREDM	-0.015286	PARENTS EDUCATION (MISSING, I DON'T KNOW)
14	ITEMS-3	0.226880	3 ITEMS IN THE HOME
15	ITEMS-4	0.358708	4 ITEMS IN THE HOME
16	TVWATCH	0.003012	NO. HOURS OF TV WATCHED PER DAY
17	TVWATCH2	-0.004743	NO. HOURS SQUARED OF TV WATCHED PER DAY
18	HOMELANG	-0.087508	MINORITY HOME LANGUAGE SPOKEN
19	HOMEWRK-	-0.543340	DON'T HAVE HOMEWORK
20	HOMEWRKO	-0.333045	DON'T DO HOMEWORK
21	HOMEWORK	0.046222	AMOUNT OF HOMEWORK DONE
22	PCTWHLOW	-0.092577	PCT WHITE IN SCHOOL 0-49
23	PCTWHMED	-0.005133	PCT WHITE IN SCHOOL 50-79
24	= MODALG	0.402627	GRADE IN SCHOOL IS MODAL GRADE
25	> MODALG	0.624665	GRADE IN SCHOOL IS GREATER THAN MODAL GRADE
26	GRADES	0.424681	GRADES IN SCHOOL (4-POINT SCALE)
27	HSPGM-C	0.270935	HIGH SCH PROGRAM COLLEGE PREP
28	HSPGM-V	- .201353	HIGH SCH PROGRAM VOCAT'L OR OTHER NONGENERAL
29	ISS-YES	-0.022436	DISCUSS ISSUES IN COURSES
30	ISS-SOMX	0.010048	DISCUSS ISSUES IN COURSES SOMEWHAT OR MISSING
31	POL-YES	0.140908	DISCUSS POLITICS IN COURSES
32	POL-SOMX	0.086860	DISCUSS POLITICS IN CRSES SOMEWHAT OR MISSING
33	NONEW-Y	0.021515	NOTHING NEW DISCUSSED IN COURSES - TRUE
34	NONEW-SX	0.066948	NOTH. NEW DISC. IN CRSES - SOMEWHAT OR MISSING
35	KNOW-YES	-0.184577	KNOWLEDGE TO PARTICIPATE - TRUE
36	KNOW-SX	-0.027103	KNOWLEDGE TO PARTICIPATE - SOMEWHAT OR MISSING
37	NATL-D	0.212025	NATIONAL DISCUSSIONS ALMOST EVERY DAY
38	NATL-C	0 67067	NAT'L DISCUSSIONS 1-2 TIMES A WEEK OR MISSING
39	NATL-B	0.105106	NATIONAL DISCUSSIONS 3-4 TIMES A MONTH
40	INTL-D	0.054382	INTERNATIONAL DISCUSSIONS ALMOST EVERY DAY
41	INTL-C	0.032460	INT'L DISCUSSIONS 1-2 TIMES A WEEK OR MISSING
42	INTL-B	0.091534	INTERNATIONAL DISCUSSIONS 3-4 TIMES A MONTH

Table C-34

Estimated Effects for U.S. History Cross-sectional Conditioning Variables
1988 Focused-BIB and Intercorrelation Samples, Grade 4/Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-1.082049	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.092760	SEX (FEMALE)
3	ETHNIC2	-0.311443	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.302839	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.084755	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.256330	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.099060	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	0.060228	REGION (SOUTHEAST)
9	REGION3	0.015842	REGION (CENTRAL)
10	REGION4	-0.054248	REGION (WEST)
11	PARED2	0.070426	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.285145	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARI	0.296709	PARENTS EDUCATION (COLLEGE GRAD)
14	PARL	0.079094	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.086135	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.183931	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.164997	TV WATCHING
18	TV. 2	-0.243077	TV WATCHING
19	HOMELNG1	0.001665	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	-0.050667	HOMEWORK (NONE ASSIGNED)
21	HW-YES	-0.021858	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	-0.008786	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.014315	PERCENT IN PROGRAM
24	%LUNCH 2	-0.158767	BLANK
25	%WHITE 1	-0.121124	WHITE MINORITY
26	%WHITE 2	-0.039429	INTEGRATED
27	AGE/GRD1	-0.562740	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	-0.076542	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD3	0.338196	MODAL AGE, > MODAL GRADE
30	AGE/GRD4	-0.190299	> MODAL AGE, MODAL GRADE
31	SCH.TYPE	0.113039	PUBLIC V. PRIVATE SCHOOLS
32	H/HOMEWK	-0.106155	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
33	PRESCH	0.138178	WENT TO PRESCHOOL
34	PAR/HOME	0.029196	SINGLE/MULTIPLE PARENT (MULTIPLE)
35	MOTH/HOM	0.125032	MOTHER AT HOME (YES)
36	MOTH/WRK	-0.018852	MOTHER WORK OUTSIDE HOME (YES)
37	GRWN HOM	0.062107	GROWNUP AT HOME RIGHT AFTER SCHOOL
38	PAGE/RD1	0.053224	PAGES A DAY READ (6 AND UP)
39	PAGE/RD2	0.047888	PAGES A DAY READ (11 AND UP)
40	HIS TAKE	-0.015240	HOW OFTEN DO YOU HAVE A SOCIAL STUDIES CLASS
41	STUD.HIS	0.068609	STUDIED THE HISTORY OF OUR COUNTRY

(continued)

Table C-34 (continued)

Estimated Effects for U.S. History Cross-sectional Conditioning Variables
1988 Focused-BIB and Intercorrelation Samples, Grade 4/Age 9

	Variable	Estimated Effect	Description
42	STUD.EXP	0.044026	STUDIED ABOUT THE EARLY EXPLORERS
43	STUD.IND	-0.002413	STUDIED ABOUT THE AMERICAN INDIANS
44	STUD.PIL	-0.001754	STUDIED ABOUT THE PILGRIMS
45	STUD.GW	-0.045748	STUDIED ABOUT GEORGE WASHINGTON
46	STUD.PIO	0.031049	STUDIED ABOUT THE PIONEERS
47	STUD.SLA	-0.011719	STUDIED ABOUT SLAVERY
48	STUD.INV	-0.029672	STUDIED ABOUT PEOPLE WHO INVENTED THINGS

Table C-35

Estimated Effects for U.S. History Cross-sectional Conditioning Variables
1988 Focused-BIB and Intercorrelation Samples, Grade 8/Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-0.833818	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.150065	SEX (FEMALE)
3	ETHNIC2	-0.175538	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.235940	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.047478	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.086050	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.044790	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	-0.068347	REGION (SOUTHEAST)
9	REGION3	0.008025	REGION (CENTRAL)
10	REGION4	-0.025248	REGION (WEST)
11	PARED2	0.046675	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.163076	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.162114	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.040112	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.073651	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.139044	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.063579	TV WATCHING
18	TV. 2	-0.107249	TV WATCHING
19	HOMELNG1	-0.015749	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	0.09[illegible]64	HOMEWORK (NONE ASSIGNED)
21	HW-YES	0.253431	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	0.023776	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.011917	PERCENT IN PROGRAM
24	%LUNCH 2	-0.020277	BLANK
25	%WHITE 1	-0.024245	WHITE MINORITY
26	%WHITE 2	0.007425	INTEGRATED
27	AGE/GRD1	-0.387768	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	-0.124035	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD3	0.170160	MODAL AGE, > MODAL GRADE
30	AGE/GRD4	-0.208910	> MODAL AGE, MODAL GRADE
31	SCH.TYPE	0.056116	PUBLIC V. PRIVATE SCHOOLS
32	H/HOMEWK	-0.142969	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
33	PAR/HOME	0.004462	SINGLE/MULTIPLE PARENT (MULTIPLE)
34	MOTH/HOM	0.075370	MOTHER AT HOME (YES)
35	MOTH/WRK	-0.022306	MOTHER WORK OUTSIDE HOME (YES)
36	PAGE/RD1	0.078121	PAGES A DAY READ (6 AND UP)
37	PAGE/RD2	-0.007233	PAGES A DAY READ (11 AND UP)
38	GRAD HS	0.262266	DO YOU EXPECT TO GRADUATE HIGH SCHOOL
39	DAYS/MIS	0.054914	DAYS OF SCHOOL MISSED LAST MONTH (NONE)
40	GRADES	0.221464	GRADES IN SCHOOL
41	HIS TAKE	-0.003141	HISTORY COURSES TAKEN IN GRADES 5-8

(continued)

Table C-35 (continued)

Estimated Effects for U.S. History Cross-sectional Conditioning Variables
1988 Focused-BIB and Intercorrelation Samples, Grade 8/Age 13

	Variable	Estimated Effect	Description
42	EARLY H	0.338573	STUDIED EARLY US HISTORY
43	POST C H	-0.174067	STUDIED POST CIVIL WAR
44	TEACH H	-0.150739	THINGS TEACHER ASKS FOR SOCIAL STUDIES CLASS
45	TEXTBOOK	-0.077157	DIFFICULTY READING US HISTORY TEXTBOOK

Table C-36

Estimated Effects for U.S. History Cross-sectional Conditioning Variables
1988 Focused-BIL and Intercorrelation Samples, Grade 12/Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-0.535861	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.171101	SEX (FEMALE)
3	ETHNIC2	-0.277811	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.249072	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.145208	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.181126	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.086993	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	-0.030869	REGION (SOUTHEAST)
9	REGION3	0.029813	REGION (CENTRAL)
10	REGION4	0.015950	REGION (WEST)
11	PARED2	0.048557	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.125807	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.147278	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.093690	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.058395	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.108175	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.048061	TV WATCHING
18	TV. 2	-0.006601	TV WATCHING
19	HOMELNG1	-0.027066	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	0.311514	HOMEWORK (NONE ASSIGNED)
21	HW-YES	0.438620	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	-0.013176	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	-0.000294	PERCENT IN PROGRAM
24	%LUNCH 2	0.006432	BLANK
25	%WHITE 1	-0.057512	WHITE MINORITY
26	%WHITE 2	-0.022398	INTEGRATED
27	AGE/GRD1	-0.191419	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	-0.067168	MODAL AGE, MODAL GRADE, MISSING
29	AGE/GRD4	-0.122420	> MODAL AGE, MODAL GRADE
30	SCH.TYPE	-0.062657	PUBLIC V. PRIVATE SCHOOLS
31	H/HOMEWK	-0.190038	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
32	PAR/HOME	0.005518	SINGLE/MULTIPLE PARENT (MULTIPLE)
33	MOTH/HOM	0.165298	MOTHER AT HOME (YES)
34	MOTH/WRK	-0.008933	MOTHER WORK OUTSIDE HOME (YES)
35	PAGE/RD1	0.069346	PAGES A DAY READ (6 AND UP)
36	PAGE/RD2	0.059610	PAGES A DAY READ (11 AND UP)
37	DAYS/MIS	0.037586	DAYS OF SCHOOL MISSED LAST MONTH (NONE)
38	GRADES	0.148587	GRADES IN SCHOOL
39	HS PROG1	0.157191	COLLEGE
40	HS PROG2	-0.049570	VOCATIONAL, TECHNICAL
41	POST S 1	0.041032	TWO-YEAR COLLEGE

(continued)

Table C-36 (continued)

Estimated Effects for U.S. History Cross-sectional Conditioning Variables
1988 Focused-BIB and Intercorrelation Samples, Grade 12/Age 17

	Variable	Estimated Effect	Description
42	POST S 2	0.163704	FOUR-YEAR COLLEGE
43	HRS WORK	-0.001431	HOURS OF OUTSIDE WORK
44	ENG CL 1	0.128965	TYPE OF ENGLISH CLASS ADVANCED/COLLEGE
45	ENG CL 2	-0.196378	TYPE OF ENGLISH CLASS REMEDIAL
46	HIS TAKE	-0.082710	HISTORY COURSES TAKEN IN GRADES 9-12
47	HIS COUR	-0.000022	US HISTORY COURSE COMPLETE
48	AP HIST	0.034608	TAKEN ADVANCED PLACEMENT AMERICAN HISTORY
49	AM HIST	0.195041	STUDIED AMERICAN HISTORY
50	WORLD H	-0.088624	STUDIED WORLD HISTORY
51	MIN/WOM	-0.071197	STUDIED HISTORY OF MINORITIES AND WOMEN
52	LIKE H	0.112180	DO YOU LIKE STUDYING US HISTORY
53	HAD HIST	0.045186	EVER HAD A US HISTORY CLASS
54	HOMEWK H	0.023091	TIME SPENT EACH WEEK ON US HIST. HOMEWORK
55	GRADES H	0.071975	GRADES YOU HAVE GOTTEN ON US HISTORY
56	TEACH/H	-0.104300	THINGS TEACHER ASKS FOR US HISTORY CLASS
57	TEXTBOOK	-0.080254	DIFFICULTY READING US HISTORY TEXTBOOK

Table C-37

Estimated Effects for Geography Cross-sectional Conditioning Variables
1988 Focused-BIB and Intercorrelation Samples, Grade 12/Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-0.332830	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.420687	SEX (FEMALE)
3	ETHNIC2	-0.685879	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.418485	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.184878	OBSERVED ETHNICITY (ASIAN)
6	STOC2	0.269361	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC3	0.169736	SIZE AND TYPE OF COMMUNITY (NOT HIGH OR LOW)
8	REGION2	-0.011230	REGION (SOUTHEAST)
9	REGION3	0.024789	REGION (CENTRAL)
10	REGION4	0.047082	REGION (WEST)
11	PARED2	0.113712	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.279869	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.287398	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.158917	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	HOMEITM3	0.117676	ARTICLES IN HOME (YES TO 3)
16	HOMEITM4	0.196076	ARTICLES IN HOME (YES TO 4)
17	TV. 1	0.042435	TV WATCHING
18	TV. 2	-0.087285	TV WATCHING
19	HOMELNG1	0.021254	OTHER LANGUAGE AT HOME (SOMETIMES, ALWAYS)
20	HW-NO	-0.659565	HOMEWORK (NONE ASSIGNED)
21	HW-YES	-0.460317	HOMEWORK (YES - SOME AMOUNT)
22	HW-2345	0.024182	HOMEWORK AMOUNT (LINEAR)
23	%LUNCH 1	0.000708	PERCENT IN PROGRAM
24	%LUNCH 2	-0.089351	BLANK
25	%WHITE 1	-0.107201	WHITE MINORITY
26	%WHITE 2	-0.068820	INTEGRATED
27	AGE/GRD1	-0.415062	MODAL AGE, < MODAL GRADE
28	AGE/GRD2	-0.251628	MODAL AGE. MODAL GRADE, MISSING
29	AGE/GRD4	-0.282149	> MODAL AGE, MODAL GRADE
30	SCH.TYPE	-0.138944	PUBLIC V. PRIVATE SCHOOLS
31	H/HOMEWK	-0.245424	HELP WITH HOMEWORK (EVERY DAY, TWICE A WEEK)
32	PAR/HOME	-0.014887	SINGLE/MULTIPLE PARENT (MULTIPLE)
33	MOTH/HOM	0.119619	MOTHER AT HOME (YES)
34	MOTH/WRK	-0.027926	MOTHER WORK OUTSIDE HOME (YES)
35	PAGE/RD1	0.031073	PAGES A DAY READ (6 AND UP)
36	PAGE/RD2	0.091392	PAGES A DAY READ (11 AND UP)
37	DAYS/MIS	0.065005	DAYS OF SCHOOL MISSED LAST MONTH (NONE)
38	GRADES	0.213982	GRADES IN SCHOOL
39	HS PROG1	0.212703	COLLEGE
40	HS PROG2	0.057015	VOCATIONAL, TECHNICAL
41	POST S 1	0.042615	TWO-YEAR COLLEGE

(continued)

Table C-37 (continued)

Estimated Effects for Geography Cross-sectional Conditioning Variables
1988 Focused-BIB and Intercorrelation Samples Grade 12/Age 17

	Variable	Estimated Effect	Description
42	POST S 2	0.232319	FOUR-YEAR COLLEGE
43	HRS WORK	-0.002[illegible]	HOURS OF OUTSIDE WORK
44	ENG CL 1	0.114107	TYPE OF ENGLISH CLASS ADVANCED/COLLEGE
45	ENG CL 2	-0.303248	TYPE OF ENGLISH CLASS REMEDIAL
46	GEO TAKE	0.004362	GEOGRAPHY COURSES TAKEN IN GRADES 9-12
47	W GEO	-0.073306	WORLD GEOGRAPHY COURSE COMPLETE
48	W HIS/G	-0.105087	WORLD HISTORY/GEO COURSE COMPLETE
49	STAT/REG	0.105754	STATE/REGIONAL COURSE COMPLETE
50	US GEO	-0.140865	UNITED STAES COURSE COMPLETE
51	US H/G	0.180866	UNITED SATES HIST/GEO COURSE COMPLETE
52	PHY/GEO	-0.082555	PHYSICAL GEO/EARTH SCIENCE COURSE COMPLETE
53	ECON/P	-0.048159	ECONOMIC AND POLITICAL COURSE COMPLETE
54	HUM/CULT	0.190482	HUMAN AND CULTURAL COURSE COMPLETE
55	URBAN G	-0.144868	URBAN COURSE COMPLETE
56	PHYSICAL	0.083896	SUM OF PHYSICAL GEO TOPICS STUDIED
57	SOCIAL	-0.089169	SUM OF SOCIAL GEO TOPICS STUDIED

Table C-38

Estimated Effects for Mathematics Trend Conditioning Variables
1988 Bridge to 1986, Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-0.279547	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.047747	GENDER (FEMALE)
3	ETHNIC2	-0.706632	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	0.209298	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.762678	OBSERVED ETHNICITY (ASIAN)
6	STOC3	0.186615	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC1	0.087756	SIZE AND TYPE OF COMMUNITY (NOT HI&NOT LO)
8	REGION2	0.007280	REGION (SOUTHEAST)
9	REGION3	0.123942	REGION (CENTRAL)
10	REGION4	-0.035032	REGION (WEST)
11	PARED2	0.251057	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.223869	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.454556	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.136615	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	<MODAL GRADE	-0.728308	MODAL GRADE (LESS THAN MODAL GRADE)
16	>MODAL GRADE	0.631198	MODAL GRADE (GREATER THAN MODAL GRADE)
17	ITEMS2	0.239816	ITEMS IN THE HOME (YES TO 3)
18	ITEMS3	0.367498	ITEMS IN THE HOME (YES TO ALL 4)
19	E2 X SEX	0.087308	ETHNICITY BY GENDER (BLACK, FEMALE)
20	E3 X SEX	-0.066049	ETHNICITY BY GENDER (HISPANIC, FEMALE)
21	E4 X SEX	-0.231095	ETHNICITY BY GENDER (ASIAN AMERICAN, FEMALE)
22	E2 X PE2	0.063586	ETHNICITY BY PARENT'S ED (BLACK, HS GRAD)
23	E2 X PE3	0.375105	ETHNICITY BY PARENT'S ED (BLACK, POST HS)
24	E2 X PE4	0.039552	ETHNICITY BY PARENT'S ED (BLACK, COLLEGE)
25	E2 X PE_	0.191412	ETHNICITY BY PARENT'S ED (BLACK, UNKNOWN)
26	E3 X PE2	-0.354255	ETHNICITY BY PARENT'S ED (HISPANIC, HS GRAD)
27	E3 X PE3	0.237226	ETHNICITY BY PARENT'S ED (HISPANIC POST HS)
28	E3 X PE4	-0.256883	ETHNICITY BY PARENT'S ED (HISPANIC, COLLEGE)
29	E3 X PE_	-0.246003	ETHNICITY BY PARENT'S ED (HISPANIC, UNKNOWN)
30	E4 X PE2	-1.034833	ETHNICITY BY PARENT'S ED (ASIAN AM, HS GRAD)
31	E4 X PE3	-0.690193	ETHNICITY BY PARENT'S ED (ASIAN AM, POST HS)
32	E4 X PE4	-0.786758	ETHNICITY BY PARENT'S ED (ASIAN AM, COLLEGE)
33	E4 X PE_	-0.518339	ETHNICITY BY PARENT'S ED (ASIAN AM, UNKNOWN)
34	SCH TYP2	0.158816	SCHOOL TYPE (NOT PUBLIC)
35	SCH TYP_		SCHOOL TYPE (MISSING)
36	TV1	0.278883	0-2 HOURS OF TV WATCHING
37	TV2	0.434684	3-5 HOURS OF TV WATCHING
38	TV3	0.259356	6+ HOURS OF TV WATCHING
39	LANGHOM3	-0.283533	LANGUAGE IN HOME OTHER THAN ENG ? (ALWAYS)
40	LANGHOM2	0.088718	LANGUAGE IN HOME OTHER THAN ENG.? SOMETIMES
41	E2 X LH[illegible]	0.143997	ETHNICITY BY LANGUAGE IN HOME (BLACK, OFTEN)

(continued)

Table C-38 (continued)

Estimated Effects for Mathematics Trend Conditioning Variables
1988 Bridge to 1986, Age 9

	Variable	Estimated Effect	Description
42	E2 X LH2	0.080093	ETHNICITY BY LANG IN HOME (BLACK, SOMETIMES)
43	E3 X LH1	0.390581	ETHNICITY BY LANGUAGE IN HOME(HISPANIC,OFTEN)
44	E3 X LH2	-0.117348	ETHNICITY BY LANG IN HOME(HISPANIC,SOMETIMES)
45	E4 X LH1	0.411867	ETHNICITY BY LANGUAGE IN HOME(ASIAN AM,OFTEN)
46	E4 X LH2	0.238582	ETHNICITY BY LANG IN HOME(ASIAN AM,SOMETIMES)
47	TIME ASS		TIME OF ASSM'T(APPLICABLE FOR Y17, N/AY19)
48	STUDYCMP	-0.057134	ARE YOU STUDYING COMPUTERS? B004501 (YES)
49	DRACE2	-0.069875	DERIVED RACE/ETHNICITY (BLACK)
50	DRACE3	-0.341651	DERIVED RACE/ETHNICITY (HISPANIC)
51	DRACE4	0.185246	DERIVED RACE/ETHNICITY (ASIAN AMERICAN)

Table C-39

Estimated Effects for Mathematics Trend Conditioning Variables
1988 Bridge to 1986, Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-1.504811	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.228401	GENDER (FEMALE)
3	ETHNIC2	-0.242682	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	0.086195	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.378006	OBSERVED ETHNICITY (ASIAN)
6	STOC3	0.534516	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
7	STOC1	0.298905	SIZE AND TYPE OF COMMUNITY (NOT HI&NOT LO)
8	REGION2	-0.121025	REGION (SOUTHEAST)
9	REGION3	-0.063070	REGION (CENTRAL)
10	REGION4	-0.107134	REGION (WEST)
11	PARED2	0.140058	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.197777	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.278975	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.021061	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	<MODAL GRADE	-0.480949	MODAL GRADE (LESS THAN MODAL GRADE)
16	>MODAL GRADE	0.541153	MODAL GRADE (GREATER THAN MODAL GRADE)
17	ITEMS2	0.122176	ITEMS IN THE HOME (YES TO 3)
18	ITEMS3	0.177230	ITEMS IN THE HOME (YES TO ALL 4)
19	E2 X SEX	0.020985	ETHNICITY BY GENDER (BLACK, FEMALE)
20	E3 X SEX	0.099927	ETHNICITY BY GENDER (HISPANIC, FEMALE)
21	E4 X SEX	-0.0[illegible]259	ETHNICITY BY GENDER (ASIAN AMERICAN, FEMALE)
22	E2 X PE2	-0.181870	ETHNICITY BY PARENT'S ED (BLACK, HS GRAD)
23	E2 X PE3	-0.179468	ETHNICITY BY PARENT'S ED (BLACK, POST HS)
24	E2 X PE4	-0.397062	ETHNICITY BY PARENT'S ED (BLACK, COLLEGE)
25	E2 X PE_	0.090978	ETHNICITY BY PARENT'S ED (BLACK, UNKNOWN)
26	E3 X PE2	-0.033586	ETHNICITY BY PARENT'S ED (HISPANIC, HS GRAD)
27	E3 X PE3	-0.035114	ETHNICITY BY PARENT'S ED (HISPANIC, POST HS)
28	E3 X PE4	-0.359408	ETHNICITY BY PARENT'S ED (HISPANIC, COLLEGE)
29	E3 X PE_	-0.15[illegible]07	ETHNICITY BY PARENT'S ED (HISPANIC, UNKNOWN)
30	E4 X PE2	-0.412270	ETHNICITY BY PARENT'S ED (ASIAN AM, HS GRAD)
31	E4 X PE3	-1.023135	ETHNICITY BY PARENT'S ED (ASIAN AM, POST HS)
32	E4 X PE4	0.005724	ETHNICITY BY PARENT'S ED (ASIAN AM, COLLEGE)
33	E4 X PE_	-0.148864	ETHNICITY BY PARENT'S ED (ASIAN AM, UNKNOWN)
34	SCH TYP2	0.019369	SCHOOL TYPE (NOT PUBLIC)
35	SCH TYP_		SCHOOL TYPE (MISSING)
36	TV1	-0.192841	0-2 HOURS OF TV WATCHING
37	TV2	-0.259867	3-5 HOURS OF TV WATCHING
38	TV3	0.391540	6+ HOURS OF TV WATCHING
39	HW-NO	0.143508	HOMEWORK (NONE ASSIGNED)
40	HW-YES	0.295564	HOMEWORK (YES - SOME AMOUNT)
41	HW-345	-0.046762	HOMEWORK (LINEAR AMOUNT)

(continued)

Table C-39 (continued)

Estimated Effects for Mathematics Trend Conditioning Variables
1988 Bridge to 1986, Age 13

	Variable	Estimated Effect	Description
42	LANGHOM3	-0.142210	LANGUAGE IN HOME OTHER THAN ENGLISH? (ALWAYS)
43	LANGHOM2	0.050961	LANGUAGE IN HOME OTHER THAN ENG.? (SOMETIMES)
44	E2 X LH1	0.100579	ETHNICITY BY LANGUAGE IN HOME (BLACK, OFTEN)
45	E2 X LH2	0.051984	ETHNICITY BY LANG IN HOME (HISP., SOMETIMES)
46	E3 X LH1	0.032823	ETHNICITY BY LANGUAGE IN HOME (HISP., OFTEN)
47	E3 X LH2	-0.081489	ETHNICITY BY LANG IN HOME (HISP., SOMETIMES)
48	E4 X LH1	-0.295872	ETHNICITY BY LANG IN HOME (ASIAN AM,OFTEN)
49	E4 X LH2	-0.351225	ETHN. BY LANG IN HOME (ASIAN AM,SOMETIMES)
50	GRADES	0.329379	GRADES IN SCHOOL
51	TYPEMAT2	0.557133	TYPE OF MATH CLASS (REGULAR MATH)
52	TYPEMAT3	0.860079	TYPE OF MATH CLASS (PRE-ALGEBRA)
53	TYPEMAT4	1.067878	TYPE OF MATH CLASS (ALGEBRA, OTHER)
54	TIME ASS		TIME OF ASSESSMENT (APPLICABLE Y17, N/A Y19)
55	STUDYCMP	0.000685	ARE YOU STUDYING COMPUTERS? B004501 (YES)
56	DRACE2	0.021696	DERIVED RACE/ETHNICITY (BLACK)
57	DRACE3	-0.262241	DERIVED RACE/ETHNICITY (HISPANIC)
58	DRACE4	0.239560	DERIVED RACE/ETHNICITY (ASIAN AMERICAN)

Table C-40

Estimated Effects for Mathematics Trend Conditioning Variables
1988 Bridge to 1986, Age 17

	Variable	Estimated Effect	Description
1	OVERALL	0.466202	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.227644	GENDER (FEMALE)
3	ETHNIC2	-0.326424	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.125207	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.542147	OBSERVED ETHNICITY (ASIAN)
6	STOC3	0.355679	SIZE AND TYPE OF COMMUNITY (HIGH METR.,
7	STOC1	0.268174	SIZE AND TYPE OF COMMUNITY (NOT HI&NOT LO)
8	REGION2	-0.035567	REGION (SOUTHEAST)
9	REGION3	0.092946	REGION (CENTRAL)
10	REGION4	0.041544	REGION (WEST)
11	PARED2	-0.009106	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.276562	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.215802	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.039054	PARENTS EDUCATION (MISSING, I DON'T KNOW)
15	<MODAL GRADE	-0.212266	MODAL GRADE (LESS THAN MODAL GRADE)
16	>MODAL GRADE	0.091063	MODAL GRADE (GREATER THAN MODAL GRADE)
17	ITEMS2	0.032057	ITEMS IN THE HOME (YES TO 3)
18	ITEMS3	0.089343	ITEMS IN THE HOME (YES TO ALL 4)
19	E2 X SEX	0.130167	ETHNICITY BY GENDER (BLACK, FEMALE)
20	E3 X SEX	0.294555	ETHNICITY BY GENDER (HISPANIC, FEMALE)
21	E4 X SEX	-0.190247	ETHNICITY BY GENDER (ASIAN AMERICAN, FEMALE)
22	E2 X PE2	-0.014269	ETHNICITY BY PARENT'S ED (BLACK, HS GRAD)
23	E2 X PE3	-0.186204	ETHNICITY BY PARENT'S ED (BLACK, POST HS)
24	E2 X PE4	-0.163440	ETHNICITY BY PARENT'S ED (BLACK, COLLEGE)
25	E2 X PE_	-0.256462	ETHNICITY BY PARENT'S ED (BLACK, UNKNOWN)
26	E3 X PE2	0.037801	ETHNICITY BY PARENT'S ED (HISPANIC, HS GRAD)
27	E3 X PE3	-0.197622	ETHNICITY BY PARENT'S ED (HISPANIC, POST HS)
28	E3 X PE4	-0.148578	ETHNICITY BY PARENT'S ED (HISPANIC, COLLEGE)
29	E3 X PE_	0.076608	ETHNICITY BY PARENT'S ED (HISPANIC, UNKNOWN)
30	E4 X PE2	1.148569	ETHNICITY BY PARENT'S ED (ASIAN AM, HS GRAD)
31	E4 X PE3	0.548141	ETHNICITY BY PARENT'S ED (ASIAN AM, POST HS)
32	E4 X PE4	-0.003476	ETHNICITY BY PARENT'S ED (ASIAN AM, COLLEGE)
33	E4 X PE_	0.555852	ETHNICITY BY PARENT'S ED (ASIAN AM, UNKNOWN)
34	SCH TYP2	-0.130104	SCHOOL TYPE (NOT PUBLIC)
35	SCH TYP_		SCHOOL TYPE (MISSING)
36	TV1	-1.980878	0-2 HOURS OF TV WATCHING
37	TV2	-1.992986	3-5 HOURS OF TV WATCHING
38	TV3	-2.079726	6+ HOURS OF TV WATCHING
39	HW-NO	-0.243494	HOMEWORK (NONE ASSIGNED)
40	HW-YES	0.104266	HOMEWORK (YES - SOME AMOUNT)
41	HW-345	-0.024606	HOMEWORK (LINEAR AMOUNT)

(continued)

Table C-40 (continued)

Estimated Effects for Mathematics Trend Conditioning Variables
1988 Bridge to 1986, Age 17

	Variable	Estimated Effect	Description
42	LANGHOM3	-0.306630	LANGUAGE IN HOME OTHER THAN ENGLISH? (ALWAYS)
43	LANGHOM2	-0.027324	LANGUAGE IN HOME OTHER THAN ENG.? (SOMETIMES)
44	E2 X LH1	0.234334	ETHNICITY BY LANGUAGE IN HOME (BLACK, OFTEN)
45	E2 X LH2	-0.085786	ETHNICITY BY LANG IN HOME (HISP, SOMETIMES)
46	E3 X LH1	0.372056	ETHNICITY BY LANGUAGE IN HOME (HISP, OFTEN)
47	E3 X LH2	0.068137	ETHNICITY BY LANG IN HOME (HISP, SOMETIMES)
48	E4 X LH1	0.542742	ETHNICITY BY LANG IN HOME (ASIAN AM, OFTEN)
49	E4 X LH2	0.390736	ETHN. BY LANG IN HOME (ASIAN AM,SOMETIMES)
50	NMATH1	-0.221100	HIGHEST LEVEL MATH TAKEN (PRE-ALGEBRA)
51	NMATH2	0.252774	HIGHEST LEVEL MATH TAKEN (ALGEBRA)
52	NMATH3	0.354687	HIGHEST LEVEL MATH TAKEN (GEOMETRY)
53	NMATH4	0.700470	HIGHEST LEVEL MATH TAKEN (ALGEBRA-2)
54	NMATH5	1.208891	HIGHEST LEVEL MATH TAKEN (CALCULUS)
55	COMPUTER	-0.009892	COMPUTER CLASS TAKEN ? (YES)
56	GRADES	0.293596	GRADES IN SCHOOL
57	HSPROG2	0.196396	HIGH SCHOOL PROGRAM (COLLEGE PREP)
58	HSPROG3	-0.090029	HIGH SCHOOL PROGRAM (VOC/TECH)
59	DRACE2	0.119675	DERIVED RACE/ETHNICITY (BLACK)
60	DRACE3	-0.202548	DERIVED RACE/ETHNICITY (HISPANIC)
61	DRACE4	-0.056777	DERIVED RACE/ETHNICITY (ASIAN AMERICAN)

Table C-41

Estimated Effects for Science Trend Conditioning Variables
1988 Bridge to 1986, Age 9

	Variable	Estimated Effect	Description
1	OVERALL	-0.167629	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.160032	GENDER (FEMALE)
3	ETHNIC2	-0.716027	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.677694	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	-0.143962	OBSERVED ETHNICITY (ASIAN)
6	STOC2	-0.400385	SIZE AND TYPE OF COMMUNITY (LOW METRO)
7	STOC3	0.114765	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
8	REGION2	0.105314	REGION (SOUTHEAST)
9	REGION3	0.202669	REGION (CENTRAL)
10	REGION4	0.08181[illegible]	REGION (WEST)
11	PARED2	0.200699	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.279235	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.435635	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	0.172272	PARENTS EDUCATION (MISSING, DON'T KNOW)
15	<MODAL GRADE	-0.498134	MODAL GRADE (LESS THAN MODAL GRADE)
16	>MODAL GRADE	1.050936	MODAL GRADE (GREATER THAN MODAL GRADE)
17	ITEMS2	0.289243	ITEMS IN THE HOME (YES TO [illegible]
18	ITEMS3	0.478227	ITEMS IN THE HOME (YES TO ALL 4)
19	SCH TYP2	0.076284	SCHOOL TYPE (NOT PUBLIC)

646

Table C-42

Estimated Effects for Science Trend Conditioning Variables
1988 Bridge to 1986, Age 13

	Variable	Estimated Effect	Description
1	OVERALL	-0.048884	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.267412	GENDER (FEMALE)
3	ETHNIC2	-0.719052	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.524609	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.161636	OBSERVED ETHNICITY (ASIAN)
6	STOC2	-0.395130	SIZE AND TYPE OF COMMUNITY (LOW METRO)
7	STOC3	-0.007911	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
8	REGION2	-0.077003	REGION (SOUTHEAST)
9	REGION3	0.046762	REGION (CENTRAL)
10	REGION4	-0.102571	REGION (WEST)
11	PARED2	0.107733	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.3[illegible]7308	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.412279	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.047971	PARENTS EDUCATION (MISSING, DON'T KNOW)
15	<MODAL GRADE	-0.530171	MODAL GRADE (LESS THAN MODAL GRADE)
16	>MODAL GRADE	0.969538	MODAL GRADE (GREATER THAN MODAL GRADE)
17	ITEMS2	0.222418	ITEMS IN THE HOME (YES TO 3)
18	ITEMS3	0.404732	ITEMS IN THE HOME (YES TO ALL 4)
19	SCH TYP2	0.128735	SCHOOL TYPE (NOT PUBLIC)

Table C-43

Estimated Effects for Science Trend Conditioning Variables
1988 Bridge to 1986, Age 17

	Variable	Estimated Effect	Description
1	OVERALL	-0.018353	OVERALL CONSTANT '1' FOR EVERYONE
2	GENDER2	-0.422265	GENDER (FEMALE)
3	ETHNIC2	-0.675393	OBSERVED ETHNICITY (BLACK)
4	ETHNIC3	-0.028940	OBSERVED ETHNICITY (HISPANIC)
5	ETHNIC4	0.105174	OBSERVED ETHNICITY (ASIAN)
6	STOC2	-0.215624	SIZE AND TYPE OF COMMUNITY (LOW METRO)
7	STOC3	0.200910	SIZE AND TYPE OF COMMUNITY (HIGH METRO)
8	REGION2	-0.078230	REGION (SOUTHEAST)
9	REGION3	-0.145136	REGION (CENTRAL)
10	REGION4	-0.156447	REGION (WEST)
11	PARED2	0.277744	PARENTS EDUCATION (HIGH SCHOOL GRAD)
12	PARED3	0.506933	PARENTS EDUCATION (POST HIGH SCHOOL)
13	PARED4	0.724225	PARENTS EDUCATION (COLLEGE GRAD)
14	PARED_	-0.353136	PARENTS EDUCATION (MISSING, DON'T KNOW)
15	<MODAL GRADE	-0.540566	MODAL GRADE (LESS THAN MODAL GRADE)
16	>MODAL GRADE	0.345666	MODAL GRADE (GREATER THAN MODAL GRADE)
17	ITEMS2	0.091730	ITEMS IN THE HOME (YES TO 3)
18	ITEMS3	0.208488	ITEMS IN THE HOME (YES TO ALL 4)
19	SCH TYP2	-0.094395	SCHOOL TYPE (NOT PUBLIC)

APPENDIX D

1988 NAEP Composite and Derived Conditioning Variables

Appendix D

1988 NAEP COMPOSITE AND DERIVED CONDITIONING VARIABLES

Some conditioning variables used in the creation of plausible values were constructed by recoding the values of a data variable or by combining and recoding data from two or more variables. This appendix describes how these conditioning variables were derived for the 1988 assessments of reading writing, civics, U.S. history, geography, mathematics and science.

Reading Cross-sectional Variables

Seven composites of background, attitude, and instructional variables were created for use as conditioning variables for cross-sectional studies in reading. The following list shows the items included in the composites. In some cases, the constituent items vary across age cohorts, as indicated. In each case, items within a composite were reflected so that they were coded "in the same direction" (e.g., more frequent use of a teaching technique received a higher score than less frequent use, etc.). The composite score for each student is a weighted sum of the item scores. (Weights were based on principal component analyses of each set of items; the weights were those associated with the first principal component.) Students with scores more than half a standard deviation above the mean on the composite were considered "high," those with scores more than half a standard deviation below the mean were considered low, and the remaining students were considered "medium." Composite scores were not calculated for students who responded to fewer than two of the constituent items. If the student responded to at least two items but did not respond to all items in the composite, the composite score was taken to be the mean on the completed items. See Table C-2 ("Contrast Codings for 1988 Reading Cross-sectional Conditioning Variables") for the coding of the constituent items. (Missing responses were treated as stated above and not recoded as indicated in Table C-2.)

RHOME (Home environment and support)

		Ages		
R800101	Magazines at home	9		
R800201	Magazines at home		13	17
R800301	Books at home	9	13	17
R800401	Read stories	9	13	17

RINDRDG (Students' independent reading)

		Ages		
R800601	Talk about reading		13	17
S003501	Read for fun	9	13	17
S003502	Talk about book	9	13	17
S003503	Books from library	9	13	17
S003504	Buy own books	9	13	17
S003505	Read after TV show		13	17
S003506	Author you like		13	17
L801101	Read in spare time		13	17
R800701	Read stories	9		
S004301	Read story,novel			17

RUSEMAT (Students' use of materials)

		Ages	
R800801	Use dictionary	13	17
S007309	Use encyclopedia	13	17
S004304	Read newspaper		17
S004305	Read magazine		17

RTEACH (Teacher instructional behaviors)

		Ages	
S004602	Teacher-main idea	13	17
S004601	Teacher-vocabulary	13	17
S004701	Teacher-questions	13	17
R801201	Analyze reading	13	17
S008501	Write on reading	13	17

RBEHAV (Students' behaviors)

		Ages
R800901	Time on literature	17
R801001	Time on homework	17

RSTRATG (Reading strategies)

		Ages
S005101	Read material over	17
S005102	Take notes	17
S005103	Make outlines	17
S005104	Answer questions	17

RSCHWRK (Students' school/coursework)

		Ages
R801101	English grades	17
S006403	Advanced placement English	17
S006401	Remedial English	17

Writing Cross-sectional Variables

WENJOY (Enjoy Writing)

For grades 4, 8, and 12, responses to item S001201 were recoded as follows:

Always	- 5
Half +	- 4
Half	- 3
Half -	- 2
None	- 1
Missing	- 3

WPOSFB (Positive Teacher Feedback)

For grades 4, 8, and 12, responses to items S001703, S001706, and S002503 were recoded as indicated below and then summed. Values ranged from 0 to 3.

Mostly all	- 1
Half +	- 1
Half	- 0
Half -	- 0
Never	- 0
Missing	- 0

WNEGFB (Negative Teacher Feedback)

For grades 4, 8, and 12, responses to items S001708 and S002505 were recoded as indicated below and then summed. Values ranged from 0 to 2.

Mostly all	- 0
Half +	- 0
Half	- 0
Half -	- 1
Never	- 1
Missing	- 0

WAMTWE (Amount Written)

For grades 8 and 12, responses to items W800302 (1 or 2 pages) and W800303 (3 pages) were recoded as follows:

Daily	= 1
Weekly	= 2
Monthly	= 3
Yearly	= 4
Never	= 5
Missing	= Missing

The variables were then coded as six dummy variables as follows:

W800302	W800303	Dummy Variables
1,2	5	100000
1,2	1,2	010000
1,2	3,4	001000
3,4,missing	3,4,missing	000100
3,4	5	000010
5	5	000000
5	1,2,3,4	000001
3,4	1,2	000001

WAMTWH (Writing in History Class)

For grades 8 and 12, responses to items W800802 (1 or 2 pages) and W800803 (3 pages) were recoded as follows:

Daily	= 1
Weekly	= 2
Monthly	= 3
Yearly	= 4
Never	= 5
Missing	= Missing

The variables were then coded as six dummy variables as follows:

W800802	W800803	Dummy Variables
1,2	5	100000
1,2	1,2	010000
1,2	3,4	001000
3,4,missing	3,4,missing	000100
3,4	5	000010
5	5	000000
5	1,2,3,4	000001
3,4	1,2	000001

WINST2 (Instruction on Writing Process II)

For grades 8 and 12, responses to items S000606, S000608, S000609, and S000610 were recoded as indicated below; the recoded responses for the four items were then averaged.

Always	- 5
Half +	- 4
Half	- 3
Half -	- 2
Never	- 1

WSUC2 (Success in English Class II)

For grades 8 and 12, responses to item S008601 were coded as follows:

A lot	- 1
Some	- 2
None	- 3
Missing	- 2

WSUC3 (Success in English Class III)

For grade 12 responses to item W800701 were recoded as follows:

Most A	- 8
Half A-B	- 7
Most B	- 6
Half B-C	- 5
Most C	- 4
Half C-D	- 3
Most D	- 2
Most < D	- 1
Missing	- 5

WTIMSP3 (Time Spent III)

For grade 12, responses to four items were recoded as indicated below. The recoded responses to all four items were then averaged. If responses to all four items were missing, WTIMSP3 was given a value of 2.

Items Used	Codings	
W800501	Weekly	- 4
W800502	Monthly	- 3
	Yearly	- 2
	Never	- 1

W800601	None	- 2
	Don't do	- 1
	< 1 hour	- 2
	1 hour	- 3
	2 hour	- 4
	3+ hours	- 4
S000201	None	- 1
	1 to 2	- 2
	3 to 4	- 3
	5 to 10	- 4
	11+	- 4

WREVISE (Revision)

For grade 12, responses to items S001302, S001303, S001304, S001308, and S001301 were recoded as indicated below. The recoded responses to all five items were then averaged. If responses to all five items were missing, WREVISE was given a value of 3.

Always	- 5
Half +	- 4
Half	- 3
Half -	- 2
Never	- 1

WPLANNG (Planning)

For grade 12, responses to items S000901 and S000902 were recoded as indicated below. The recoded responses to both items were then averaged. If responses to both items were missing, WPLANNG was given a value of 3.

Most, all	- 5
Half +	- 4
Half	- 3
Half -	- 2
Never	- 1

WTIMSP2 (Time Spent II)

For grade 8, responses to item W800502 were recoded as follows:

Weekly	- 4
Monthly	- 3
Yearly	- 2
Never	- 1
Missing	- 3

WTIMSP1 (Time Spent I)

For grade 4, three dummy variables were created from the responses to item W800201 as follows:

1	Weekly	100
2	Monthly	010
3	Yearly	001
4	Never	000
	Missing	010

WSUC1 (Success in English Class I)

For grade 4, four dummy variables were created from the responses to item W800401 as follows:

1	Daily	1000
2	Weekly	0100
3	Monthly	0010
4	Yearly	0001
5	Never	0000
	Missing	0010

WINST1 (Instruction on Writing Process 1)

For grade 4, responses to items S000604 and S000605 were recoded as indicated below. The recoded responses to both items were then averaged. If responses to both items were missing, WINST1 was given a value of 3.

Always	= 5
Half +	= 4
Half	= 3
Half -	= 2
Never	= 1

Writing Trend Variables

NUMREP (Number of Reports Written for English Class)

For grades 8 and 12, if the student had an English class (i.e., response to items B001208 is not 1), responses to items B001201-B001207 were recoded as follows:

None	= 0
1 or 2 written	= 1
3 or more written	= 1
BLK missing	= 0

The sum of the seven responses was then computed and the number of reports written for English class became 0 (for 0 reports or missing) or 1 to 7 (for 1 to 7 reports).

Civics Cross-sectional Variables

ST-ALOT (Number of Subjects Studied a Lot)

For grade 4, items P800601, P800701, P800801, P800901, P801001, and P801101 asked students to indicate whether they had studied particular civics-related topics a lot, some, or not at all. Each of the six items was coded as follows:

A lot = 1
Some = 0
Not at all = 0
Missing = 0

The sum of the six recoded variables was defined as ST-ALOT.

STCV-INC (Studied Civics—Incomplete Set of Answers)

For grade 8, items P800101 to P800104 asked students whether they had studied American government or civics in grades 5 to 8, respectively. Each of the four items was coded as:

Yes = 0
No = 0
I don't know = 1
Missing = 1

The sum of the four variables was then assigned the codes:

0 = 0
1-4 = 1

ST-ALOT (Number of Subjects Studied a Lot)

For grade 8, items P801201 to P801210 asked students to indicate whether they had studied particular civics-related topics a lot, some or not at all. Each of the ten items was coded as follows:

A lot = 1
Some = 0
Not at all = 0
Missing = 0

The sum of the ten recoded variables was defined as ST-ALOT.

CVA-ACT (Active in 4 Areas of Civics, 1, 2, 5, and 8)

For grade 8, items P802201, P802202, P802205, and P802208 asked students to indicate how often they participated in particular educational activities for social studies class. These items were grouped together because of the similarity of responses to them. Each of the four items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
A few times a year	= 0
Never	= 0
Missing	= 0

The sum of the four variables was then assigned the codes:

4	= 1
0-3	= 0

CVB-ACT (Active in 2 Areas of Civics, 3 and 10)

For grade 8, items P802203 and P802210 asked students to indicate how often they participated in particular educational activities for social studies class. These items were grouped together because of the similarity of responses to them. Each of the two items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
Never	= 0
Missing	= 0

The sum of the two variables was then assigned the codes:

2	= 1
0-1	= 0

CVC-ACT (Active in 2 Areas of Civics, 4 and 9)

For grade 8, items P802204 and P802209 asked students to indicate how often they participated in particular educational activities for social studies class. These items were grouped together because of the similarity of responses to them. Each of the two items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
A few times a year	= 0
Never	= 0
Missing	= 0

The sum of the two variables was then assigned the codes:

2	= 1
0-1	= 0

CVD-ACT (Active in 2 Areas of Civics, 6 and 7)

For grade 8, items P802206 and P802207 asked students to indicate how often they participated in particular educational activities for social studies class. These items were grouped together because of the similarity of responses to them. Each of the two items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
A few times a year	= 0
Never	= 0
Missing	=

The sum of the two variables was then assigned the codes:

2	= 1
0-1	= 0

CVD-NOT (Not very Active in these 2 Areas of Civics)

For grade 8, items P802206 and P802207 asked students to indicate how often they participated in particular educational activities for social studies class. These items were grouped together because of the similarity of responses to them. Each of the two items was recoded as:

Almost every day	= 0
Once or twice a week	= 0
Once or twice a month	= 0
A few times a year	= 1
Never	= 1
Missing	= 1

The sum of the two variables was then assigned the codes:

2	= 1
0-1	= 0

STCV-INC (Studied Civics-Incomplete set of Answers)

For grade 12, items P800201 to P800204 asked students whether they had studied American government or civics in grades 9 to 12, respectively. Each of the four items was coded as:

Yes	= 0
No	= 0
I don't know	= 1
Missing	= 1

The sum of the four variables was then assigned the codes:

0	= 0
1-4	= 1

GOVT (How Many Years of Government Study have You Had)

For grade 12, item P800301 asked for the number of years the student had studied American government or civics since the beginning of ninth grade. The responses were recoded as:

None	= 0
Less than 1/2 year	= .25
1/2 year	= .5
Between 1/2 year and 1 year	= .75
1 year	= 1.0
More than 1 year	= 2.0
Missing	= 0

ST-ALOT (Number of Subjects Studied a Lot)

For grade 12, items P801301 to P801310 asked students to indicate whether they had studied particular civics-related topics a lot, some or not at all. Each of the ten items was coded as follows:

A lot	= 1
Some	= 0
Not at all	= 0
Missing	= 0

The sum of the ten recoded variables was defined as ST-ALOT.

ST-AVE (Average Level of Study for These Ten)

For grade 12, responses to items P801301 to P801310 were recoded as:

A lot	= 2
Some	= 2
Not at all	= 1
Missing	= 0

The average of the ten recoded variables was defined as ST-ALOT.

CIV-INT (Level of Interest in Civics)

For grade 12, item P801401 asked students how much they like studying American government or civics. The responses were recoded as:

One of my favorite subjects	= 3
Interesting	= 2
Like several other subjects better	= 1
Never studied	= 0

GOVATTN (Degree of Attention Government Pays)

For grade 12, items P801601 and P801701 asked students to give their opinions about the responsiveness of government to the people. Each of the two items was coded as:

A good deal	= 2
Some	= 1
Not much	= 0

The average of the two variables was defined as GOVATTN.

The responses to HADCIV (yes, have had a civics class) influenced the coding of the following variables. This item (P801801) asked students if they had ever had an American government or civics class. If they had, they were to continue answering questions in the civics background variable block. If not, they were asked to stop. If students continued to answer questions although they had not said that they had taken a civics course, the rest of the responses were coded as missing.

CVHWORK (Amount of Civics Homework)

For grade 12, responses to item P801901 were recoded as:

No assignments	= 0
Don't do assignment	= 0
Less than 1 hour	=.5
1 hour	= 1
2 hours	= 2
3 hours	= 3
4 hours	= 4
5 hours or more	= 5
Did not say that Civics was taken	= 0
Missing	= 0

CVGRADES (Civics Grades in School)

For grade 12, responses to item P802001 were recoded as:

Mostly A	= 4.0
Half A and half B	= 3.5
Mostly B	= 3.0
Half B and half C	= 2.5
Mostly C	= 2.0
Half C and half D	= 1.5
Mostly D	= 1.0
Mostly below D	= .5
Did not say that civics had been taken	= 2.0
Missing	= 2.0

CVA-ACT (Active in 4 Areas of Civics, 1, 2, 5, and 8)

For grade 12, items P802401, P802402, P802405, and P802408 asked students to indicate how often they participated in particular educational activities for civics class. These items were grouped together because of the similarity of responses to them. Each of the four items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
A few times a year	= 0
Never	= 0
Did not say that civics was taken	= 0
Missing	= 0

The sum of the four variables was then assigned the codes:

4	= 1
0-3	= 0

CVB-ACT (Active in 2 Areas of Civics, 3 and 10)

For grade 12, items P802403 and P802410 asked students to indicate how often they participated in particular educational activities for civics class. These items were grouped together because of the similarity of responses to them. Each of the two items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
A few times a year	= 0
Never	= 0
Did not say that civics was taken	= 0
Missing	= 0

The sum of the two variables was then assigned the codes:

2	= 1
0-1	= 0

CVC-ACT (Active in 2 Areas of Civics, 4 and 9)

For grade 12, items P802404 and P802409 asked students to indicate how often they participated in particular educational activities for civics class. These items were grouped together because of the similarity of responses them. Each of the two items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
A few times a year	= 0
Never	= 0
Did not say that civics was taken	= 0
Missing	= 0

The sum of the two variables was then assigned the codes:

2 = 1
0-1 = 0

CVD-ACT (Active in 2 Areas of Civics, 6 and 7)

For grade 12, items P802406 and P802407 asked students to indicate how often they participated in particular educational activities for civics class. These items were grouped together because of the similarity of responses to them. Each of the two items was recoded as:

Almost every day	= 1
Once or twice a week	= 1
Once or twice a month	= 0
A few times a year	= 0
Never	= 0
Did not say that civics was taken	= 0
Missing	= 0

The sum of the two variables was then assigned the codes:

2 = 1
0-1 = 0

CVD-NOT (Not Very Active in These 2 Areas of Civics)

For grade 12, items P802406 and P802407 asked students to indicate how often they participated in particular educational activities for civics class. These items were grouped together because of the similarity of responses to them. Each of the two items was recoded as:

Almost every day	= 0
Once or twice a week	= 0
Once or twice a month	= 0
A few times a year	= 1
Never	= 1
Did not say that civics was taken	= 1
Missing	= 1

The sum of the two variables was then assigned the codes:

2 = 1
0-1 = 0

RDDIF (At Least Some Difficulty in Reading or Missing)

For grade 12, responses to item P802301 were recoded as:

A lot	= 1
Some	= 1
None	= 0
Did not say that civics was taken	= 1
Missing	= 1

Civics Trend Variables

PPARED (Parent's Education - 2nd Set of Categories)

For ages 13 and 17, a parents' education variable was derived for 1982 data, so that comparisons with 1988 data be made for students having a parent who graduated from college. This variable is based on the 1982 variables PARED, FCLGRAD, and MCLGRAD (whether the father or mother graduated from college). The categories for PPARED are coded as:

1	< High school	if PARED = 1 (<HS)
2	High school graduate	if PARED = 2 (HS Grad)
3	Post high school	if PARED = 3 (post-HS), FCLGRAD = 2(No), and MCLGRAD = 2 (No)
4	College graduate	if PARED = 3 and either FCLGRAD = 1 (Yes) or MCLGRAD = 1 (Yes)
	Missing	if otherwise

U.S. History Cross-sectional Variables

HIS TAKE (How often do you have a social studies class?)

For grade 4, responses to item S008701 were recoded as follows:

Never, hardly ever or missing	= 0
Less than once a week	= 1
One or two times a week	= 2
Three or four times a week	= 3
Every day	= 4

STUD.HIS (Have you ever studied the history of our country?)

For grade 4, responses to item H800601 were recoded as follows:

No or missing = 0
Yes = 1

STUD.EXP (Have you studied the early explorers of the new world?)

For grade 4, responses to item H801101 were recoded as follows:

Not at all or missing = 0
Some = 1
A lot = 2

STUD.IND (Have you studied the Indians who first lived in our country?)

For grade 4, responses to item H801201 were recoded as follows:

Not at all or missing = 0
Some = 1
A lot = 2

STUD.PIL (How much have you studied about the Pilgrims and the first colonies in America?)

For grade 4, responses to item H801301 were recoded as follows:

Not at all or missing = 0
Some = 1
A lot = 2

STUD.GW (How much have you studied George Washington and the war to gain independence from England?)

For grade 4, responses to item H801401 were recoded as follows:

Not at all or missing = 0
Some = 1
A lot = 2

STUD.PIO (How much have you studied the pioneers who settled the western part of the country?)

For grade 4, responses to item H801501 were recoded as follows:

Not at all or missing	= 0
Some	= 1
A lot	= 2

STUD.SLA (How much have you studied slavery and the war between the Northern and Southern states?)

For grade 4, responses to item H801601 were recoded as follows:

Not at all or missing	= 0
Some	= 1
A lot	= 2

STUD.INV (How much have you studied about people who invented things and made new discoveries?)

For grade 4, responses to item H801701 were recoded as follows:

Not at all or missing	= 0
Some	= 1
A lot	= 2

HIS TAKE (Did you study or expect to study U.S. history in the following grades?)

For grade 8, responses to items H800701 to H800704 (study in fifth, sixth, seventh, eighth grade) were recoded as follows:

Yes	= 1
No	= 2
I don't know	= 3

The variable HIS TAKE was then coded as the sum of all "Yes" responses, ranging from 0 to 4.

EARLY H (How much have you studied these periods of U.S. history?)

For grade 8, responses to items H801801 to H801806 were recoded as follows:

A lot	= 2
Some	= 1
Not at all	= 0

The average of the six recoded variables formed the variable EARLY H, whose values ranged from 0 to 2.

POST CH (How much have you studied these periods of U.S. history?)

For grade 8, responses to items H801807 to H801811 were recoded as follows:

A lot	= 2
Some	= 1
Not at all	= 0

The average of the five recoded variables formed the variable POST CH, whose values ranged from 0 to 2.

TEACH H (How often does your social studies teacher ask you to do these things?)

For grade 8, responses to items H802501 to H802511 were recoded as follows:

Almost every day	= 4
Once or twice a week	= 3
Once or twice a month	= 2
A few times a year	= 1
Never	= 0

The average of the 11 recoded variables formed the variable TEACH H, whose values ranged from 0 to 4.

TEXTBOOK (How much difficulty do you have reading your social studies textbooks?)

For grade 8, responses to item S008801 were recoded as follows:

A lot	= 2
Some	= 1
None	= 0

HIS TAKE (Did you study or expect to study U.S. history in the following grades?)

For grade 12, responses to items H800801 to H800804 (ninth, tenth, eleventh, twelfth grade) were recoded as follows:

Yes	= 1
No	= 2
I don't know	= 3

The variable HIS TAKE was then coded as the sum of all "Yes" responses, ranging from 0 to 4.

HIS COUR (Since the beginning of ninth grade, how much U.S. history coursework have you completed up to now?)

For grade 12, responses to item H800901 were recoded as follows:

None	= 00
Less than half a year	= 25
Half a year	= 50
Between half a year and one year	= 75
One year	= 100
More than one year	= 200

AP HIST (Have you taken or are you taking Advanced Placement American history?)

For grade 12, responses to item H801001 were recoded as follows:

Yes	= 1
No	= 0

AM HIST (Since the beginning of ninth grade, how much have you studied the following periods of U.S. history?)

For grade 12, responses to items H801901-H801909 and H801913 were recoded as follows:

A lot	= 2
Some	= 1
Not at all	= 0

The average of the 10 recoded variables formed the variable AM HIST, whose values ranged from 0 to 2.

WORLD H (Since the beginning of ninth grade, how much have you studied the following periods of U.S. history?)

For grade 12, responses to items H801910-H801912 were recoded as follows:

A lot	- 2
Some	- 1
Not at all	- 0

The average of the three recoded variables formed the variable WORLD H, whose values ranged from 0 to 2.

MIN/WOM (To what extent have you studied the history and contributions of minorities and women?)

For grade 12, responses to item H802001 were recoded as follows:

A lot	- 2
Some	- 1
Not at all	- 0

LIKE H (How much do you like studying United States history?)

For grade 12, responses to item H802101 were recoded as follows:

It is one of my favorite subjects	- 3
It is interesting	- 2
I like several other subjects better	- 1
I have never studied United States history	- 0

HAD HIST (Have you ever had a United States history class?)

For grade 12, responses to item H802201 were recoded as follows:

Yes	- 1
No	- 0

HOMEWK H (How much time have you usually spent each week on homework for your U.S. history class?)

For grade 12, responses to item H802301 were recoded as follows:

I usually haven't had homework assigned	= 0
I have had homework, but I usually haven't done it	= 0
Less than 1 hour	= .5
1 hour	= 1
2 hours	= 2
3 hours	= 3
4 hours	= 4
5 hours or more	= 5

GRADES H (What kind of grades have you gotten in your U. S. history class?)

For grade 12, responses to item H802401 were recoded as follows:

Mostly A	= 4.0
About half A and half B	= 3.5
Mostly B	= 3.0
About half B and half C	= 2.5
Mostly C	= 2.0
About half C and half D	= 1.5
Mostly D	= 1.0
Mostly below D	= 0.0

TEACH/H (How often has your U.S. history teacher asked you to do the following things in class?)

For grade 12, responses to items H802601-H802611 were recoded as follows:

Almost every day	= 4
Once or twice a week	= 3
Once or twice a month	= 2
A few times a year	= 1
Never	= 0

The average of the 11 recoded variables formed the variable TEACH/H, whose values ranged from 0 to 4.

TEXTBOOK (How much difficulty have you had reading your U. S. history textbook?)

For grade 12, responses to item H802801 were recoded as follows:

A lot	= 2
Some	= 1
None	= 0

Geography Cross-sectional Variables (Grade 12)

GEO TAKE (Did you take or do you expect to take a geography course in the following grades?)

Responses to items G800301-G800304 (ninth, tenth, eleventh, twelfth grades) were recoded as follows:

Yes	= 1
No	= 2
I don't know	= 3

The variable GEO TAKE was then coded as the sum of all "Yes" responses, ranging from 0 to 4.

PHYSICAL (How much have you studied the following topics in geography?)

Responses to items G800201-G800209 (physical geography topics) were recoded as follows:

A lot	= 3
Some	= 2
Very little	= 1
Not at all	= 0

The average of the nine recoded variables formed the variable PHYSICAL, whose values ranged from 0 to 3.

SOCIAL (How much have you studied the following topics in geography?)

Responses to items G800210-G800215 (social geography topics) were recoded as follows:

A lot	= 3
Some	= 2
Very little	= 1
Not at all	= 0

The average of the five recoded variables formed the variable SOCIAL, whose values ranged from 0 to 3.

The following nine variables were created from responses to nine items that stemmed from the question "Since the beginning of ninth grade, how much of the following geography coursework have you completed up to now?"

W GEO	(World geography coursework taken)
W HIS/G	(World history and geography coursework taken)
STAT/REG	(State/regional geography coursework taken)
US GEO	(United States geography coursework taken)
US H/G	(United States history and geography coursework taken)
PHY/GEO	(Physical geography/earth science coursework taken)
ECON/P	(Economic and political geography coursework taken)
HUM/CULT	(Human and cultural geography coursework taken)
URBAN G	(Urban geography coursework taken)

These variables were created, respectively, from items G800101-G800109. Responses to these items were first recoded as:

Yes, in grade 9	= 1
Yes, in grade 10	= 1
Yes, in grade 11	= 1
Yes, in grade 12	= 1
No, I haven't taken	= 0

To form the new variables, the recoded responses were then collapsed as follows:

Any "Yes"	= 1
No, I haven't taken	= 0

Mathematics Trend Variables

Observed Race x Observed Sex

This variable is defined as the following combinations of the variables RACE and SEX:

	RACE	SEX
1	White, American Indian, Other, or Unclassified	Male
2	Black	Male
3	Hispanic	Male
4	Asian American	Male
5	White, American Indian, Other, or Unclassified	Female
6	Black	Female
7	Hispanic	Female
8	Asian American	Female

Observed Race x Parents' Education

This variable is defined as the following combinations of the variables RACE and PARED:

	RACE	PARED
1	White, American Indian, Other, or Unclassified	Less than high school education
2	White, American Indian, Other, or Unclassified	High school graduate
3	White, American Indian, Other, or Unclassified	Some education after high school
4	White, American Indian, Other, or Unclassified	College graduate
5	White, American Indian, Other, or Unclassified	Missing
6	Black	Less than high school education
7	Black	High school graduate
8	Black	Some education after high school
9	Black	College graduate
10	Black	Missing
11	Hispanic	Less than high school education
12	Hispanic	High school graduate
13	Hispanic	Some education after high school
14	Hispanic	College graduate
15	Hispanic	Missing
16	Asian American	Less than high school education
17	Asian American	High school graduate
18	Asian American	Some education after high school
19	Asian American	College graduate
20	Asian American	Missing

Observed Race x Language in the Home

Items B003201 (age 9) and B003301 (ages 13 and 17) asked students how often people in their home speak a language other than English. This conditioning variable is defined by the following combinations of the variables RACE and B003201 (for age 9) or RACE and B003301 (ages 13 and 17):

Age 9

	RACE	B003201
1	White, American Indian, Other, or Unclassified	Always
2	White, American Indian, Other, or Unclassified	Sometimes
3	White, American Indian, Other or Unclassified	Never
4	Black	Always
5	Black	Sometimes
6	Black	Never
7	Hispanic	Always
8	Hispanic	Sometimes
9	Hispanic	Never
10	Asian American	Always
11	Asian American	Sometimes
12	Asian American	Never

Ages 13 and 17

	RACE	B003301
1	White, American Indian, Other, or Unclassified	Always or mostly
2	White, American Indian, Other, or Unclassified	Occasionally or half the time
3	White, American Indian, Other, or Unclassified	Never
4	Black	Always or mostly
5	Black	Occasionally or half the time
6	Black	Never
7	Hispanic	Always or mostly
8	Hispanic	Occasionally or half the time
9	Hispanic	Never
10	Asian American	Always or mostly
11	Asian American	Occasionally or half the time
12	Asian American	Never

APPENDIX E

Revision of 1984 NAEP Poststratification Weights for Grade 4/Age 9 and Grade 8/Age 13

Appendix E

REVISION OF 1984 NAEP POSTSTRATIFICATION WEIGHTS FOR GRADE 4/AGE 9 AND GRADE 8/AGE 13

Keith F. Rust

Westat, Inc.

A comparison of the proportions of 9-year-old students who were in grade 4, based on weighted data, revealed an inconsistency between the 1984 main sample results and those for bridge studies in subsequent years. In 1984, the percentage of 9-year-old students in grade 4 was 74.9. For three subsequent bridges, the percentage ranged from 62.6 to 66.1.

A consideration of the method of obtaining the separate poststratification factors for those students both grade and age eligible, those eligible by age alone, and those eligible by grade alone, used in 1984 but not for subsequent bridges, revealed the possibility of improving the approach used to derive the independent estimates which constitute the major component of the numerators of each poststratification factor. This improvement pertained to the poststratification procedure for grade 4/age 9 and grade 8/age 13, but not grade 11/age 17.

The possibility of improvement arose because the independent estimates were derived using Current Population Survey data on the distribution over grades of the population by whole years of age. These ages are as of early October, the time each year the Current Population Survey in which this information is collected is conducted. The age definition for ages 9 and 13 used in 1984 means that this distribution is required as of January 1. (For age 17, and for all three ages for the main samples in 1986, the appropriate date is October 1, consistent with the Current Population Survey data, consequently, the problem of inconsistency between NAEP and the Current Population Survey did not occur for these cases.)

Evidence from the 1984 and 1988 NAEP samples shows clearly that the proportion of 9-year-olds who were in grade 4 and 13-year-olds who were in grade 8 declined between October 1 and the following January 1. That is, there were more fourth graders who had their tenth birthday during this period than there were fourth graders who had their ninth birthday. The difference was sufficiently great as to decrease the percentage of 9-year-olds who were age-eligible by about 10 percentage points. A similar but less marked decrease also occurred at age 13.

Independent estimates and the resulting poststratification factors were recomputed in a way that recognized this shift. The magnitude in the shift was estimated from NAEP data, this being the only source of information available. We note that the shift proved very consistent between the 1984 and 1988 samples, when the same age and grade definitions were used.

The 1988 poststratification procedure, which differed from that used in 1984 and 1986 in a number of ways, was performed in a manner that also accounted for this shift in the grade/age distribution. Hence, no revision of the 1988 poststratification factors is required.

APPENDIX F

1988 NAEP IRT Parameters

Appendix F

1988 NAEP IRT PARAMETERS

Appendix F contains 13 tables of IRT (item response theory) parameters for NAEP items that were used in each subject area and study (cross-sectional or trend) for which IRT scales were created.

For each NAEP item used in scaling, the tables show the corresponding IRT parameters (A, B, and C) and standard errors (S.E.), the block in which the item appears for each age class (BLOCK), and the order in which the item appears within the block (ITEM).

IRT parameters for items used in cross-sectional scales are shown for reading in Table F-1, for civics in Table F-3, for U.S. history in Table F-6, and for geography in Table F-7. IRT parameters for items used in trend scales are shown for reading in Table F-2, for civics in Tables F-4 and F-5, for mathematics in Tables F-8 to F-10, and for science in Tables F-11 to F-13.

Note that item parameters shown in this appendix are in the metrics used for the original calibration of the scale. The transformations needed to represent these parameters in terms of the metric of the final reporting scales are given in Chapters 10 through 1[illegible].

Table F-1

1988 IRT Parameters, Reading Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
N001501	1.733	(0.028)	-1.036	(0.029)	0.221	(0.010)	R3	4	R3	4	--	--
N001502	2.190	(0.064)	-0.211	(0.023)	0.256	(0.007)	R3	5	R3	5	--	--
N001503	1.732	(0.046)	-0.536	(0.026)	0.319	(0.008)	R3	6	R3	6	--	--
N001504	1.798	(0.052)	-0.240	(0.022)	0.305	(0.007)	R3	7	R3	7	--	--
N001601	0.589	(0.017)	-0.804	(0.027)	0.181	(0.009)	R4	4	R4	4	--	--
N001602	1.387	(0.040)	-0.346	(0.020)	0.258	(0.007)	R4	5	R4	5	--	--
N001603	1.305	(0.039)	0.325	(0.022)	0.312	(0.006)	R4	6	R4	6	--	--
N001604	1.400	(0.041)	0.182	(0.019)	0.272	(0.006)	R4	7	R4	7	--	--
N002001	1.775	(0.051)	0.163	(0.023)	0.176	(0.009)	--	--	R4	8	--	--
N002002	1.553	(0.048)	0.155	(0.023)	0.234	(0.009)	--	--	R4	9	--	--
N002003	1.878	(0.073)	-0.091	(0.027)	0.254	(0.011)	--	--	R4	10	--	--
N002101	1.113	(0.018)	0.830	(0.021)	0.207	(0.006)	--	--	R5	13	R5	13
N002102	1.922	(0.022)	0.838	(0.022)	0.188	(0.005)	--	--	R5	14	R5	14
N002401	1.285	(0.059)	0.027	(0.023)	0.148	(0.007)	R4	1	--	--	--	--
N002401	1.383	(0.052)	-0.399	(0.028)	0.225	(0.012)	--	--	R4	1	--	--
N002801	1.803	(0.039)	-0.579	(0.023)	0.186	(0.008)	R3	2	R3	2	--	--
N002802	1.591	(0.032)	-0.735	(0.024)	0.196	(0.009)	R3	3	R3	3	--	--
N003001	1.605	(0.020)	1.026	(0.023)	0.210	(0.005)	--	--	R5	10	R5	10
N003002	0.509	(0.015)	0.649	(0.022)	0.141	(0.007)	--	--	R5	11	R5	11
N003003	2.506	(0.021)	1.456	(0.0[illegible])	0.096	(0.003)	--	--	R5	12	R5	12
N003101	1.335	(0.050)	-0.424	(0.029)	0.232	(0.014)	--	--	R3	8	--	--
N003102	1.739	(0.062)	-0.150	(0.026)	0.201	(0.012)	--	--	R3	9	--	--
N003201	1.312	(0.035)	-0.463	(0.022)	0.252	(0.011)	--	--	R5	1	R5	1
N003202	1.552	(0.049)	0.204	(0.025)	0.271	(0.009)	--	--	R5	2	--	--
N003202	1.139	(0.041)	-0.114	(0.025)	0.316	(0.015)	--	--	--	--	R5	2
N003203	1.165	(0.022)	0.390	(0.018)	0.298	(0.007)	--		R5	3	R5	3
N003204	1.162	(0.039)	0.497	(0.029)	0.285	(0.008)	--	--	R5	4	--	--
N003204	1.494	(0.056)	-0.071	(0.029)	0.368	(0.014)	--	--	--	--	R5	4
N003501	0.998	(0.024)	0.146	(0.020)	0.187	(0.014)	--	--	--	--	R3	7
N003701	0.941	(0.024)	-0.768	(0.025)	0.293	(0.009)	R4	2	R4	2	--	--
N003702	0.968	(0.032)	-0.182	(0.017)	0.318	(0.007)	R4	3	R4	3	--	--
N004201	1.454	(0.029)	0.218	(0.018)	0.295	(0.007)	--	--	R5	5	R5	5
N004202	1.080	(0.024)	0.606	(0.023)	0.410	(0.006)	--	--	R5	6	R5	6
N004801	1.025	(0.025)	-1.112	(0.033)	0.339	(0.011)	R3	1	R3	1	--	--
N005001	2.826	(0.025)	1.223	(0.032)	0.233	(0.004)	--	--	R5	7	R5	7
N005002	1.516	(0.022)	1.215	(0.030)	0.307	(0.005)	--	--	R5	8	R5	8
N005003	1.143	(0.018)	1.528	(0.031)	0.141	(0.004)	--	--	R5	9	R5	9
N005101	0.716	(0.018)	-2.688	(0.075)	0.237	(0.025)	R2	2	--	--	--	--
N005503	0.835	(0.019)	0.499	(0.019)	0.280	(0.007)	--	--	R2	1	R2	1
N005504	1.778	(0.021)	0.759	(0.020)	0.237	(0.006)	--	--	R2	2	R2	2
N005505	1.457	(0.040)	-0.582	(0.027)	0.300	(0.012)	--	--	R2	3	R2	3
N007301	1.403	(0.031)	0.020	(0.017)	0.276	(0.008)	--	--	R2	4	R2	4
N007302	1.097	(0.018)	0.717	(0.020)	0.254	(0.006)	--	--	R2	5	R2	5
N007303	1.703	(0.030)	0.224	(0.018)	0.232	(0.007)	--	--	R2	6	R2	6
N007304	1.117	(0.022)	0.352	(0.017)	0.264	(0.007)	--	--	R2	7	R2	7
N007305	0.951	(0.017)	0.671	(0.019)	0.212	(0.006)	--	-	R2	8	R2	8
N007306	1.368	(0.024)	0.184	(0.016)	0.206	(0.008)	-	--	R2	9	R2	9
N007401	2.069	(0.022)	0.748	(0.020)	0.159	(0.005)	--	--	R2	10	R2	10
N007402	1.601	(0.031)	0.254	(0.020)	0.288	(0.008)	--	--	R2	11	R2	11
N007403	2.302	(0.035)	0.353	(0.021)	0.203	(0.007)	--	--	R2	12	R2	12
N007404	1.852	(0.028)	0.531	(0.022)	0.296	(0.007)	--	--	R2	13	R2	13
N007405	1.432	(0.020)	1.183	(0.028)	0.220	(0.005)	--	--	R2	14	R2	14
N008701	0.715	(0.018)	-2.941	(0.079)	0.242	(0.026)	R2	1	--	--	--	--
N010501	1.525	(0.033)	-1.160	(0.039)	0.285	(0.012)	R2	8	--	--	--	--
N010502	0.913	(0.026)	-1.050	(0.037)	0.213	(0.012)	R2	9	--	--	--	--
N010503	1.479	(0.033)	-1.117	(0.039)	0.268	(0.012)	R2	10	--	--	--	--
N010504	1.843	(0.049)	-0.627	(0.032)	0.189	(0.010)	R2	11	--	--	--	--
N013101	1.706	(0.036)	-1.145	(0.040)	0.325	(0.012)	R2	4	--	--	--	--
N013102	1.251	(0.040)	-0.460	(0.025)	0.178	(0.009)	R2	5	--	--	--	--
N013103	1.038	(0.040)	-0.335	(0.025)	0.200	(0.009)	R2	6	--	--	--	--
N013104	0.700	(0.035)	-0.116	(0.022)	0.235	(0.009)	R2	7	--	--	--	--
N013301	1.143	(0.022)	-1.717	(0.044)	0.252	(0.017)	R2	3	--	--	--	--
N014301	1.609	(0.084)	-0.292	(0.032)	0.197	(0.008)	R2	12	--	--	--	--
N014302	1.511	(0.081)	-0.052	(0.029)	0.227	(0.008)	R2	13	--	--	--	--

Table F-1 (continued)

1988 IRT Parameters, Reading Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
N014303	2.049	(0.056)	-0.683	(0.038)	0.243	(0.010)	R2	14	--	--	--	--
N015901	1.331	(0.031)	0.389	(0.025)	0.250	(0.012)	--	--	--	--	R3	8
N015902	1.229	(0.025)	0.719	(0.028)	0.315	(0.010)	--	--	--	--	R3	9
N015903	2.311	(0.027)	1.020	(0.032)	0.226	(0.008)	--	--	--	--	R3	10
N015905	0.684	(0.019)	0.561	(0.023)	0.000	(0.000)	--	--	--	--	R3	11A
N016001	1.040	(0.031)	-0.168	(0.022)	0.210	(0.016)	--	--	--	--	R3	1
N016002	0.917	(0.022)	0.598	(0.024)	0.22[illegible]	(0.011)	--	--	--	--	R3	2
N016003	1.319	(0.029)	0.49[illegible]	(0.026)	0.283	(0.011)	--	--	--	--	R3	3
N016004	1.159	(0.030)	0.139	(0.022)	0.240	(0.014)	--	--	--	--	R3	4
N016005	2.167	(0.051)	0.240	(0.029)	0.247	(0.013)	--	--	--	--	R3	5
N016006	1.295	(0.027)	0.489	(0.024)	0.252	(0.011)	--	--	--	--	R3	6
N021101	0.757	(0.017)	-1.807	(0.046)	0.206	(0.016)	R7	4	--	--	--	--
N021102	0.943	(0.022)	-0.997	(0.031)	0.166	(0.011)	R7	5	--	--	--	--
N021103	1.096	(0.025)	-1.082	(0.033)	0.207	(0.011)	R7	6	--	--	--	--
N021401	1.337	(0.050)	-0.297	(0.026)	0.208	(0.008)	R7	12	--	--	--	--
N021402	2.202	(0.074)	-0.354	(0.032)	0.214	(0.007)	R7	13	--	--	--	--
N021403	1.425	(0.032)	-1.130	(0.038)	0.282	(0.012)	R7	14	--	--	--	--
N021404	2.006	(0.113)	0.081	(0.032)	0.205	(0.006)	R7	15	--	--	--	--
N021501	0.859	(0.025)	-0.780	(0.031)	0.246	(0.010)	R5	1	--	--	--	--
N021502	1.116	(0.02[illegible])	-1.607	(0.037)	0.201	(0.014)	R5	2	--	--	--	--
N021503	1.111	(0.019)	-1.975	(0.043)	0.223	(0.018)	R5	3	--	--	--	--
N021504	1.036	(0.030)	-0.706	(0.029)	0.248	(0.009)	R5	4	--	--	--	--
N021505	1.582	(0.033)	-0.837	(0.030)	0.198	(0.009)	R5	5	--	--	--	--
N021601	0.781	(0.026)	-0.009	(0.021)	0.291	(0.013)	--	--	--	--	R4	15
N021602	0.708	(0.019)	1.098	(0.035)	0.188	(0.009)	--	--	--	--	R4	16
N021603	0.631	(0.032)	2.241	(0.119)	0.381	(0.007)	--	--	--	--	R4	17
N021604	1.488	(0.033)	0.393	(0.026)	0.278	(0.011)	--	--	--	--	R4	18
N021605	1.242	(0.023)	0.869	(0.029)	0.276	(0.009)	--	--	--	--	R4	19
R000101	1.095	(0.024)	-0.970	(0.030)	0.176	(0.010)	R5	6	--	--	--	--
R000102	1.391	(0.064)	-0.035	(0.024)	0.200	(0.007)	R5	7	--	--	--	--
R000103	1.702	(0.043)	-0.699	(0.032)	0.276	(0.009)	R5	8	--	--	--	--
R000104	1.766	(0.031)	-1.362	(0.041)	0.313	(0.012)	R5	9	--	--	--	--
R000201	0.983	(0.030)	-0.839	(0.034)	0.304	(0.010)	R5	10	--	--	--	--
R000202	2.351	(0.044)	-0.873	(0.037)	0.252	(0.009)	R5	11	--	--	--	--
R000203	1.462	(0.027)	-1.363	(0.039)	0.273	(0.013)	R5	12	--	--	--	--
R000204	2.011	(0.035)	-1.165	(0.040)	0.273	(0.011)	R5	13	--	--	--	--
R000205	1.389	(0.027)	-1.305	(0.038)	0.255	(0.013)	R5	14	--	--	--	--
R000206	1.068	(0.039)	0.130	(0.021)	0.000	(0.000)	R5	15A	--	--	--	--
R000301	1.510	(0.053)	-0.355	(0.028)	0.257	(0.008)	R6	1	--	--	--	--
R000302	1.134	(0.027)	-0.862	(0.029)	0.191	(0.010)	R6	2	--	--	--	--
R000303	1.076	(0.034)	-0.567	(0.029)	0.258	(0.009)	R6	3	--	--	--	--
R000304	1.035	(0.027)	-0.895	(0.032)	0.241	(0.010)	R6	4	--	--	-	--
R000401	1.504	(0.056)	-0.260	(0.026)	0.217	(0.007)	R6	5	--	--	--	--
R000402	0.848	(0.017)	-1.994	(0.047)	0.222	(0.017)	R6	6	--	--	--	--
R000403	1.052	(0.054)	0.112	(0.025)	0.238	(0.007)	R6	7	--	--	--	--
R000404	1.336	(0.025)	-1.258	(0.034)	0.186	(0.012)	R6	8	--	--	--	--
R000501	1.583	(0.030)	-1.197	(0.036)	0.236	(0.011)	R6	9	--	--	--	--
R000502	1.656	(0.060)	-0.395	(0.030)	0.270	(0.008)	R6	10	--	--	--	--
R000503	1.194	(0.067)	0.449	(0.037)	0.162	(0.006)	R6	11	--	--	--	--
R000504	0.726	(0.027)	-0.673	(0.032)	0.261	(0.010)	R6	12	--	--	--	--
R000505	1.435	(0.054)	-0.397	(0.030)	0.245	(0.008)	R6	13	--	--	--	--
R000601	1.079	(0.022)	-1.326	(0.035)	0.197	(0.012)	R7	1	--	--	--	--
R000602	1.029	(0.028)	-0.829	(0.031)	0.240	(0.010)	R7	2	--	--	--	--
R000603	1.354	(0.030)	-0.973	(0.032)	0.231	(0.010)	R7	3	--	--	--	--
R000701	1.338	(0.056)	-0.054	(0.023)	0.171	(0.007)	R7	7	--	--	--	--
R000702	1.767	(0.076)	-0.152	(0.027)	0.216	(0.007)	R7	8	--	--	--	--
R000703	1.472	(0.030)	-1.083	(0.035)	0.257	(0.011)	R7	9	--	--	--	--
R000704	0.556	(0.017)	-1.821	(0.058)	0.221	(0.015)	R7	10	--	--	--	--
R000705	0.661	(0.026)	0.537	(0.030)	0.273	(0.010)	R7	11	--	--	--	--
R000801	0.741	(0.015)	-1.541	(0.036)	0.231	(0.012)	R8	6	R8	10	--	--
R000802	1.206	(0.019)	-1.311	(0.028)	0.195	(0.011)	R8	7	R8	11	--	--
R000803	0.941	(0.022)	-0.452	(0.017)	0.164	(0.007)	R8	8	R8	12	--	--
R000804	1.631	(0.035)	-0.402	(0.019)	0.139	(0.006)	R8	9	R8	13	--	--
R000805	1.462	(0.031)	-0.702	(0.024)	0.226	(0.008)	R8	10	R8	14	--	--

Table F-1 (continued)

1988 IRT Parameters, Reading Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
R000806	1.647	(0.045)	-0.799	(0.036)	0.537	(0.008)	R8	11A	R8	15A	--	--
R000807	0.881	(0.016)	-1.090	(0.025)	0.000	(0.000)	R8	11B	R8	15B	--	--
R000901	1.531	(0.025)	-0.797	(0.021)	0.200	(0.008)	R8	1	R8	1	R8	1
R000902	1.185	(0.024)	-0.808	(0.023)	0.297	(0.008)	R8	2	R8	2	R8	2
R000903	1.523	(0.022)	-1.074	(0.023)	0.165	(0.009)	R8	3	R8	3	R8	3
R000904	2.829	(0.106)	0.473	(0.045)	0.121	(0.004)	R8	4	--	--	--	--
R000904	1.973	(0.043)	0.009	(0.018)	0.176	(0.008)	--	--	R8	4	R8	4
R000905	1.448	(0.037)	-0.928	(0.033)	0.199	(0.015)	R8	5	R8	5	R8	5
R001001	1.172	(0.043)	-0.791	(0.039)	0.273	(0.015)	--	--	R8	6	--	--
R001002	1.336	(0.047)	-0.166	(0.023)	0.250	(0.011)	--	--	R8	7	--	--
R001003	1.498	(0.047)	0.144	(0.022)	0.197	(0.009)	--	--	R8	8	--	--
R001004	1.499	(0.052)	-0 165	(0.024)	0.223	(0.010)	--	--	R8	9	--	--
R001101	1.375	(0.034)	0.932	(0.035)	0.132	(0.006)	--	--	R6	16	--	--
R001102	1.797	(0.046)	0.729	(0.039)	0.351	(0.007)	--	--	R6	17	--	--
R001103	1.089	(0.035)	0.570	(0.029)	0.190	(0.008)	--	--	R6	18	--	--
R001201	1.984	(0.067)	-0.221	(0.026)	0.178	(0.010)	--	--	R6	6	--	--
R001202	1.819	(0.062)	0.038	(0.026)	0.294	(0.009)	--	--	R6	7	--	--
R001203	1.695	(0.061)	-0.356	[illegible]	0.197	(0.011)	--	--	R6	8	--	--
R001204	1.574	(0.049)	-0.123	(0.022)	0.159	(0.010)	--	--	R6	9	--	--
R001205	1.749	(0.059)	-0.541	(0.034)	0.250	(0.013)	--	--	R6	10	--	--
R001301	1.020	(0.034)	-0.080	(0.019)	0.189	(0.010)	--	--	R6	1	--	--
R001302	1.563	(0.051)	0.301	(0.027)	0.333	(0.008)	--	--	R6	2	--	--
R001303	2.042	(0.081)	-0.289	(0.032)	0.322	(0.011)	--	--	R6	3	--	--
R001304	2.428	(0.041)	0.924	(0.038)	0.197	(0.005)	--	--	R6	4	--	--
R001305	1.540	(0.057)	-0.457	(0.031)	0.282	(0.012)	--	--	R6	5	--	--
R001401	1.863	(0.057)	0.086	(0.024)	0.189	(0.009)	--	--	R6	11	--	--
R001402	2.344	(0.085)	-0.007	(0.030)	0.268	(0.009)	--	--	R6	12	--	--
R001403	2.321	(0.053)	0.398	(0.031)	0.262	(0.008)	--	--	R6	13	--	--
R001404	2.271	(0.091)	0.090	(0.031)	0.308	(0.009)	--	--	R6	14	--	--
R001405	0.773	(0.032)	0.057	(0.020)	0.238	(0.010)	--	--	R6	15	--	--
R001501	0.809	(0.021)	-0.974	(0.030)	0.202	(0.013)	--	--	R7	1	R7	1
R001502	0.612	(0.015)	0.182	(0.013)	0.146	(0.008)	--	--	R7	2	R7	2
R001503	1.199	(0.033)	-0.909	(0.033)	0.238	(0.014)	--	--	R7	3	R7	3
R001504	1.460	(0.031)	-0.032	(0.017)	0.261	(0.008)	--	--	R7	4	R7	4
R001601	1.575	(0.037)	-0.642	(0.026)	0.166	(0.012)	--	--	R7	5	R7	5
R001602	1.406	(0.037)	-0.401	(0.021)	0.207	(0.010)	--	--	R7	6	R7	6
R001603	1.245	(0.029)	-0.097	(0.016)	0.240	(0.009)	--	--	R7	7	R7	7
R001604	[illegible]	(0.020)	-0.951	(0.029)	0.172	(0.013)	--	--	R7	8	R7	8
R001605	1.039	(0.019)	0.463	(0.016)	0.207	(0.006)	--	--	R7	9	R7	9
R001701	1.132	(0.025)	0.153	(0.017)	[illegible]	(0.008)	--	--	R7	10	R7	10
R001702	1.936	(0.058)	-0.498	(0.028)	0.261	(0.011)	--	--	R7	11	R7	11
R001703	0.670	(0.016)	0.939	(0.026)	0.180	(0.006)	--	--	R7	12	R7	12
R001704	1.057	(0.019)	0.568	(0.018)	0.224	(0.006)	--	--	R7	13	R7	13
R001801	1.493	(0.033)	-0.017	(0.017)	0.210	(0.008)	--	--	R7	14	R7	14
R001802	2.274	(0.031)	0.365	(0.019)	0.149	(0.006)	--	--	R7	15	R7	15
R001803	1.087	(0.018)	0.771	(0.021)	0.235	(0.006)	--	--	R7	16	R7	16
R001804	2.053	(0.035)	0.308	(0.021)	0.294	(0.007)	--	--	R7	17	R7	17
R001805	1.840	(0.028)	0.355	(0.019)	0.225	(0.007)	--	--	R7	18	R7	18
R001806	0.955	(0.020)	0 184	(0.015)	0.215	(0.008)	--	--	R7	19	R7	19
R001901	0.692	(0.031)	-1.135	(0.057)	0.237	(0.022)	--	--	--	--	R4	10
R001902	1.535	(0.036)	0.250	(0 024)	0.241	(0.012)	--	--	--	--	R4	11
R001903	1.172	(0.027)	0.255	(0.022)	0.226	(0.012)	--	--	--	--	R4	12
R001904	1.864	(0.048)	0.130	(0.027)	0.273	(0.013)	--	--	--	--	R4	13
R001905	1.307	(0.039)	-0.237	(0.026)	0.191	(0.017)	--	--	--	--	R4	14
R002001	1.762	(0.057)	-0.080	(0.029)	0.272	(0.015)	--	--	--	--	R4	1
R002002	1.099	(0.025)	0.412	(0.023)	0.259	(0.011)	--	--	--	--	R4	2
R002003	1.463	(0.030)	0.329	(0.023)	0.198	(0.011)	--	--	--	--	R4	3
R002004	1.179	(0.032)	0.115	(0.023)	0.285	(0.013)	--	--	--	--	R4	4
R002005	0.837	(0.022)	0.716	(0.028)	0.279	(0.010)	--	--	--	--	R4	5
R002101	0.442	(0.026)	2.427	(0.145)	0.305	(0.007)	--	--	--	--	R4	6
R002102	1.160	(0.024)	0.363	(0.021)	0.169	(0.011)	--	--	--	--	R4	7
R002103	1.528	(0.047)	-0.162	(0.027)	0.213	(0.016)	--	--	--	--	R4	8
R002104	1.117	(0.030)	-0.030	(0.022)	0.245	(0.014)	--	--	--	--	R4	9
R002201	1.324	(0.027)	0.951	(0.034)	0.417	(0.008)	--	--	--	--	R6	1

Table F-1 (continued)

1988 IRT Parameters, Reading Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
R002202	1.123	(0.051)	-0.911	(0.053)	0.234	(0.024)	--	--	--	--	R6	2
R002203	1.107	(0.023)	0.535	(0.023)	0.238	(0.010)	--	--	--	--	R6	3
R002204	0.805	(0.020)	1.279	(0.039)	0.227	(0.008)	--	--	--	--	R6	4
R002205	1.078	(0.043)	2.336	(0.111)	0.319	(0.006)	--	--	--	--	R6	5
R002301	2.070	(0.033)	0.569	(0.027)	0.244	(0.009)	--	--	--	--	R6	6
R002302	0.683	(0.020)	0.839	(0.031)	0.232	(0.010)	--	--	--	--	R6	7
R002303	1.982	(0.030)	0.526	(0.025)	0.178	(0.009)	--	--	--	--	R6	8
R002304	2.068	(0.036)	0.411	(0.026)	0.203	(0.010)	--	--	--	--	R6	9
R002401	1.281	(0.057)	-0.479	(0.037)	0.287	(0.019)	--	--	--	--	R6	10
R002402	0.777	(0.026)	1.820	(0.069)	0.335	(0.007)	--	--	--	--	R6	11
R002403	1.383	(0.044)	-0.054	(0.027)	0.317	(0.015)	--	--	--	--	R6	12
R002404	1.427	(0.021)	1.073	(0.028)	0.197	(0.008)	--	--	--	--	R6	13
R002405	1.482	(0.029)	0.550	(0.026)	0.262	(0.010)	--	--	--	--	R6	14
R002406	0.808	(0.017)	1.011	(0.027)	0.000	(0.000)	--	--	--	--	R6	15A
R002501	1.564	(0.026)	0.547	(0.024)	0.227	(0.010)	--	--	--	--	R8	6
R002502	1.150	(0.026)	0.364	(0.022)	0.251	(0.011)	--	--	--	--	R8	7
R002503	1.287	(0.037)	-0.190	(0.025)	0.205	(0.016)	--	--	--	--	R8	8
R002601	1.442	(0.026)	0.581	(0.024)	0.238	(0.010)	--	--	--	--	R8	9
R002602	1.402	(0.038)	0.102	(0.025)	0.295	(0.013)	--	--	--	--	R8	10
R002603	1.539	(0.056)	-0.146	(0.029)	0.294	(0.016)	--	--	--	--	R8	11
R002604	1.338	(0.024)	0.574	(0.022)	0.163	(0.010)	--	--	--	--	R8	12
R002605	1.678	(0.026)	0.826	(0.028)	0.265	(0.008)	--	--	--	--	R8	13
R002701	1.462	(0.029)	0.269	(0.022)	0.144	(0.012)	--	--	--	--	R8	14
R002702	1.327	(0.021)	1.132	(0.029)	0.187	(0.007)	--	--	--	--	R8	15
R002703	2.013	(0.036)	0.440	(0.027)	0.205	(0.011)	--	--	--	--	R8	16
R002704	0.971	(0.021)	0.784	(0.026)	0.208	(0.010)	--	--	--	--	R8	17
R002705	1.940	(0.026)	1.460	(0.039)	0.221	(0.006)	--	--	--	--	R8	18

Table F-2

1988 IRT Parameters, Reading Bridge to 1984

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 11/AGE 17 BLOCK	GRADE 11/AGE 17 ITEM
N001101	0.344	(0.048)	-0.384	(0.104)	0.291	(0.053)	H	5	H	6	--	--
N001201	0.712	(0.183)	1.144	(0.403)	0.369	(0.053)	--	--	H	7	--	--
N001202	1.276	(0.18[illegible])	0.585	(0.197)	0.258	(0.037)	--	--	H	8	--	--
N001301	0.986	(0.11[illegible])	0.495	(0.173)	0.409	(0.043)	--	--	H	9	H	10
N001302	0.720	(0.087)	-1.548	(0.241)	0.497	(0.097)	--	--	H	10	H	11
N001303	1.534	(0.133)	0.407	(0.131)	0.281	(0.030)	--	--	H	11	H	12
N001401	0.999	(0.094)	0.001	(0.114)	0.251	(0.055)	--	--	H	12	H	13
N001501	1.808	(0.130)	-1.313	(0.152)	0.225	(0.047)	H	10	H	13	H	14
N001502	1.643	(0.098)	-0.507	(0.061)	0.182	(0.026)	H	11	H	14	H	15
N001503	1.345	(0.088)	-0.202	(0.086)	0.207	(0.043)	H	12	H	15	H	16
N001504	1.448	(0.089)	-0.650	(0.068)	0.173	(0.032)	H	13	H	16	H	17
N001601	0.622	(0.041)	-0.959	(0.082)	0.133	(0.046)	J	12	J	11	--	--
N001602	1.263	(0.079)	-0.693	(0.066)	0.250	(0.031)	J	13	J	12	--	--
N001603	0.816	(0.073)	-0.031	(0.068)	0.233	(0.033)	J	14	J	13	--	--
N001604	1.375	(0.101)	0.111	(0.063)	0.269	(0.018)	J	15	J	14	--	--
N001701	0.981	(0.066)	-0.418	(0.090)	0.231	(0.059)	--	--	J	17	J	12
N001702	0.541	(0.116)	2.651	(0.621)	0.231	(0.028)	--	--	J	18	J	13
N001703	1.081	(0.080)	0.003	(0.096)	0.291	(0.044)	--	--	J	19	J	14
N001802	1.592	(0.140)	0.727	(0.131)	0.217	(0.012)	J	20	J	21	--	--
N001901	1.644	(0.111)	0.210	(0.098)	0.331	(0.028)	--	--	J	22	J	15
N002001	1.197	(0.065)	-0.013	(0.050)	0.131	(0.020)	K	9	K	9	K	9
N002002	1.444	(0.084)	-0.042	(0.055)	0.203	(0.020)	K	10	K	10	K	10
N002003	1.583	(0.093)	-0.229	(0.054)	0.224	(0.022)	K	11	K	11	K	11
N002101	0.941	(0.094)	1.171	(0.176)	0.247	(0.019)	K	18	K	12	K	12
N002102	1.495	(0.100)	0.840	(0.118)	0.147	(0.012)	K	19	K	13	K	13
N002201	1.704	(0.118)	-0.129	(0.078)	0.200	(0.037)	--	--	K	14	K	14
N002202	1.358	(0.120)	-0.349	(0.112)	0.337	(0.059)	--	--	K	15	K	15
N002203	0.783	(0.066)	-1.139	(0.137)	0.236	(0.086)	--	--	K	16	K	16
N002401	1.449	(0.096)	-0.375	(0.057)	0.128	(0.023)	L	22	L	22	--	--
N002501	0.550	(0.053)	0.129	(0.100)	0.205	(0.057)	--	--	L	23	L	27
N002701	1.024	(0.102)	0.833	(0.164)	0.234	(0.032)	--	--	L	24	L	28
N002702	1.148	(0.077)	0.055	(0.065)	0.141	(0.023)	L	20	--	--	L	29
N002801	1.921	(0.114)	-0.767	(0.081)	0.175	(0.028)	L	24	L	25	L	30
N002802	1.896	(0.110)	-0.912	(0.092)	0.143	(0.028)	L	25	L	26	L	31
N002902	0.558	(0.050)	-0.801	(0.114)	0.229	(0.071)	--	--	M	6	M	6
N002903	2.3[illegible]	(0.180)	-0.341	(0.082)	0.253	(0.040)	--	--	M	7	M	7
N002904	1.289	(0.095)	-0.020	(0.087)	0.197	(0.041)	--	--	M	8	M	8
N002905	0.758	(0.058)	0.248	(0.083)	0.116	(0.040)	--	--	M	9	M	9
N002906	1.964	(0.148)	-0.363	(0.082)	0.230	(0.044)	--	--	M	10	M	10
N003001	1.293	(0.109)	1.153	(0.169)	0.207	(0.013)	M	10	M	11	M	11
N003002	0.309	(0.029)	0.119	(0.065)	0.168	(0.041)	M	11	M	12	M	12
N003003	2.294	(0.109)	1.724	(0.190)	0.120	(0.006)	M	12	M	13	M	13
N003101	1.571	(0.100)	-0.645	(0.073)	0.267	(0.032)	M	14	M	14	M	14
N003102	1.530	(0.083)	-0.353	(0.051)	0.145	(0.023)	M	15	M	15	M	15
N003201	1.207	(0.088)	-0.593	(0.087)	0.171	(0.056)	--	--		12	N	21
N003202	1.590	(0.124)	0.012	(0.093)	0.227	(0.038)	--	--		13	N	22
N003203	1.215	(0.101)	0.240	(0.107)	0.222	(0.039)	--	--	N	14	N	23
N003204	1.457	(0.120)	0.260	(0.112)	0.238	(0.0[illegible])	--	--	N	15	N	24
N003301	1.141	(0.081)	-0.410	(0.078)	0.158	(0.049)	--	--	N	16	N	25
N003401	1.467	(0.150)	-0.207	(0.092)	0.159	(0.047)	--	--	N	17	--	--
N003501	0.751	(0.062)	-0.448	(0.093)	0.172	(0.061)	--	--	N	18	N	27
N003601	1.452	(0.116)	-0.668	(0.099)	0.203	(0.060)	--	--	N	19	N	28
N003602	1.320	(0.109)	-0.130	(0.097)	0.241	(0.048)	--	--	N	20	N	29
N003701	0.736	(0.061)	-0.760	(0.104)	0.239	(0.060)	N	23	N	21	N	30
N003702	1.071	(0.084)	-0.010	(0.078)	0.236	(0.032)	N	24	N	22	N	31
N003801	0.891	(0.112)	1.465	(0.251)	0.309	(0.018)	O	12	O	12	O	12
N003802	0.414	(0.030)	-0.703	(0.078)	0.110	(0.047)	O	13	O	13	O	13
N003803	0.757	(0.093)	1.600	(0.245)	0.206	(0.019)	O	14	O	14	O	14
N003901	1.375	(0.192)	-1.847	(0.331)	0.232	(0.089)	--	--	O	16	--	--
N004002	615	(0.079)	-1.426	(0.214)	0.246	(0.093)	--	--	O	15	--	--
N004101	1.096	(0.087)	-1.122	(0.114)	0.229	(0.054)	O	17	O	17	--	--
N004201	1.103	(0.071)	0.031	(0.062)	0.185	(0.024)	O	18	O	18	O	21
N004202	0.762	(0.072)	0.187	(0.098)	0.291	(0.038)	O	19	O	19	O	22
N004301	1.420	(0.125)	0.404	(0.131)	0.286	(0.032)	--	--	O	.0	O	23

Table F-2 (continued)

1988 IRT Parameters, Reading Bridge to 1984

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 11/AGE 17 BLOCK	GRADE 11/AGE 17 ITEM
N004401	1.718	(0.127)	-1.774	(0.202)	0.262	(0.065)	--	--	P	7	--	--
N004402	0.876	(0.075)	-0.220	(0.066)	0.148	(0.036)	--	--	P	8	--	--
N004403	1.642	(0.128)	-1.467	(0.170)	0.228	(0.054)	--	--	P	9	--	--
N004501	0.974	(0.103)	0.493	(0.151)	0.305	(0.043)	--	--	P	10	P	20
N004502	0.680	(0.054)	-0.824	(0.105)	0.180	(0.068)	--	--	P	11	P	21
N004601	0.899	(0.078)	0.179	(0.104)	0.184	(0.048)	--		P	12	P	22
N004602	1.318	(0.103)	-0.085	(0.092)	0.249	(0.044)	--	--	P	13	P	23
N004603	1.485	(0.113)	-0.516	(0.089)	0.226	(0.034)	--	--	P	14	P	24
N004701	1.694	(0.101)	-0.515	(0.059)	0.204	(0.021)	Q	10	Q	7	--	--
N004702	0.764	(0.065)	-0.928	(0.105)	0.237	(0.057)	Q	11	Q	8	--	--
N004703	1.021	(0.065)	-0.651	(0.062)	0.153	(0.031)	Q	12	Q	9	--	--
N004801	1.257	(0.085)	-1.258	(0.108)	0.242	(0.047)	Q	13	Q	10	--	--
N004901	0.916	(0.057)	0.221	(0.060)	0.190	(0.021)	Q	14	Q	11	Q	10
N005001	1.993	(0.102)	1.380	(0.159)	0.211	(0.011)	--	--	Q	13	Q	7
N005002	0.859	(0.108)	1.288	(0.240)	0.264	(0.029)	--	--	Q	14	Q	8
N005003	0.737	(0.105)	1.905	(0.331)	0.135	(0.024)	--	--	Q	15	Q	9
N005101	0.842	(0.061)	-2.140	(0.178)	0.236	(0.083)	Q	15	Q	12	--	--
N005201	0.674	(0.107)	0.636	(0.239)	0.481	(0.054)	--	--	Q	16	Q	11
N005202	0.600	(0.071)	0.582	(0.152)	0.206	(0.058)	--	--	Q	17	Q	12
N005203	1.143	(0.121)	1.637	(0.284)	0.309	(0.015)	--	--	Q	18	Q	13
N005301	1.133	(0.146)	-0.028	(0.132)	0.283	(0.059)	--	--	Q	19	--	--
N005302	1.406	(0.145)	0.387	(0.119)	0.129	(0.030)	--	--	Q	20	--	--
N005303	0.887	(0.195)	1.008	(0.344)	0.330	(0.048)	--	--	Q	21	--	--
N005304	1.810	(0.197)	0.052	(0.114)	0.227	(0.038)	--	--	Q	22	--	--
N005305	1.086	(0.121)	-0.677	(0.130)	0.222	(0.077)	--	--	Q	23	--	--
N005403	1.347	(0.153)	-0.335	(0.115)	0.289	(0.061)	--	--	R	7	--	--
N005404	1.455	(0.138)	-1.037	(0.144)	0.187	(0.067)	--	--	R	8	--	--
N005405	2.018	(0.195)	0.068	(0.100)	0.206	(0.031)	--	--	R	9	--	--
N005406	1.210	(0.116)	-0.398	(0.094)	0.165	(0.054)	--	--	R	10	--	--
N005407	1.777	(0.201)	-0.246	(0.110)	0.326	(0.049)	--	--	R	11	--	--
N005503	0.718	(0.074)	0.358	(0.127)	0.211	(0.054)	--	--	R	14	R	14
N005504	1.316	(0.112)	0.778	(0.147)	0.219	(0.024)	--	--	R	15	R	15
N005505	1.126	(0.092)	-0.913	(0.121)	0.247	(0.079)	--	--	R	16	R	16
N005601	1.387	(0.151)	-0.653	(0.125)	0.253	(0.071)	--	--	R	17	--	--
N005602	1.715	(0.187)	0.297	(0.133)	0.208	(0.031)	--	--	R	18	--	--
N005603	1.487	(0.171)	-0.177	(0.113)	0.306	(0.051)	--	--	R	19	--	--
N008601	1.789	(0.179)	-0.972	(0.171)	0.169	(0.037)	H	6	--	--	--	--
N008602	1.368	(0.179)	-0.554	(0.122)	0.261	(0.041)	H	7	--	--	--	--
N008603	1.206	(0.118)	-0.985	(0.137)	0.140	(0.043)	H	8	--	--	--	--
N008701	1.192	(0.134)	-2.391	(0.342)	0.240	(0.088)	H	9	--	--	--	--
N008801	1.489	(0.100)	-1.789	(0.173)	0.194	(0.056)	J	18	--	--	--	--
N008901	1.328	(0.106)	-1.244	(0.138)	0.148	(0.041)	J	21	--	--	--	--
N008902	1.258	(0.102)	-1.271	(0.140)	0.156	(0.043)	J	22	--	--	--	--
N009001	1.328	(0.152)	-0.433	(0.097)	0.154	(0.031)	K	12	--	--	--	--
N009002	1.177	(0.163)	-0.093	(0.087)	0.178	(0.030)	K	13	--	--	--	--
N009003	0.844	(0.203)	0.755	(0.242)	0.226	(0.032)	K	14	--	--	--	--
N009004	1.768	(0.225)	-0.350	(0.109)	0.240	(0.027)	K	15	--	--	--	--
N009101	1.007	(0.120)	-1.451	(0.210)	0.256	(0.076)	K	16	--	--	--	--
N009201	1.795	(0.172)	-1.377	(0.216)	0.301	(0.054)	K	17	--	--	--	--
N009401	1.882	(0.127)	-1.402	(0.172)	0.105	(0.036)	L	23	--	--	--	--
N009601	1.360	(0.106)	-1.872	(0.207)	0.133	(0.053)	L	21	--	--	--	--
N009701	1.082	(0.124)	-0.654	(0.112)	0.164	(0.047)	M	5	--	--	--	--
N009702	1.959	(0.227)	-0.533	(0.131)	0.249	(0.028)	M	6	--	--	--	--
N009703	1.449	(0.211)	-0.165	(0.097)	0.258	(0.029)	M	7	--	--	--	--
N009704	1.150	(0.185)	0.033	(0.086)	0.209	(0.031)	M	8	--	--	--	--
N009705	1.957	(0.207)	-0.702	(0.147)	0.211	(0.029)	M	9	--	--	--	--
N009801	1.396	(0.134)	-2.227	(0.296)	0.259	(0.086)	N	12	--	--	-	--
N009901	0.976	(0.117)	-1.049	(0.160)	0.206	(0.059)	N	13	--	--	--	--
N010002	1.290	(0.137)	-1.094	(0.165)	0.172	(0.047)	N	18	--	--	--	--
N010003	1.657	([illegible]194)	-0.940	(0.179)	0.241	(0.042)	N	19	--	--	--	--
N010102	1.124	(0.193)	-0.050	(0.111)	0.267	(0.0[illegible]7)	N	21	--	--	--	--
N010103	1.795	(0.200)	-1.075	(0.207)	0.209	(0.042)	N	22	--	--	--	--
N010201	[illegible]243	(0.121)	-1.932	(0.245)	0.244	(0.078)	O	16	--	--	--	--
N010301	[illegible].702	(0.085)	-2.383	(0.318)	0.248	(0.093)	O	15	--	--	--	--

Table F-2 (continued)

1988 IRT Parameters, Reading Bridge to 1984

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 11/AGE 17 BLOCK	GRADE 11/AGE 17 ITEM
N010401	0.715	(0.087)	-1.487	(0.209)	0.219	(0.077)	O	20	--	--	--	--
N010402	0.928	(0.171)	0.132	(0.113)	0.222	(0.037)	O	21	--	--	--	--
N010403	1.031	(0.137)	0.465	(0.153)	0.190	(0.027)	O	22	--	--	--	--
N010801	1.084	(0.119)	-0.471	(0.087)	0.260	(0.035)	Q	16	--	--	--	--
N010902	1.564	(0.153)	-0.467	(0.087)	0.241	(0.026)	Q	18	--	--	--	--
N010903	1.850	(0.157)	-0.564	(0.096)	0.193	(0.022)	Q	19	--	--	--	--
N010904	1.522	(0.170)	-0.245	(0.080)	0.275	(0.024)	Q	20	--	--	--	--
N011001	1.279	(0.116)	-0.079	(0.113)	0.228	(0.037)	R	5	--	--	--	--
N011002	1.657	(0.167)	-0.315	(0.081)	0.252	(0.022)	R	6	--	--	--	--
N011003	2.416	(0.169)	-0.928	(0.147)	0.[illegible]41	(0.024)	R	7	--	--	--	--
N011004	1.788	(0.159)	-0.543	(0.097)	0.226	(0.023)	R	8	--	--	--	--
N011101	1.568	(0.141)	-0.541	(0.090)	0.197	(0.025)	R	9	--	--	--	--
N011201	0.911	(0.117)	-0.259	(0.085)	0.260	(0.037)	R	10	--	--	--	-
N011301	1.653	(0.143)	-0.756	(0.111)	0.211	(0.028)	R	11	--	--	--	--
N011302	0.992	(0.119)	-0.430	(0.089)	0.227	(0.039)	R	12	--	--	--	--
N011401	0.838	(0.194)	0.697	(0.227)	0.334	(0.030)	R	13	--	--	--	--
N011402	0.822	(0.139)	0.010	(0.102)	0.288	(0.041)	R	14	--	--	--	--
N011403	0.971	(0.192)	0.621	(0.186)	0.270	(0.025)	R	15	--	--	--	--
N011404	1.327	(0.215)	0.492	(0.151)	0.220	(0.019)	R	16	--	--	--	--
N013201	1.655	(0.211)	-0.693	(0.160)	0.181	(0.037)	V	29	--	--	--	--
N013301	1.232	(0.161)	-1.557	(0.268)	0.253	(0.077)	V	30	--	--	--	--
N013401	1.203	(0.177)	-0.250	(0.107)	0.157	(0.035)	V	31	--	--	--	--
N013402	1.438	(0.189)	-0.862	(0.175)	0.205	(0.048)	V	32	--	--	--	--
N013403	1.494	(0.223)	-0.278	(0.116)	0.199	(0.033)	V	33	--	--	--	--
N0[illegible]01	1.238	(0.153)	-0.857	(0.149)	0.249	(0.048)	M	13	--	--	--	--
N014101	0.758	(0.071)	-1.283	(0.142)	0.169	(0.059)	Q	21	--	--	--	--
N014201	1.207	(0.134)	-1.218	(0.189)	0.136	(0.052)	V	34	--	--	--	--
N014301	1.755	(0.191)	-0.820	(0.158)	0.190	(0.035)	N	14	--	--	--	--
N014302	1.074	(0.136)	-0.498	(0.108)	0.181	(0.041)	N	15	--	--	--	--
N014303	1.721	(0.187)	-1.041	(0.188)	0.208	(0.041)	N	16	--	--	--	--
N014501	0.432	(0.065)	-2.264	(0.348)	0.000	(0.000)	V	35A	--	--	--	--
N014502	0.934	(0.123)	-2.664	(0.406)	0.000	(0.000)	V	35B	--	--	--	--
N014503	0.624	(0.133)	-4.120	(0.903)	0.000	(0.000)	V	35C	--	--	--	--
N015101	0.932	(0.110)	0.343	(0.168)	0.234	(0.067)	--	--	--	--	R	17
N015102	2.533	(0.236)	0.548	(0.207)	0.216	(0.030)	--	--	--	--	R	18
N015103	2.401	(0.200)	0.660	(0.197)	0.219	(0.028)	--	--	--	--	R	19
N015104	1.707	(0.184)	0.441	(0.193)	0.278	(0.045)	--	--	--	--	R	20
N015201	1.089	(0.126)	-0.766	(0.150)	0.227	(0.085)	--	--	--	--	N	26
N015502	1.273	(0.126)	0.189	(0.140)	0.209	(0.057)	--	--	--	--	P	16
N015503	0.912	(0.119)	0.756	(0.216)	0.247	(0.056)	--	--	--	--	P	17
N015504	1.189	(0.121)	0.110	(0.138)	0.220	(0.062)	--	--	--	--	P	18
N015505	0.683	(0.083)	-0.175	(0.145)	0.247	(0.087)	--	--	--	--	P	19
N015901	1.021	(0.133)	0.371	(0.204)	0.333	(0.068)	--	--	--	--	Q	14
N015902	1.380	(0.165)	0.726	(0.234)	0.317	(0.043)	--	--	--	--	Q	15
N015903	1.182	(0.129)	1.101	(0.224)	0.153	(0.032)	--	--	--	--	Q	16
N016001	1.043	(0.122)	0.033	(0.164)	0.285	(0.078)	--	--	--	--	O	15
N016002	1.386	(0.154)	1.247	(0.276)	0.456	(0.028)	--	--	--	--	O	16
N016003	0.906	(0.103)	0.354	(0.157)	0.205	(0.065)	--	--	--	--	O	17
N016004	1.095	(0.126)	0.103	(0.165)	0.271	(0.074)	--	--	--	--	O	18
N016005	1.734	(0.175)	0.156	(0.152)	0.230	(0.054)	--	--	--	--	O	19
N016006	1.357	(0.137)	0.424	(0.161)	0.203	(0.049)	--	--	--	--	O	20
N017001	1.518	(0.157)	0.484	(0.175)	0.213	(0.042)	--	--	--	--	H	7
N017002	1.935	(0.138)	1.100	(0.193)	0.196	(0.022)	--	--	--	--	H	8
N017003	1.833	(0.129)	1.770	(0.248)	0.177	(0.016)	--	--	--	--	H	9

Table F-3

1988 IRT Parameters, Civics Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	ITEM	GRADE 8/AGE 13 BLOCK	ITEM	GRADE 12/AGE 17 BLOCK	ITEM
P000101	0.663	(0.038)	0.123	(0.025)	0.468	(0.008)	C4	1	C4	1	--	--
P000201	0.587	(0.032)	-0.679	(0.044)	0.485	(0.010)	C4	2	C4	2	--	--
P000301	0.462	(0.025)	-1.126	(0.064)	0.443	(0.012)	C4	3	C4	3	--	--
P000401	0.282	(0.029)	0.377	(0.045)	0.510	(0.008)	C4	4	C4	4	--	--
P000501	1.324	(0.029)	-0.931	(0.029)	0.251	(0.012)	C4	5	C4	5	--	--
P000601	1.043	(0.044)	-0.597	(0.036)	0.509	(0.009)	C4	6	C4	6	--	--
P000701	1.691	(0.056)	0.256	(0.027)	0.313	(0.006)	C4	7	C4	7	--	--
P000801	1.709	(0.056)	-0.252	(0.024)	0.258	(0.009)	C4	8	C4	8	--	--
P000802	1.839	(0.062)	0.171	(0.026)	0.230	(0.007)	C4	9	C4	9	--	--
P000901	0.822	(0.039)	-0.269	(0.026)	0.504	(0.008)	C4	10	C4	10	--	--
P001001	1.079	(0.050)	0.129	(0.029)	0.519	(0.007)	C4	11	C4	11	--	--
P001101	0.836	(0.028)	-1.332	(0.051)	0.438	(0.014)	C4	12	C4	12	--	--
P001201	0.446	(0.029)	-0.514	(0.039)	0.496	(0.009)	C4	13	C4	13	--	--
P001301	1.191	(0.065)	0.130	(0.030)	0.471	(0.008)	C4	14	C4	14	--	--
P001401	0.690	(0.031)	-0.376	(0.027)	0.432	(0.009)	C4	15	C4	15	--	--
P001501	1.225	(0.064)	-0.003	(0.028)	0.500	(0.008)	C4	16	C4	16	--	--
P001601	[illegible]	(0.034)	-0.791	(0.038)	0.463	(0.011)	C4	17	C4	17	--	--
P001701	0.743	(0.029)	-0.939	(0.043)	0.421	(0.011)	C4	18	C4	18	--	--
P001801	1.074	(0.041)	-0.308	(0.025)	0.424	(0.009)	C4	19	C4	19	--	--
P001802	1.898	(0.080)	0.135	(0.034)	0.410	(0.007)	C4	20	C4	20	--	--
P001901	2.306	(0.059)	1.142	(0.067)	0.320	(0.007)	--	--	C4	21	--	--
P002001	2.067	(0.089)	-0.011	(0.030)	0.252	(0.012)	--	--	C4	22	--	--
P002101	2.981	(0.089)	1.196	(0.101)	0.434	(0.008)	--	--	C4	23	--	--
P002201	0.785	(0.048)	-0.915	(0.065)	0.444	(0.020)	--	--	C4	24	--	--
P002301	1.533	(0.055)	0.132	(0.026)	0.209	(0.012)	--	--	C4	25	--	--
P002401	1.557	(0.064)	1.680	(0.101)	0.189	(0.007)	--	--	C4	26	--	--
P002501	1.371	(0.049)	1.091	(0.058)	0.355	(0.008)	--	--	C4	27	--	--
P002601	1.568	(0.060)	0.059	(0.027)	0.242	(0.013)	--	--	C4	28	--	--
P002701	0.775	(0.039)	0.573	(0.038)	0.297	(0.010)	--	--	C4	29	--	--
P002702	0.555	(0.086)	2.894	(0.455)	0.405	(0.007)	--	--	C4	30	--	--
P003201	1.053	(0.020)	0.544	(0.019)	0.195	(0.007)	--	--	C2	1	C2	1
P003401	1.765	(0.021)	1.159	(0.028)	0.272	(0.005)	--	--	C2	2	C2	2
P003501	1.676	(0.013)	1.525	(0.031)	0.171	(0.004)	--	--	C2	3	C2	3
P003601	1.545	(0.019)	0.974	(0.024)	0.230	(0.006)	--	--	C2	4	C2	4
P003701	1.090	(0.023)	0.287	(0.017)	0.265	(0.006)	--	--	C2	5	C2	5
P003801	1.815	(0.023)	1.233	(0.031)	0.271	(0.005)	--	--	C2	6	C2	6
P003901	1.523	(0.019)	0.942	(0.022)	0.211	(0.006)	--	--	C2	17	C2	17
P004001	1.795	(0.067)	-0.068	(0.028)	0.469	(0.010)	--	--	C2	7	C2	7
P004002	0.447	(0.020)	0.191	(0.020)	0.474	(0.008)	--	--	C2	8	C2	8
P004003	1.649	(0.055)	-0.571	(0.036)	0.436	(0.014)	--	--	C2	9	C2	9
P004004	0.547	(0.022)	1.263	(0.055)	0.447	(0.007)	--	--	C2	10	C2	10
P004005	1.215	(0.026)	1.491	(0.048)	0.523	(0.005)	--	--	C2	11	C2	11
P004007	2.269	(0.034)	0.899	(0.037)	0.494	(0.006)	--	-	C2	12	C2	12
P004008	1.336	(0.047)	-0.367	(0.028)	0.428	(0.012)	--	--	C2	13	C2	13
P004010	1.805	(0.061)	-0.582	(0.037)	0.442	(0.014)	--	--	C2	14	C2	14
P004011	0.506	(0.020)	1.061	(0.047)	0.451	(0.007)	--	--	C2	15	C2	15
P004012	0.697	(0.038)	2.638	(0.153)	0.509	(0.005)	--	--	C2	16	C2	16
P004101	1.270	(0.033)	-0.231	(0.019)	0.264	(0.010)	--	-	C2	18	C2	18
P004201	1.631	(0.043)	-0.266	(0.021)	0.190	(0.011)	--	--	C2	19	C2	19
P004301	1.033	(0.019)	0.733	(0.021)	0.214	(0.007)	--	--	C2	26	C2	26
P004401	0.703	(0.028)	0.037	(0.022)	0.545	(0.009)	--	--	C2	20	C2	20
P004501	2.455	(0.033)	1.019	(0.038)	0.470	(0.006)	--	--	C2	21	C2	21
P004601	0.782	(0.029)	-0.145	(0.021)	0.497	(0.009)	--	--	C2	22	C2	22
P004701	0.788	(0.027)	0.520	(0.030)	0.567	(0.007)	--	--	C2	23	C2	23
P004801	1.069	(0.032)	-0.001	(0.021)	0.443	(0.010)	--	--	C2	24	C2	24
P004901	1.306	(0.038)	-0.193	(0.024)	0.430	(0.011)	--	--	C2	25	C2	25
P005301	1.248	(0.037)	0.328	(0.029)	0.500	(0.009)	--	--	C5	1	C5	1
P005401	1.362	(0.061)	[illegible]	(0.035)	0.479	(0.014)	--	--	C5	2	C5	2
P005501	0.955	(0.039)	-0.644	(0.035)	0.466	(0.013)	--	--	C5	5	C5	5
P005502	0.898	(0.039)	-1.497	(0.073)	0.451	(0.022)	--	--	C5	6	C5	6
P005504	1.549	(0.064)	-1.127	(0.063)	0.468	(0.023)	--	--	C5	7	C5	7
P005505	0.878	(0.043)	-0.853	(0.050)	0.457	(0.017)	--	--	C5	8	C5	8
P005601	1.852	(0.060)	-0.591	(0.034)	0.474	(0.016)	--	--	C5	3	C5	3
P005701	1.032	(0.027)	-0.209	(0.018)	0.220	(0.011)	--	--	C5	4	C5	4

Table F-3 (continued)

1938 IRT Parameters, Civics Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE .2/AGE 17 BLOCK	GRADE .2/AGE 17 ITEM
P005801	0.690	(0.028)	0.002	(0.022)	0.539	(0.009)	--	--	C5	10	C5	10
P005802	1.178	(0.029)	0.944	(0.038)	0.516	(0.007)	--	--	C5	11	C5	11
P005803	1.435	(0.033)	0.562	(0.030)	0.495	(0.008)	--	--	C5	12	C5	12
P005901	0.936	(0.029)	-0.198	(0.022)	0.426	(0.011)	--	--	C5	14	C5	14
P006101	1.579	(0.040)	0.431	(0.032)	0.550	(0.008)	--	--	C5	16	C5	16
P006201	1.491	(0.039)	0.218	(0.027)	0.493	(0.009)	--	--	C5	17	C5	17
P006301	1.659	(0.029)	0.529	(0.024)	0.311	(0.008)	--	--	C5	18	C5	18
P006401	1.085	(0.027)	1.045	(0.040)	0.490	(0.007)	--	--	C5	19	C5	19
P006501	1.200	(0.036)	0.193	(0.026)	0.470	(0.010)	--	--	C5	20	C5	20
P006601	0.849	(0.018)	0.929	(0.026)	0.226	(0.007)	--	--	C5	21	C5	21
P006701	2.024	(0.031)	0.669	(0.027)	0.346	(0.007)	--	--	C5	22	C5	22
P006801	1.613	(0.037)	0.326	(0.027)	0.417	(0.009)	--	--	C5	23	C5	23
P006802	0.387	(0.019)	0.359	(0.026)	0.441	(0.009)	--	--	C5	24	C5	24
P006803	1.343	(0.031)	1.982	(0.069)	0.396	(0.006)	--	--	C5	25	C5	25
P006804	1.347	(0.026)	1.248	(0.040)	0.446	(0.006)	--	--	C5	26	C5	26
P007001	2.819	(0.029)	1.680	(0.050)	0.249	(0.005)	--	--	C5	27	C5	27
P008201	1.107	(0.047)	-0.168	(0.037)	0.496	(0.019)	--	--	--	--	C3	1
P008202	1.684	(0.083)	-0.090	(0.045)	0.469	(0.020)	--	--	--	--	C3	2
P006301	1.469	(0.032)	0.417	(0.030)	0.255	(0.014)	--	--	--	--	C3	3
P008401	1.597	(0.034)	2.311	(0.081)	0.236	(0.006)	--	--	--	--	C3	5
P008402	1.863	(0.030)	0.807	(0.034)	0.342	(0.011)	--	--	--	--	C3	6
P008403	1.900	(0.032)	0.771	(0.035)	0.358	(0.011)	--	--	--	--	C3	7
P008404	2.168	(0.028)	1.608	(0.048)	0.343	(0.008)	--	--	--	--	C3	8
P008405	1.898	(0.036)	0.739	(0.037)	0.405	(0.011)	--	--	--	--	C3	9
P008406	1.773	(0.025)	1 372	(0.040)	0.346	(0.009)	--	--	--	--	C3	10
P008407	1.789	(0.026)	1.195	(0.038)	0.286	(0.010)	--	--	--	--	C3	11
P008501	0.901	(0.020)	1.748	0.049)	0.241	(0.008)	--	--	--	--	C3	12
P008601	3.427	(0.031)	1.999	(0.060)	0.100	(0.005)	--	--	--	--	C3	14
P008701	1.343	(0.020)	1.312	(0.033)	0.205	(0.009)	--	--	--	--	C3	4
P008801	1.149	(0.019)	1.322	(0.032)	0.171	(0.009)	--	--	--	--	C3	13
P008901	0.747	(0.020)	1.547	(0.048)	0.237	(0.010)	--	--	--	--	C3	15
P009001	2.364	(0.025)	1.507	(0.041)	0.255	(0.008)	--	--	--	--	C3	16
P009101	1.794	(0.022)	1.214	(0.032)	0.177	(0.009)	--	--	--	--	C3	17
P009201	2.104	(0.025)	1.712	(0.045)	0.239	(0.007)	--	--	--	--	C3	18
P009301	1.470	(0.021)	1.693	(0.041)	0.162	(0.008)	--	--	--	--	C3	19
P009401	0.957	(0.020)	0.807	(0 028)	0.212	(0.011)	--	--	--	--	C3	20
P009501	1.792	(0.023)	1.719	(0.043)	0.210	(0.007)	--	--	--	--	C3	21
P009601	1.059	(0.049)	-0.684	(0.046)	0.247	(0.025)	--	--	--	--	C3	22
P009801	1.356	(0.028)	0.401	(0.029)	0.231	(0.013)	--	--	--	--	C3	24
P009901	1.525	(0.023)	1.298	(0.037)	0.342	(0.009)	--	--	--	--	C3	25
P010001	1.615	(0.022)	1.325	(0.035)	0.248	(0.009)	--	--	--	--	C3	26
P010101	1.231	(0.022)	1.342	(0.038)	0.310	(0.009)	--	--	--	--	C3	23
P010901	2.147	(0.032)	-2.091	(0.063)	0.219	(0.026)	C2	1	--	--	--	--
P011001	1.501	(0.025)	-1.559	(0.040)	0.283	(0.017)	C2	2	--	--	--	--
P011101	0.842	(0.020)	-2.265	(0.061)	0.233	(0.023)	C2	3	--	--	--	--
P011201	1.090	(0.023)	-1.455	(0.038)	0.240	(0.015)	C2	4	--	--	--	--
P011301	1.226	(0.023)	-1.753	(0.043)	0.227	(0.019)	C2	5	--	--	--	--
P011401	1.761	(0.044)	-0.822	(0.038)	0.407	(0.011)	C2	6	--	--	--	--
P011402	1.006	(0.043)	-0.744	(0.045)	0.488	(0.011)	C2	7	--	--	--	--
P011403	1.042	(0.032)	-1.157	(0.046)	0.435	(0.014)	C2	8	--	--	--	--
P011404	1.243	(0.036)	-1.022	(0.043)	0.409	(0.014)	C2	9	--	--	--	--
P011405	1.216	(0.048)	-0.163	(0.023)	0.426	(0.007)	C3	3	C3	3	--	--
P011406	1.155	(0.027)	-1.218	(0.036)	0.386	(0.012)	C3	4	C3	4	--	--
P011407	1.098	(0.030)	-1.021	(0.036)	0.453	(0.011)	C3	5	C3	5	--	--
P011408	1.110	(0.032)	-0.894	(0.035)	0.484	(0.010)	C3	6	3	6	--	--
P011501	1.577	(0.034)	-0.747	(0.030)	0.237	(0.010)	C2	10	--	--	--	--
P011601	1.388	(0.035)	-0.681	(0.030)	0.268	(0.010)	C2	11	--	--	--	--
P011701	1.337	(0.046)	-0.237	(0.027)	0.181	(0.008)	C2	12	--	--	--	--
P011801	1.960	(0.056)	-0.379	(0.031)	0.202	(0.008)	C2	13	--	--	--	--
P011901	1.353	(0.025)	-1.121	(0.031)	0.187	(0.013)	C2	14	--	--	--	--
P012001	0.878	(0.043)	-0.075	(0.028)	0.289	(0.008)	C2	15	--	--	--	--
P012101	1.374	(0.022)	-1.53[illegible]	(0.034)	0.246	(0.015)	C3	1	C3	1	--	--
P012201	1.352	(0.020)	-1.7[illegible]4	(0.036)	0.186	(0.01[illegible])	C3	2	C3	2	--	--
P012501	1.573	(0.027)	-1.013	(0.027)	0.273	(0.011)	C3	7	C3	7	--	--

Table F-3 (continued)

1988 IRT Parameters, Civics Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
P012502	1.149	(0.027)	-0.674	(0.024)	0.260	(0.009)	C3	8	C3	8	--	--
P012601	1.559	(0.045)	-0.220	(0.021)	0.262	(0.007)	C3	9	C3	9	--	--
P012701	1.506	(0.034)	-0.584	(0.023)	0.268	(0.008)	C3	10	C3	10	--	--
P012801	1.035	(0.023)	-0.549	(0.019)	0.149	(0.008)	C3	11	C3	11	--	--
P012901	1.262	(0.028)	-0.500	(0.020)	0.172	(0.008)	C3	12	C3	12	--	--
P013001	0.591	(0.020)	-0.590	(0.025)	0.238	(0.009)	C3	13	C3	13	--	--
P013101	1.148	(0.021)	-1.012	(0.025)	0.151	(0.011)	C3	14	C3	14	--	--
P013201	1.244	(0.041)	0.961	(0.042)	0.176	(0.004)	C3	15	C3	15	--	--
P013301	0.936	(0.035)	0.440	(0.025)	0.237	(0.006)	C3	16	C3	16	--	--
P013401	1.359	(0.052)	0.037	(0.023)	0.216	(0.011)	--	--	C3	17	--	--
P013501	1.102	(0.047)	-0.317	(0.026)	0.203	(0.013)	--	--	C3	19	--	--
P013601	0.861	(0.039)	0.263	(0.025)	0.238	(0.010)	--	--	C3	18	--	--
P013701	1.289	(0.046)	0.533	(0.032)	0.248	(0.008)	--	--	C3	20	--	--
P013801	2.061	(0.056)	0.882	(0.050)	0.325	(0.007)	--	--	C3	21	--	--
P013901	1.035	(0.042)	0.169	(0.023)	0.238	(0.010)	--	--	C3	22	--	--
P013902	1.236	(0.043)	1.119	(0.052)	0.230	(0.007)	--	--	C3	23	--	--
P014001	2.049	(0.097)	0.016	(0.029)	0.289	(0.011)	--	--	C3	24	--	--
P014101	1.152	(0.034)	-0.475	(0.023)	0.328	(0.011)	--	--	C6	1	C6	1
P014201	1.185	(0.027)	-0.234	(0.016)	0.212	(0.010)	--	--	C6	2	C6	2
P014301	0.862	(0.018)	0.303	(0.015)	0.240	(0.007)	--	--	C6	3	C6	3
P014401	1.503	(0.022)	0.372	(0.016)	0.199	(0.006)	--	--	C6	4	C6	4
P014501	1.114	(0.023)	0.024	(0.015)	0.242	(0.008)	--	--	C6	5	C6	5
P014601	1.681	(0.028)	0.245	(0.016)	0.178	(0.007)	--	--	C6	6	C6	6
P014701	0.829	(0.016)	0.393	(0.015)	0.190	(0.007)	--	--	C6	7	C6	7
P014801	1.426	(0.025)	0.349	(0.019)	0.349	(0.007)	--	--	C6	8	C6	8
P014901	0.999	(0.020)	0.174	(0.014)	0.240	(0.008)	--	--	C6	9	C6	9
P015001	0.602	(0.015)	0.069	(0.012)	0.200	(0.008)	--	--	C6	10	C6	10
P015101	1.511	(0.021)	1.752	(0.040)	0.307	(0.004)	--	--	C6	11	C6	11
P015201	1.365	(0.019)	0.728	(0.019)	0.272	(0.006)	--	--	C6	12	C6	12
P015202	1.562	(0.039)	-0.238	(0.018)	0.205	(0.010)	--	--	C6	13	C6	13
P015301	1.057	(0.014)	0.955	(0.019)	0.174	(0.005)	--	--	C6	14	C6	14
P015401	1.324	(0.015)	1.240	(0.022)	0.159	(0.004)	--	--	C6	22	C6	22
P015501	1.232	(0.016)	1.238	(0.024)	0.235	(0.005)	--	--	C6	16	C6	16
P015601	2.674	(0.021)	1.037	(0.024)	0.227	(0.004)	--	--	C6	17	C6	17
P015701	0.563	(0.014)	1.053	(0.029)	0.226	(0.006)	--	--	C6	18	C6	18
P015702	0.518	(0.014)	1.500	(0.044)	0.214	(0.005)	--	--	C6	19	C6	19
P015703	1.592	(0.018)	1.452	(0.029)	0.262	(0.004)	--	--	C6	20	C6	20
P015704	1.241	(0.020)	2.044	(0.045)	0.195	(0.004)	--	--	C6	21	C6	21
P015801	1.164	(0.015)	0.959	(0.019)	0.163	(0.005)	--	--	C6	15	C6	15
P015901	1.641	(0.021)	0.617	(0.018)	0.223	(0.006)	--	--	C6	23	C6	23
P015902	2.711	(0.024)	0.779	(0.022)	0.181	(0.005)	--	--	C6	24	C6	24
P016001	1.768	(0.033)	0.021	(0.018)	0.302	(0.008)	--	--	C7	1	C7	1
P016101	1.349	(0.032)	-0.163	(0.017)	0.216	(0.009)	--	--	C7	2	C7	2
P016201	1.682	(0.020)	0.846	(0.022)	0.226	(0.005)	--	--	C7	3	C7	3
P016301	1.445	(0.027)	0.472	(0.021)	0.356	(0.007)	--	--	C7	4	C7	4
P016401	0.978	(0.021)	0.204	(0.015)	0.232	(0.008)	--	--	C7	5	C7	5
P016501	1.881	(0.031)	0.350	(0.021)	0.272	(0.007)	--	--	C7	6	C7	6
P016601	0.779	(0.020)	0.053	(0.015)	0.279	(0.008)	--	--	C7	[illegible]	C7	7
P016701	1.992	(0.024)	0.785	(0.023)	0.250	(0.005)	--	--	C7	8	C7	8
P016801	0.885	(0.018)	0.585	(0.019)	0.244	(0.007)	--	--	C7	9	C7	9
P016901	1.729	(0.024)	1.151	(0.031)	0.400	(0.005)	--	--	C7	10	C7	10
P017001	0.632	(0.015)	1.013	(0.027)	0.174	(0.006)	--	--	C7	11	C7	11
P017101	1.998	(0.022)	0.831	(0.022)	0.206	(0.005)	--	--	C7	12	C7	12
P017201	1.170	(0.017)	0.971	(0.022)	0.210	(0.005)	--	--	C7	13	C7	13
P017301	1.809	(0.022)	0.854	(0.023)	0.272	(0.005)	--	--	C7	14	C7	14
P017401	1.188	(0.022)	0.233	(0.015)	0.172	(0.007)	--	--	C7	15	C7	15
P017501	1.616	(0.020)	1.187	(0.027)	0.244	(0.005)	--	--	C7	16	C7	16
P017601	1.238	(0.018)	1.064	(0.025)	0.281	(0.005)	--	--	C7	17	C7	17
P017701	1.649	(0.020)	1.580	(0.035)	0.247	(0.004)	--	--	C7	18	C7	18
P017801	1.690	(0.021)	1.441	(0.035)	0.321	(0.005)	--	--	C7	19	C7	19
P017901	0.701	(0.016)	0.842	(0.025)	0.248	(0.006)	--	--	C7	20	C7	20
P018001	1.508	(0.017)	1.222	(0.025)	0.183	(0.004)	--	--	C7	21	C7	21
P018101	1.443	(0.019)	1.418	(0.031)	0.266	(0.005)	--	--	C7	22	C7	22
P018201	0.406	(0.021)	-3.026	(0.158)	0.000	(0.000)	--	--	C8	1	C8	1

Table F-3 (continued)

1988 IRT Parameters, Civics Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
P018202	0.861	(0.014)	0.685	(0.017)	0.000	(0.000)	--	--	C8	1	C8	1
P018301	0.996	(0.035)	-0.724	(0.037)	0.218	(0.021)	--	--	--	--	C4	2
P018401	1.816	(0.026)	0.928	(0.032)	0.419	(0.008)	--	--	--	--	C4	3
P018501	1.303	(0.047)	-0.395	(0.032)	0.249	(0.018)	--	--	--	--	C4	4
P018601	0.950	(0.020)	0.450	(0.022)	0.228	(0.011)	--	--	--	--	C4	5
P018701	1.502	(0.023)	0.761	(0.027)	0.305	(0.009)	--	--	--	--	C4	6
P018801	1.067	(0.022)	0.955	(0.032)	0.419	(0.008)	--	--	--	--	C4	7
P018901	0.511	(0.016)	0.353	(0.020)	0.216	(0.011)	--	--	--	--	C4	8
P019001	1.322	(0.018)	0.748	(0.022)	0.153	(0.009)	--	--	--	--	C4	9
P019101	1.935	(0.023)	0.864	(0.025)	0.180	(0.008)	--	--	--	--	C4	10
P019201	1.061	(0.018)	1.336	(0.032)	0.257	(0.007)	--	--	--	--	C4	11
P019301	0.799	(0.019)	0.245	(0.019)	0.197	(0.012)	--	--	--	--	C4	12
P019401	1.853	(0.022)	1.013	(0.028)	0.284	(0.008)	--	--	--	--	C4	13
P019501	1.463	(0.021)	0.921	(0.027)	0.274	(0.008)	--	--	--	--	C4	14
P019601	2.148	(0.021)	1.100	(0.028)	0.210	(0.007)	--	--	--	--	C4	15
P019701	2.713	(0.027)	1.096	(0.033)	0.297	(0.007)	--	--	--	--	C4	16
P019801	0.772	(0.020)	2.422	(0.070)	0.157	(0.005)	--	--	--	--	C4	17
P019901	1.238	(0.019)	0.887	(0.025)	0.209	(0.008)	--	--	--	--	C4	18
P020001	1.470	(0.018)	1.271	(0.028)	0.188	(0.007)	--	--	--	--	C4	19
P020101	1.614	(0.024)	1.827	(0.047)	0.308	(0.006)	--	--	--	--	C4	20
P020201	0.574	(0.017)	2.146	(0.069)	0.190	(0.007)	--	--	--	--	C4	21
P020301	1.958	(0.020)	1.115	(0.027)	0.184	(0.007)	--	--	--	--	C4	22
P020401	2.228	(0.025)	1.838	(0.047)	0.229	(0.005)	--	--	--	--	C4	23
P020501	1.922	(0.021)	1.613	(0.035)	0.195	(0.006)	--	--	--	--	C4	24
P020601	0.677	(0.029)	2.947	(0.135)	0.251	(0.006)	--	--	--	--	C4	25
P020701	1.691	(0.027)	0.704	(0.026)	0.324	(0.007)	--	--	C5	9	C5	9
P021301	0.840	(0.031)	-0.676	(0.035)	0.239	(0.019)	--	--	--	--	C4	1

Table F-4

1988 IRT Parameters, Civics Trend, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 13 BLOCK	AGE 13 ITEM
P000101	0.531	(0.057)	0.212	(0.044)	0.455	(0.014)	C9	1
P000201	0.421	(0.048)	-1.449	(0.171)	0.463	(0.021)	C9	2
P000401	0.188	(0.037)	-0.185	(0.049)	0.472	(0.015)	C9	4
P000501	0.715	(0.053)	-2.550	(0.194)	0.296	(0.040)	C9	5
P000601	0.393	(0.057)	-2.956	(0.43[illegible])	0.454	(0.035)	C9	6
P000801	0.886	(0.054)	-1.638	(0.112)	0.257	(0.034)	C9	8
P000802	0.826	(0.070)	-0.506	(0.057)	0.236	(0.021)	C9	9
P000901	0.413	(0.050)	-1.529	(0.191)	0.456	(0.024)	C9	10
P001001	0.621	(0.071)	-0.805	(0.104)	0.492	(0.022)	C9	11
P001101	0.394	(0.054)	-3.152	(0.435)	0.446	(0.036)	C9	12
P001201	0.232	(0.037)	-2.410	(0.382)	0.466	(0.024)	C9	13
P001301	0.700	(0.061)	0.331	(0.049)	0.406	(0.013)	C9	14
P001401	0.530	(0.051)	-1.558	(0.153)	0.440	(0.023)	C9	15
P001501	0.718	(0.086)	-0.195	(0.054)	0.429	(0.019)	C9	16
P001601	0.415	(0.052)	-2.313	(0.292)	0.442	(0.031)	C9	17
P001701	0.647	(0.079)	-0.769	(0.105)	0.475	(0.022)	C9	18
P001801	0.415	(0.048)	-1.627	(0.192)	0.418	(0.026)	C9	19
P001802	0.931	(0.105)	-0.343	(0.066)	0.418	(0.021)	C9	20
P001901	0.519	(0.057)	2.265	(0.258)	0.290	(0.011)	C9	21
P002101	1.066	(0.077)	1.622	(0.154)	0.418	(0.012)	C9	23
P002201	0.383	(0.050)	-2.438	(0.321)	0.436	(0.031)	C9	24
P002301	0.831	(0.056)	-0.527	(0.047)	0.236	(0.018)	C9	25
P002401	0.516	(0.056)	3.028	(0.335)	0.148	(0.008)	C9	26
P002501	0.764	(0.050)	2.057	(0.147)	0.198	(0.008)	C9	27
P002601	0.515	(0.050)	-0.517	(0.060)	0.239	(0.021)	C9	28
P002701	0.530	(0.048)	0.037	(0.029)	0.291	(0.015)	C9	29
P002702	0.308	(0.045)	2.316	(0.336)	0.296	(0.010)	C9	30
P002901	0.536	(0.063)	-0.814	(0.106)	0.447	(0.023)	C9	32
P003001	0.496	(0.051)	0.266	(0.044)	0.250	(0.017)	C9	33
P003201	0.598	(0.052)	0.727	(0.072)	0.226	(0.013)	C1	1
P003401	0.880	(0.063)	2.057	(0.171)	0.234	(0.010)	C1	2
P003501	0.523	(0.066)	3.496	(0.450)	0.172	(0.008)	C1	3
P003601	0.819	(0.054)	1.907	(0.143)	0.173	(0.010)	C1	4
P003701	0.628	(0.051)	-0.004	(0.033)	0.206	(0.017)	C1	5
P003801	0.629	(0.068)	2.658	(0.300)	0.221	(0.010)	C1	6
P004001	0.767	(0.082)	-0.819	(0.100)	0.438	(0.025)	C1	7
P004002	0.349	(0.040)	-0.707	(0.089)	0.403	(0.020)	C1	8
P004003	1.917	(0.086)	-1.322	(0.106)	0.333	(0.035)	C1	9
P004004	0.329	(0.064)	2.782	(0.548)	0.450	(0.013)	C1	10
P004005	0.573	(0.064)	1.007	(0.126)	0.436	(0.014)	C1	11
P004007	0.822	(0.087	1.966	(0.238)	0.463	(0.012)	C1	12
P004008	0.927	(0.073)	-1.153	(0.104)	0.394	(0.028)	C1	13
P004010	1.894	(0.084)	-1.375	(0.105)	0.343	(0.035)	C1	14
P004011	0.464	(0.055)	0.764	(0.102)	0.445	(0.015)	C1	15
P004012	0.568	(0.085)	2.629	(0.409)	0.434	(0.013)	C1	16
P004101	0.300	(0.060)	-0.958	(0.081)	0.218	(0.026)	C1	18
P004201	0.904	(0.066)	-0.774	(0.066)	0.207	(0.021)	C1	19
P004401	0.292	(0.048)	-1.817	(0.304)	0.460	(0.026)	C1	20
P004501	0.860	(0.085)	1.698	(0.198)	0.499	(0.013)	C1	21
P004601	0.370	(0.055)	-1.003	(0.156)	0.468	(0.023)	C1	22
P004701	0.372	(0.050)	-1.013	(0.142)	0.440	(0.023)	C1	23
P004901	0.605	(0.054)	-0.554	(0.064)	0.369	(0.022)	C1	25
P005001	0.921	(0.079)	2.293	(0.232)	0.293	(0.010)	C1	27
P005301	0.587	(0.055)	0.321	(0.048)	0.417	(0.013)	C1	1
P005401	0.789	(0.064)	-1.649	(0.142)	0.447	(0.029)	C1	2
P005501	0.508	(0.049)	-1.922	(0.192)	0.458	(0.028)	C1	5
P005502	0.859	(0.078)	-2.671	(0.258)	0.427	(0.049)	C1	6
P005504	1.082	(0.081)	-2.432	(0.214)	0.457	(0.051)	C1	7
P005505	0.505	(0.055)	-1.923	(0.215)	0.436	(0.032)	C1	8
P005601	0.844	(0.057)	-1.505	(0.112)	0.275	(0.032)	C1	3
P005801	0.533	(0.051)	-1.771	(0.175)	0.431	(0.025)	C1	10
P005802	0.857	(0.084)	1.932	(0.216)	0.462	(0.010)	C1	11
P005803	0.649	(0.073)	0.378	(0.061)	0.462	(0.014)	C1	12
P006001	0.487	(0.088)	3.509	(0.647)	0.385	(0.011)	C1	15

Table F-4 (continued)

1988 IRT Parameters, Civics Trend, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 13 BLOCK	AGE 13 ITEM
P006101	0.934	(0.104)	0.409	(0.079)	0.534	(0.015)	C1	16
P006201	0.980	(0.118)	-0.034	(0.058)	0.526	(0.017)	C1	17
P006301	0.763	(0.063)	0.933	(0.093)	0.317	(0.013)	C1	18
P006401	0.590	(0.061)	1.005	(0.114)	0.427	(0.013)	C1	19
P006501	0.752	(0.096)	0.273	(0.059)	0.490	(0.014)	C1	20
P006601	0.417	(0.042)	0.935	(0.099)	0.247	(0.013)	C1	21
P006701	0.696	(0.071)	0.189	(0.044)	0.259	(0.017)	C1	22
P006801	0.748	(0.079)	0.036	(0.049)	0.390	(0.019)	C1	23
P006802	0.237	(0.039)	-0.567	(0.101)	0.466	(0.019)	C1	24
P006803	0.565	(0.070)	3.069	(0.394)	0.288	(0.008)	C1	25
P006804	0.744	(0.062)	1.522	(0.142)	0.423	(0.010)	C1	26
P007201	0.791	(0.058)	0.181	(0.039)	0.203	(0.016)	C1	29
P007301	0.434	(0.060)	-0.403	(0.068)	0.485	(0.017)	C1	30
P007401	0.429	(0.049)	-1.774	(0.205)	0.442	(0.023)	C1	31
P007501	0.514	(0.056)	-2.363	(0.264)	0.445	(0.035)	C1	32
P007502	0.776	(0.095)	0.074	(0.052)	0.487	(0.016)	C1	33
P007503	0.559	(0.076)	-0.242	(0.058)	0.492	(0.019)	C1	34
P020701	0.530	(0.056)	0.223	(0.043)	0.288	(0.017)	C1	9
P021101	0.509	(0.037)	-1.882	(0.142)	0.000	(0.000)	C1	28

Table F-5

1988 IRT Parameters, Civics Trend, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 17 BL	AGE 17 ITEM
P000101	0.683	(0.066)	-0.408	(0.055)	0.502	(0.01[illegible])	C9	27
P000201	0.526	(0.053)	-1.490	(0.154)	0.479	(0.021)	C9	28
P000401	0.174	(0.033)	-0.160	(0.045)	0.525	(0.014)	C9	30
P000801	1.335	(0.050)	-1.886	(0.098)	0.217	(0.040)	C9	32
P000802	1.111	(0.050)	-1.510	(0.085)	0.203	(0.032)	C9	33
P003201	0.622	(0.047)	-0.414	(0.042)	0.226	(0.017)	C1	1
P003401	1.014	(0.087)	0.239	(0.049)	0.266	(0.015)	C1	2
P003501	1.262	(0.059)	.013	(0.068)	0.206	(0.010)	C1	3
P003601	1.009	(0.068)	0.421	(0.051)	0.194	(0.014)	C1	4
P003701	0.659	(0.050)	-1.089	(0.090)	0.240	(0.025)	C1	5
P003801	1.174	(0.079)	0.739	(0.073)	0.252	(0.014)	C1	6
P004001	1.245	(0.072)	-1.578	(0.118)	0.451	(0.036)	C1	7
P004002	0.421	(0.047)	-1.761	(0.200)	0.472	(0.026)	C1	8
P004003	1.327	(0.063)	-1.957	(0.128)	0.427	(0.042)	C1	9
P004004	0.475	(0.056)	-0.183	(0.050)	0.462	(0.018)	C1	10
P004005	0.933	(0.079)	0.253	(0.060)	0.494	(0.015)	C1	11
P004007	1.188	(0.105)	0.403	(0.074)	0.463	(0.014)	C1	12
P004008	0.988	(0.062)	-1.768	(0.129)	0.459	(0.035)	C1	13
P004010	1.394	(0.067)	-1.883	(0.124)	0.423	(0.041)	C1	14
P004011	0.571	(0.057)	0.403	(0.062)	0.449	(0.015)	C1	15
P004012	0.504	(0.057)	0.674	(0.091)	0.478	(0.015)	C1	16
P004201	1.106	(0.050)	-1.548	(0.084)	0.277	(0.030)	C1	19
P004401	0.505	(0.069)	-1.418	(0.199)	0.502	(0.027)	C1	20
P004501	1.045	(0.122)	-0.056	(0.060)	0.494	(0.018)	C1	21
P004601	0.778	(0.084)	-1.104	(0.132)	0.511	(0.026)	C1	22
P004701	0.577	(0.063)	-1.164	(0.135)	0.462	(0.025)	C1	23
P004901	0.597	(0.062)	-1.539	(0.166)	0.452	(0.028)	C1	25
P005001	1.264	(0.074)	0.776	(0.073)	0.277	(0.012)	C1	32
P005101	0.904	(0.083)	-0.542	(0.069)	0.405	(0.022)	C1	27
P005102	1.342	(0.091)	-0.509	(0.061)	0.345	(0.019)	C1	28
P005103	1.276	(0.101)	0.279	(0.058)	0.371	(0.014)	C1	29
P005104	1.788	(0.123)	1.442	(0.183)	0.559	(0.010)	C1	30
P005105	1.421	(0.120)	-0.233	(0.058)	0.415	(0.017)	C1	31
P005301	1.269	(0.089)	-0.923	(0.085)	0.505	(0.022)	C1	1
P005401	1.086	(0.065)	-1.986	(0.137)	0.479	(0.038)	C1	2
P005501	1.407	(0.072)	-1.663	(0.116)	0.437	(0.039)	C1	5
P005502	0.699	(0.070)	-3.130	(0.326)	0.462	(0.051)	C1	6
P005504	0.936	(0.123)	-3.212	(0.458)	0.468	(0.062)	C1	7
P005505	0.851	(0.076)	-2.015	(0.195)	0.460	(0.042)	C1	8
P005601	1.573	(0.064)	-1.697	(0.100)	0.232	(0 041)	C1	3
P005801	0.577	(0.053)	-1.573	(0.151)	0.493	(0.023)	C1	10
P005802	0.724	(0.075)	0.235	(0.051)	0.466	(0.014)	C1	11
P005803	0.936	(0.077)	-0.701	(0.072)	0.461	(0.020)	C1	[illegible]
P005804	0.620	(0.092)	3.051	(0.473)	0.408	(0.009)	C1	13
P006101	0.845	(0.072)	-0.839	(0.087)	0.443	(0.025)	C1	16
P006201	1.037	(0.083)	-0.750	(0.079)	0.463	(0.023)	C1	17
P006301	1.294	(0.100)	-0.429	(0.059)	0.314	(0.020)	C1	18
P006401	0.504	(0.055)	0.351	(0.056)	0.475	(0.014)	C1	19
P006501	0.899	(0.084)	-0.675	(0.077)	0.506	(0.019)	C1	20
P006601	0.552	(0.043)	-0.187	(0.031)	0.247	(0.016)	C1	21
P006701	0.968	(0.072)	-0.878	(0.078)	0.242	(0.027)	C1	22
P006801	0.843	(0.080)	-0.721	(0.085)	0.431	(0.024)	C1	23
P006802	0.381	(0.053)	-0.363	(0.066)	0.498	(0.018)	C1	24
P006803	1.061	(0.066)	1.607	(0.128)	0.369	(0.010)	C1	25
P006804	0.887	(0.063)	0.492	(0.055)	0.389	(0.012)	C1	26
P007201	1.490	(0.090)	-0.704	(0.065)	0.244	(0.024)	C1	29
P007301	0.839	(0.086)	-0.840	(0.099)	0.530	(0.021)	C1	30
P007401	0.963	(0.060)	-1.555	(0.110)	0.473	(0.029)	C1	31
P007501	0.410	(0.063)	-3.651	(0.562)	0.476	(0.041)	C1	32
P007502	0.505	(0.053)	-1.014	(0.114)	0.473	(0.022)	C1	33
P007503	0.924	(0.078)	-1.071	(0.105)	0.479	(0.027)	C1	34
P008201	0.923	(0.068)	-1.318	(0.112)	0.506	(0.027)	C9	1
P008202	1.300	(0.076)	-1.343	(0.106)	0.476	(0.030)	C9	2
P008301	0.852	(0.053)	-1.108	(0.080)	0.252	(0.026)	C9	3

Table F-5 (continued)

1988 IRT Parameters, Civics Trend, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 17 BLOCK	AGE 17 ITEM
P008401	1.166	(0.068)	2.014	(0.154)	0.216	(0.009)	C9	5
P008402	2.414	(0.284)	0.010	(0.074)	0.393	(0.015)	C9	6
P008403	1.406	(0.105)	-0.131	(0.051)	0.371	(0.017)	C9	7
P008404	2.417	(0.100)	1.215	(0.125)	0.364	(0.011)	C9	8
P008405	1.664	(0.186)	-0.067	(0.064)	0.463	(0.017)	C9	9
P008406	1.277	(0.069)	0.763	(0.071)	0.348	(0.013)	C9	10
P008407	1.637	(0.108)	0.401	(0.071)	0.353	(0.015)	C9	11
P008501	0.796	(0.061)	0.638	(0.066)	0.279	(0.015)	C9	12
P008601	1.271	(0.065)	2.356	(0.164)	0.089	(0.007)	C9	14
P008701	0.888	(0.058)	0.816	(0.070)	0.245	(0.012)	C9	4
P009001	1.482	(0.066)	0.874	(0.067)	0.211	(0.011)	C9	16
P009101	0.948	(0.061)	0.994	(0.082)	0.201	(0.013)	C9	17
P009201	1.138	(0.057)	1.310	(0.088)	0.188	(0.010)	C9	18
P009301	0.886	(0.059)	1.579	(0.124)	0.199	(0.012)	C9	19
P009401	0.474	(0.039)	-1.105	(0.096)	0.228	(0.024)	C9	20
P009501	1.027	(0.055)	1.579	(0.105)	0.168	(0.009)	C9	21
P009601	0.653	(0.060)	-3.201	(0.303)	0.236	(0.047)	C9	22
P009801	0.669	(0.055)	-0.991	(0.090)	0.239	(0.026)	C9	24
P009901	1.397	(0.069)	0.420	(0.048)	0.315	(0.011)	C9	25
P010101	0.550	(0.045)	0.595	(0.058)	0.295	(0.012)	C9	23
P020701	1.052	(0.089)	-0.435	(0.056)	0.267	(0.021)	C1	9
P021001	1.071	(0.049)	-1.238	(0.069)	0.000	(0.000)	C1	33
P021101	0.593	(0.042)	-2.269	(0.165)	0.000	(0.000)	C1	34

Table F-6

1988 IRT Parameters, U.S. History Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
H000101	1.562	(0.014)	1.061	(0.019)	0.223	(0.005)	--	--	--	--	H3	1
H000201	1.575	(0.014)	1.549	(0.024)	0.205	(0.004)	--	--	--	--	H3	3
H000301	0.760	(0.016)	0.019	(0.014)	0.272	(0.009)	--	--	--	--	H3	4
H000401	1.558	(0.021)	1.679	(0.039)	0.496	(0.004)	--	--	--	--	H3	5
H000601	0.723	(0.012)	0.529	(0.015)	0.151	(0.008)	--	--	--	--	H3	6
H000801	1.197	(0.034)	-0.459	(0.024)	0.286	(0.013)	--	--	--	--	H3	7
H000901	1.687	(0.013)	1.444	(0.021)	0.171	(0.004)	--	--	--	--	H3	8
H001101	1.561	(0.016)	1.036	(0.020)	0.297	(0.005)	--	--	--	--	H3	9
H001301	2.153	(0.022)	0.683	(0.021)	0.334	(0.006)		--	--	--	H3	13
H001401	0.624	(0.012)	0.458	(0.015)	0.230	(0.008)	--	--	--	--	H3	14
H001501	1.257	(0.014)	0.602	(0.015)	0.144	(0.007)	--	--	--	--	H3	15
H001701	1.582	(0.014)	1.301	(0.021)	0.218	(0.005)	--	--	--	--	H3	17
H001801	1.452	(0.016)	1.833	(0.031)	0.245	(0.004)	--	--	--	--	H3	18
H001901	0.774	(0.012)	0.990	(0.020)	0.196	(0.006)	--	--	--	--	H3	19
H002001	1.058	(0.013)	0.843	(0.017)	0.215	(0.006)	--	--	--	--	H3	20
H002101	1.309	(0.013)	1.422	(0.023)	0.229	(0.005)	--	--	--	--	H3	21
H002201	1.064	(0.016)	0.494	(0.017)	0.299	(0.008)	--	--	--	--	H3	22
H002301	2.722	(0.024)	2.094	(0.049)	0.198	(0.003)	--	--	--	--	H3	23
H002601	1.964	(0.016)	1.483	(0.025)	0.253	(0.004)	--	--	--	--	H3	24
H002701	1.188	(0.013)	1.466	(0.023)	0.192	(0.005)	--	--	--	--	H3	25
H003001	1.201	(0.016)	0.078	(0.009)	0.232	(0.005)	--	--	H5	1	H5	1
H003101	1.119	(0.011)	0.872	(0.012)	0.152	(0.003)	--	--	H5	2	H5	2
H003201	1.825	(0.013)	1.594	(0.022)	0.204	(0.002)	--	--	H5	3	H5	3
H003301	0.818	(0.010)	0.933	(0.015)	0.202	(0.004)	--	--	H5	4	H5	4
H003601	1.555	(0.012)	1.113	(0.016)	0.261	(0.003)	--	--	H5	7	H5	7
H003701	1.467	(0.016)	1.778	(0.031)	0.349	(0.003)	--	--	H5	8	H5	8
H003801	1.538	(0.012)	1.624	(0.021)	0.164	(0.002)	--	--	H5	9	H5	9
H004001	0.937	(0.011)	1.234	(0.018)	0.259	(0.003)	--	--	H5	10	H5	10
H004301	0.649	(0.015)	2.217	(0.056)	0.384	(0.003)	--	--	H5	12	H5	12
H004401	1.359	(0.014)	0.761	(0.014)	0.327	(0.004)	--	--	H5	13	H5	13
H004501	0.526	(0.011)	-0.775	(0.018)	0.204	(0.007)	--	--	H5	14	H5	14
H004502	0.723	(0.011)	0.166	(0.008)	0.191	(0.005)	--	--	H5	15	H5	15
H004601	0.747	(0.009)	1.097	(0.016)	0.142	(0.003)	--	--	H5	17	H5	17
H004901	0.825	(0.011)	0.180	(0.008)	0.163	(0.005)	--	--	H5	18	H5	18
H005001	2.204	(0.018)	0.577	(0.014)	0.280	(0.004)	--	--	H5	19	H5	19
H005201	1.451	(0.011)	0.970	(0.013)	0.191	(0.003)	--	--	H5	22	H5	22
H005301	0.539	(0.010)	1.340	(0.026)	0.211	(0.004)	--	--	H5	23	H5	23
H005401	1.318	(0.013)	0.573	(0.011)	0.216	(0.004)	--	--	H5	24	H5	24
H005501	1.526	(0.016)	1.865	(0.031)	0.284	(0.003)	--	--	H5	25	H5	25
H005801	0.882	(0.014)	-0.130	(0.009)	0.231	(0.006)	--	--	H5	26	H5	26
H005901	1.632	(0.015)	0.939	(0.018)	0.215	(0.005)	--	--	--	--	H4	1
H006001	1.470	(0.017)	0.541	(0.016)	0.205	(0.007)	--	--	--	--	H4	2
H006101	1.367	(0.029)	-0.138	(0.018)	0.228	(0.011)	--	--	--	--	H4	3
H006401	1.551	(0.013)	1.920	(0.027)	0.113	(0.003)	--	--	--	--	H4	6
H006601	1.445	(0.013)	1.442	(0.023)	0.223	(0.004)	--	--	--	--	H4	7
H006701	1.858	(0.020)	2.208	(0.044)	0.164	(0.003)	--	--	--	--	H4	8
H006801	1.864	(0.018)	0.771	(0.020)	0.303	(0.006)	--	--	--	--	H4	9
H007001	1.845	(0.015)	1.090	(0.020)	0.231	(0.005)	--	--	--	--	H4	13
H007101	1.159	(0.017)	2.117	(0.041)	0.250	(0.004)	--	--	--	--	H4	10
H007102	1.787	(0.015)	1.037	(0.019)	0.238	(0.005)	--	--	--	--	H4	11
H007103	1.196	(0.016)	0.790	(0.019)	0.339	(0.006)	--	--	--	--	H4	12
H007201	0.671	(0.013)	1.467	(0.033)	0.314	(0.005)	--	--	--	--	H4	14
H007301	1.757	(0.015)	1.882	(0.029)	0.153	(0.003)	--	--	--	--	H4	15
H007401	1.795	(0.016)	1.905	(0.033)	0.219	(0.004)	--	--	--	--	H4	16
H007501	1.342	(0.013)	1.201	(0.020)	0.219	(0.005)	--	--	--	--	H4	17
H007601	1.341	(0.014)	1.626	(0.026)	0.233	(0.004)	--	--	--	--	H4	18
H007701	1.965	(0.021)	0.684	(0.020)	0.312	(0.006)	--	--	--	--	H4	19
H007801	1.129	(0.012)	1.083	(0.018)	0.164	(0.005)	--	--	--	--	H4	20
H008101	0.771	(0.011)	1.072	(0.019)	0.145	(0.006)	--	--	--	--	H4	22
H008201	0.878	(0.013)	1.846	(0.032)	0.208	(0.004)	--	--	--	--	H4	23
H008501	1.140	(0.031)	-0.624	(0.027)	0.235	(0.014)	--	--	--	--	H4	24
H008701	1.362	(0.014)	1.588	(0.025)	0.233	(0.004)	--	--	--	--	H4	25
H008801	1.094	(0.013)	1.143	(0.021)	0.285	(0.005)	--	--	--	--	H3	2
H008801	2.411	(0.015)	1.046	(0.017)	0.292	(0.003)	--	--	H5	5	H5	5

Table F-6 (continued)

1988 IRT Parameters, U.S. History Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
H009201	0.813	(0.014)	[illegible]	(0.016)	0.254	(0.008)	--	--	--	--	H3	10
H009401	0.684	(0.014)	0.295	(0.014)	0.238	(0.008)	--	--	--	--	H4	4
H009501	1.341	(0.014)	1.205	(0.022)	0.299	(0.005)	--	--	--	--	H3	11
H009601	2.409	(0.017)	1.754	(0.030)	0.193	(0.004)	--	--	--	--	H3	12
H009801	1.854	(0.025)	0.439	(0.020)	0.346	(0.008)	--	--	--	--	H3	16
H009901	1.044	(0.022)	-0.324	(0.018)	0.214	(0.012)	--	--	--	--	H4	5
H010101	1.536	(0.015)	0.982	(0.019)	0.270	(0.006)	--	--	--	--	H4	21
H010201	1.125	(0.013)	1.091	(0.018)	0.362	(0.003)	--	--	H5	11	H5	11
H010401	1.787	(0.012)	1.403	(0.017)	0.163	(0.002)	--	--	H5	16	H5	16
H010801	0.648	(0.011)	1.634	(0.029)	0.226	(0.003)	--	--	H5	20	H5	20
H011001	1.737	(0.013)	0.697	(0.012)	0.145	(0.003)	--	--	H5	21	H5	21
H011501	0.761	(0.010)	-1.141	(0.018)	0.308	(0.006)	H2	1	H2	1	--	--
H011601	1.233	(0.010)	-1.617	(0.017)	0.166	(0.008)	H2	2	H2	2	--	--
H011701	1.074	(0.009)	-2.023	(0.021)	0.167	(0.011)	H2	3	H2	3	--	--
H011801	0.912	(0.009)	-1.591	(0.018)	0.158	(0.008)	H2	4	H2	4	--	--
H011901	1.412	(0.020)	-0.806	(0.017)	0.487	(0.005)	H2	5	H2	5	--	--
H012001	0.831	(0.009)	-1.619	(0.020)	0.202	(0.008)	H2	6	H2	6	--	--
H012101	2.009	(0.024)	-0.471	(0.013)	0.264	(0.004)	H2	7	H2	7	--	--
H012201	1.246	(0.014)	-0.641	(0.012)	0.213	(0.004)	H2	8	H2	8	--	--
H012301	1.654	(0.019)	-0.400	(0.010)	0.170	(0.003)	H2	9	H2	9	--	--
H012401	1.931	(0.018)	-0.655	(0.012)	0.184	(0.004)	H2	10	H2	10	--	--
H012501	2.601	(0.032)	-0.384	(0.013)	0.236	(0.003)	H2	11	H2	11	--	--
H012701	0.803	(0.012)	-0.529	(0.011)	0.214	(0.004)	H2	13	H2	13	--	--
H012801	1.581	(0.020)	-0.471	(0.012)	0.255	(0.004)	H2	14	H2	14	--	--
H012901	2.293	(0.035)	-0.170	(0.012)	0.210	(0.003)	H2	15	H2	15	--	--
H013001	1.421	(0.013)	-0.954	(0.014)	0.214	(0.005)	H3	1	H3	1	--	--
H013101	0.885	(0.012)	-0.908	(0.016)	0.322	(0.005)	H3	2	H3	2	--	--
H013201	1.757	(0.013)	-0.980	(0.013)	0.146	(0.005)	H3	3	H3	3	--	--
H013301	0.860	(0.013)	-0.242	(0.009)	0.183	(0.004)	H3	4	H3	4	--	--
H013401	1.413	(0.025)	-0.127	(0.011)	0.278	(0.003)	H3	5	H3	5	--	--
H013501	0.868	(0.011)	-0.580	(0.011)	0.179	(0.005)	H3	6	H3	6	--	--
H013502	1.374	(0.012)	-1.055	(0.014)	0.216	(0.006)	H3	7	H3	7	--	--
H013701	0.426	(0.009)	-1.059	(0.023)	0.220	(0.006)	H3	8	H3	8	--	--
H013801	1.434	(0.026)	-0.030	(0.011)	0.267	(0.003)	H3	9	H3	9	--	--
H013901	1.342	(0.016)	-0.590	(0.012)	0.220	(0.004)	H3	10	H3	10	--	--
H014001	1.609	(0.022)	-0.248	(0.011)	0.158	(0.003)	H3	11	H3	11	--	--
H014101	1.637	(0.017)	-0.666	(0.013)	0.208	(0.004)	H3	12	H3	12	--	--
H014201	1.763	(0.033)	0.132	(0.013)	0.248	(0.003)	H3	13	H3	13	--	--
H014301	1.635	(0.028)	0.478	(0.017)	0.276	(0.003)	H3	14	H3	14	--	--
H014401	1.666	(0.023)	-0.198	(0.010)	0.139	(0.003)	H3	15	H3	15	--	--
H014501	1.867	(0.016)	-1.167	(0.018)	0.385	(0.006)	H4	1	H4	1	--	--
H014601	2.344	(0.018)	-1.123	(0.019)	0.355	(0.006)	H4	2	H4	2	--	--
H014701	1.277	(0.010)	-1.587	(0.017)	0.163	(0.009)	H4	5	H4	5	--	--
H014702	0.902	(0.009)	-1.613	(0.018)	0.122	(0.008)	H4	6	H4	6	--	--
H014703	1.590	(0.022)	-0.301	(0.011)	0.197	(0.004)	H4	7	H4	7	--	--
H015001	0.904	(0.010)	-0.831	(0.012)	0.094	(0.005)	H4	3	H4	3	--	--
H015101	0.785	(0.014)	-0.030	(0.009)	0.199	(0.004)	H4	4	H4	4	--	--
H015201	1.356	(0.022)	-0.383	(0.013)	0.330	(0.004)	H4	8	H4	8	--	--
H015301	0.801	(0.010)	-0.872	(0.014)	0.176	(0.006)	H4	9	H4	9	--	--
H015401	1.283	(0.014)	-0.544	(0.010)	0.118	(0.004)	H4	10	H4	10	--	--
H015501	2.399	(0.044)	0.258	(0.018)	0.323	(0.003)	H4	11	H4	11	--	--
H015601	1.046	(0.021)	0.236	(0.012)	0.269	(0.003)	H4	12	H4	12	--	--
H015701	1.040	(0.013)	-0.514	(0.010)	0.130	(0.004)	H4	13	H4	13	--	--
H015801	1.023	(0.019)	0.506	(0.014)	0.174	(0.003)	H4	14	H4	14	--	--
H015901	1.613	(0.028)	0.038	(0.012)	0.245	(0.003)	H4	15	H4	15	--	--
H016001	0.665	(0.017)	-1.459	(0.040)	0.221	(0.012)	--	--	H2	16	--	--
H016101	2.267	(0.051)	-0.368	(0.018)	0.227	(0.007)	--	--	H2	17	--	--
H016201	0.456	(0.015)	-0.612	(0.022)	0.200	(0.008)	--	--	H2	18	--	--
H016301	2.096	(0.035)	0.333	(0.018)	0.352	(0.005)	--	--	H2	19	--	--
H016401	1.511	(0.035)	-0.279	(0.015)	0.279	(0.007)	--	--	H2	20	--	--
H016501	1.922	(0.042)	0.113	(0.016)	0.357	(0.005)	--	--	H2	21	--	--
H016601	1.983	(0.036)	0.137	(0.014)	0.230	(0.005)	--	--	H2	22	--	--
H016701	1.551	(0.026)	0.166	(0.013)	0.176	(0.005)	--	--	H2	23	--	--
H016801	2.766	(0.035)	0.422	(0.019)	0.253	(0.004)	--	--	H2	24	--	--

Table F-6 (continued)

1988 IRT Parameters, U.S. History Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
H016901	1.500	(0.022)	0.695	(0.018)	0.209	(0.004)	--	--	H2	25	--	--
H017001	0.488	(0.023)	2.198	(0.105)	0.222	(0.004)	--	--	H2	26	--	--
H017101	1.120	(0.023)	0.500	(0.018)	0.245	(0.005)	--	--	H3	16	--	--
H017102	1.529	(0.038)	-0.593	(0.023)	0.294	(0.010)	--	--	H3	17	--	--
H017301	1.310	(0.027)	0.638	(0.023)	0.340	(0.005)	--	--	H3	18	--	--
H017401	1.119	(0.024)	0.445	(0.018)	0.253	(0.006)	--	--	H3	19	--	--
H017501	1.167	(0.029)	-0.372	(0.017)	0.215	(0.009)	--	--	H3	20	--	--
H017601	1.757	(0.031)	0.359	(0.018)	0.234	(0.005)	--	--	H3	21	--	--
H017701	1.387	(0.032)	1.459	(0.048)	0.331	(0.004)	--	--	H3	22	--	--
H017801	0.941	(0.022)	0.564	(0.020)	0.238	(0.006)	--	--	H3	23	--	--
H017901	1.952	(0.028)	1.369	(0.037)	0.186	(0.003)	--	--	H3	24	--	--
H018001	2.571	(0.026)	1.109	(0.029)	0.154	(0.003)	--	--	H3	25	--	--
H018101	1.319	(0.023)	1.219	(0.030)	0.186	(0.004)	--	--	H3	26	--	--
H018201	0.818	(0.020)	-0.289	(0.014)	0.152	(0.008)	--	--	H4	16	--	--
H018301	1.724	(0.036)	0.075	(0.015)	0.191	(0.006)	--	--	H4	17	--	--
H018401	1.890	(0.035)	0.266	(0.018)	0.271	(0.006)	--	--	H4	18	--	--
H018501	1.016	(0.025)	1.112	(0.034)	0.295	(0.005)	--	--	H4	19	--	--
H018601	1.258	(0.025)	0.567	(0.019)	0.235	(0.005)	--	--	H4	20	--	--
H018701	0.923	(0.022)	[illegible]	[illegible]	0.203	(0.006)	--	--	H4	21	--	--
H018801	2.143	(0.027)	[illegible]	[illegible]	0.207	(0.004)	--	--	H4	22	--	--
H018901	2.108	(0.026)	0.772	(0.022)	0.198	(0.004)	--	--	H4	23	--	--
H019001	1.139	(0.023)	0.955	(0.025)	0.188	(0.004)	--	--	H4	24	--	--
H019101	1.373	(0.026)	[illegible]	(0.030)	0.288	(0.004)	--	--	H4	25	--	--
H019201	2.390	(0.025)	[illegible]	(0.028)	0.122	(0.003)	--	--	H4	26	--	--
H019301	0.507	(0.012)	[illegible]	(0.043)	0.246	(0.009)	--	--	H6	1	H6	1
H019401	1.048	(0.013)	-0.026	(0.008)	0.181	(0.005)	--	--	H6	2	H6	2
H019501	1.206	(0.017)	-0.115	(0.009)	0.243	(0.005)	--	--	H6	3	H6	3
H019601	1.068	(0.013)	0.922	(0.016)	0.452	(0.003)	--	--	H6	4	H6	4
H019701	1.007	(0.017)	-0.564	(0.013)	0.229	(0.007)	--	--	H6	5	H6	5
H019801	0.893	(0.012)	-0.112	(0.009)	0.287	(0.005)	--	--	H6	6	H6	6
H019901	1.060	(0.020)	-1.263	(0.028)	0.259	(0.011)	--	--	H6	10	H6	10
H019902	0.735	(0.012)	-0.850	(0.016)	0.162	(0.008)	--	--	H6	11	H6	11
H020101	1.121	(0.014)	0.319	(0.011)	0.146	(0.004)	--	--	H6	8	H6	8
H020201	1.959	(0.022)	0.224	(0.011)	0.276	(0.004)	--	--	H6	7	H6	7
H020301	1.657	(0.013)	0.816	(0.014)	0.350	(0.003)	--	--	H6	12	H6	12
H020401	1.327	(0.010)	0.875	(0.012)	0.210	(0.003)	--	--	H6	9	H6	9
H020501	1.414	(0.014)	0.517	(0.012)	0.357	(0.004)	--	--	H6	13	H6	13
H020601	0.751	(0.010)	0.087	(0.008)	0.217	(0.005)	--	--	H6	14	H6	14
H020701	1.081	(0.011)	0.572	(0.010)	0.202	(0.004)	--	--	H6	15	H6	15
H020801	1.500	(0.012)	0.436	(0.009)	0.118	(0.003)	--	--	H6	16	H6	16
H020901	[illegible]	(0.014)	0.426	(0.010)	0.181	(0.004)	--	--	H6	17	H6	17
H021001	1.866	(0.014)	0.638	(0.013)	0.299	(0.003)	--	--	H6	18	H6	18
H021101	1.376	(0.011)	0.627	(0.010)	0.169	(0.003)	--	--	H6	19	H6	19
H021201	0.879	(0.011)	0.589	(0.012)	0.349	(0.004)	--	--	H6	20	H6	20
H021301	1.862	(0.013)	0.690	(0.011)	0.198	(0.003)	--	--	H6	21	H6	21
H021401	1.595	(0.017)	0.232	(0.010)	0.249	(0.004)	--	--	H6	22	H6	22
H021501	1.022	(0.008)	1.337	(0.014)	0.115	(0.002)	--	--	H6	23	H6	23
H021601	1.390	(0.011)	1.077	(0.015)	0.289	(0.003)	--	--	H6	24	H6	24
H021701	2.606	(0.016)	1.713	(0.028)	0.295	(0.002)	--	--	H6	25	H6	25
H021801	1.705	(0.012)	1.496	(0.020)	0.287	(0.003)	--	--	H6	26	H6	26
H021901	2.025	(0.012)	1.186	(0.016)	0.255	(0.003)	--	--	H6	27	H6	27
H022001	1.657	(0.013)	1.886	(0.026)	0.225	(0.002)	--	--	H6	28	H6	28
H022101	1.157	(0.015)	0.102	(0.010)	0.233	(0.005)	--	--	H7	1	H7	1
H022201	0.647	(0.013)	-1.047	(0.024)	0.200	(0.009)	--	--	H7	2	H7	2
H022301	0.896	(0.011)	0.674	(0.012)	0.179	(0.004)	--	--	H7	3	H7	3
H022501	1.479	(0.017)	0.244	(0.011)	0.224	(0.005)	--	--	H7	5	H7	5
H022601	1.379	(0.021)	0.006	(0.012)	0.303	(0.006)	--	--	H7	6	H7	6
H022701	0.981	(0.011)	0.576	(0.011)	0.167	(0.004)	--	--	H7	7	H7	7
H022801	1.064	(0.017)	0.003	(0.011)	0.370	(0.006)	--	--	H7	8	H7	8
H022802	1.082	(0.013)	0.665	(0.014)	0.352	(0.004)	--	--	H7	9	H7	9
H022901	1.861	(0.014)	0.712	(0.012)	0.139	(0.003)	--	--	H7	10	H7	10
H023101	2.186	(0.015)	0.869	(0.016)	0.248	(0.003)	--	--	H7	11	H7	11
H023201	0.936	(0.012)	0.308	(0.010)	0.193	(0.005)	--	--	H7	12	H7	12
H023301	1.065	(0.013)	0.617	(0.013)	0.285	(0.004)	--	--	H7	13	H7	13

Table F-6 (continued)

1988 IRT Parameters, U.S. History Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
H023401	0.799	(0.011)	0.677	(0.014)	0.253	(0.004)	--	--	H7	14	H7	14
H023501	1.351	(0.012)	0.810	(0.013)	0.194	(0.004)	--	--	H7	15	H7	15
H023601	0.766	(0.011)	0.547	(0.012)	0.194	(0.005)	--	--	H7	16	H7	16
H023701	1.843	(0.012)	1.101	(0.016)	0.204	(0.003)	--	--	H7	17	H7	17
H023801	2.350	(0.016)	0.793	(0.015)	0.221	(0.003)	--	--	H7	18	H7	18
H023901	1.500	(0.013)	0.804	(0.014)	0.265	(0.004)	--	--	H7	19	H7	19
H024001	2.335	(0.016)	1.683	(0.028)	0.252	(0.003)	--	--	H7	20	H7	20
H024101	1.069	(0.010)	0.843	(0.012)	0.079	(0.003)	--	--	H7	21	H7	21
H024201	3.021	(0.016)	1.390	(0.024)	0.238	(0.003)	--	--	H7	22	H7	22
H024301	1.282	(0.013)	1.595	(0.023)	0.221	(0.003)	--	--	H7	23	H7	23
H024401	1.776	(0.015)	0.800	(0.015)	0.259	(0.004)	--	--	H7	24	H7	24
H024501	1.669	(0.013)	1.103	(0.017)	0.250	(0.003)	--	--	H7	25	H7	25
H024601	1.775	(0.013)	1.687	(0.024)	0.182	(0.003)	--	--	H7	26	H7	26
H024701	1.622	(0.014)	1.682	(0.026)	0.258	(0.003)	--	--	H7	27	H7	27
H024801	3.042	(0.021)	1.942	(0.041)	0.235	(0.002)	--	--	H7	28	H7	28
H024901	1.194	(0.031)	-0.153	(0.016)	0.296	(0.008)	--	--	H8	1	--	--
H025003	1.018	(0.013)	1.776	(0.028)	0.168	(0.004)	--	--	--	--	H8	1
H025101	0.915	(0.012)	0.439	(0.013)	0.218	(0.007)	--	--	--	--	H2	3
H025201	1.866	(0.014)	1.311	(0.022)	0.349	(0.004)	--	--	--	--	H2	4
H025301	1.247	(0.012)	1.050	(0.017)	0.275	(0.005)	--	--	-	--	H2	5
H025401	1.303	(0.012)	1.641	(0.024)	0.280	(0.004)	--	--	--	--	H2	6
H025501	0.891	(0.015)	0.170	(0.013)	0.288	(0.008)	--	--	--	--	H2	7
H025502	0.955	(0.013)	0.306	(0.013)	0.220	(0.007)	--	--	--	--	H2	8
H025701	1.115	(0.013)	1.344	(0.023)	0.385	(0.005)	--	--	--	--	H2	9
H025801	1.006	(0.011)	0.744	(0.014)	0.164	(0.006)	--	--	--	--	H2	10
H025901	1.389	(0.012)	0.934	(0.015)	0.195	(0.005)	--	--	--	--	H2	11
H026001	0.910	(0.010)	1.020	(0.016)	0.178	(0.005)	--	--	--	--	H2	12
H026101	1.168	(0.012)	1.203	(0.019)	0.285	(0.005)	--	--	--	--	H2	13
H026201	1.188	(0.013)	0.479	(0.013)	0.172	(0.007)	--	--	--	-	[illegible]	14
H026301	1.218	(0.013)	0.617	(0.013)	0.150	(0.006)	--	--	--	--	[illegible]	15
H026401	1.316	(0.014)	0.760	(0.016)	0.314	(0.006)	--	--	--	--	H2	16
H026501	1.337	(0.012)	1.421	(0.020)	0.250	(0.004)	--	--	--	--	H2	17
H026601	1.224	(0.012)	1.021	(0.016)	0.218	(0.005)	--	--	--	--	H2	18
H026701	0.977	(0.012)	1.400	(0.022)	0.280	(0.005)		--	--	--	H2	20
H026801	0.744	(0.010)	[illegible]	(0.015)	0.156	(0.006)	--	--	--	--	H2	19
H026901	1.001	(0.011)	[illegible]	(0.013)	0.130	(0.006)	--	--	-	--	H2	21
H027001	1.223	(0.011)	1.257	(0.018)	0.215	(0.005)	--	--	--	--	H2	22
H027101	2.452	(0.014)	1.717	(0.026)	0.196	(0.003)	--	--	--	--	H2	23
H027201	1.205	(0.012)	1.592	(0.023)	0.268	(0.004)	--	--	--	--	H2	24
H027301	1.820	(0.014)	1.695	(0.026)	0.281	(0.004)	--	--	--	--	H2	25
H027401	1.590	(0.014)	1.924	(0.029)	0.202	(0.003)	--	--	--	--	H2	26
H027501	2.060	(0.015)	1.835	(0.029)	0.250	(0.004)	--	--	--	--	H2	27
H027601	2.026	(0.019)	2.105	(0.039)	0.226	(0.003)	--	--	--	-	H2	28
H028001	1.379	(0.022)	0.148	(0.017)	0.392	(0.008)	--	--	--	--	H2	1
H028101	0.814	(0.012)	0.372	(0.012)	0.186	(0.007)	--	--	--	--	H2	2

Table F-7

1988 IRT Parameters, Geography Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
G000101	0.823	(0.030)	-0.168	(0.022)	0.291	(0.010)	--	--	--	--	G2	1
G000201	0.622	(0.025)	-0.271	(0.022)	0.286	(0.010)	--	--	--	--	G2	2
G000301	0.567	(0.023)	-1.238	(0.055)	0.300	(0.015)	--	--	--	--	G2	3
G000401	0.722	(0.028)	-0.013	(0.020)	0.30[illegible]	(0.009)	--	--	--	--	G2	4
G000501	1.109	(0.035)	-0.578	(0.028)	0.187	(0.012)	--	--	--	--	[illegible]	5
G000601	0.870	(0.027)	-0.783	(0.031)	0.195	(0.013)	--	--	--	--	G2	6
G000701	0.853	(0.032)	-0.070	(0.021)	0.239	(0.009)	--	--	--	--	G2	7
G000801	0.938	(0.029)	0.295	(0.022)	0.201	(0.008)	--	--	--	--	G2	8
G000901	0.380	(0.021)	0.471	(0.032)	0.280	(0.009)	--	--	--	--	G2	9
G001001	0.383	(0.019)	0.130	(0.019)	0.231	(0.010)	--	--	--	--	G2	10
G001101	0.992	(0.032)	1.552	(0.062)	0.2[illegible]	(0.006)	--	--	--	--	G2	11
G001201	0.592	(0.021)	-0.392	(0.022)	0.180	(0.011)	--	--	--	--	G2	12
G001301	1.685	(0.041)	0.834	(0.043)	0.422	(0.006)	--	--	--	--	G2	14
G001302	1.060	(0.031)	0.538	(0.027)	0.223	(0.007)	--	--	--	--	G2	15
G001401	0.786	(0.031)	0.431	(0.029)	0.354	(0.008)	--	--	--	--	G2	16
G001501	1.358	(0.039)	0.421	(0.028)	0.259	(0.007)	--	--	--	--	G2	17
G001601	1.084	(0.030)	0.578	(0.029)	0.259	(0.007)	--	--	--	--	G2	13
G001701	1.050	(0.031)	0.741	(0.033)	0.275	(0.007)	--	--	--	--	G2	18
G001801	0.710	(0.022)	0.206	(0.018)	0.114	(0.008)	--	--	--	--	G2	19
G001901	0.661	(0.029)	0.490	(0.031)	0.336	(0.008)	--	--	--	--	G2	22
G002001	1.704	(0.035)	1.429	(0.053)	0.263	(0.005)	--	--	--	--	G2	20
G002101	1.252	(0.033)	1.582	(0.058)	0.245	(0.005)	--	--	--	--	G2	21
G002201	1.452	(0.038)	1.661	[illegible]	0.277	(0.005)	--	--	--	--	G2	23
G002301	1.296	(0.030)	1.278	(0.044)	0.209	(0.006)	--	--	--	--	G2	24
G002401	0.969	(0.035)	1.952	(0.063)	0.211	(0.005)	--	--	--	--	G2	25
G002501	0.886	(0.033)	-0.683	(0.034)	0.289	(0.013)	--	--	--	--	G2	26
G002601	1.001	(0.034)	0.261	(0.026)	0.256	(0.009)	--	--	--	--	G3	1
G002602	0.479	(0.024)	0.484	(0.033)	0.244	(0.010)	--	--	--	--	G3	2
G002701	0.908	(0.035)	-0.432	(0.028)	0.229	(0.012)	--	--	--	--	G3	3
G002801	0.800	(0.031)	-0.507	(0.030)	0.244	(0.013)	--	--	-	--	G3	4
G002901	1.121	(0.041)	-0.340	(0.028)	0.238	(0.012)	--	--	--	--	G3	5
G003001	0.658	(0.029)	-0.168	(0.023)	0.262	(0.011)	--	--	--	--	G3	6
G003101	1.051	(0.034)	-0.380	(0.025)	0.131	(0.012)	--	--	--	--	G3	7
G003201	1.059	(0.036)	0.846	(0.043)	0.325	(0.008)	--	--	--	--	G3	8
G003301	1.370	(0.046)	0.451	(0.037)	0.390	(0.008)	--	--	--	--	G3	9
G003401	0.404	(0.022)	-0.303	(0.026)	0.238	(0.012)	--	--	--	--	G3	10
G003501	0.858	(0.032)	0.568	(0.032)	0.233	(0.008)	--	--	--	--	G3	11
G003502	1.046	(0.029)	1.094	(0.042)	0.158	(0.006)	--	--	--	--	G3	12
G003601	1.367	(0.039)	0.549	(0.035)	0.312	(0.008)	--	--	--	--	G3	13
G003701	0.704	(0.029)	0.964	(0.048)	0.233	(0.008)	--	--	--	--	G3	14
G003801	1.225	(0.03[illegible])	0.307	(0.026)	0.170	(0.008)	--	--	--	--	G3	15
G003901	0.888	(0.034)	0.709	(0.039)	0.301	(0.008)	--	--	--	--	G3	16
G004001	1.184	(0.035)	0.880	(0.042)	0.296	(0.007)	--	--	--	--	C3	17
G004101	0.713	(0.032)	1.221	(0.062)	0.262	(0.008)	--	--	--	--	G3	18
G004201	0.949	(0.029)	1.320	(0.049)	0.142	(0.006)	--	--	--	--	G3	19
G004301	0.984	(0.032)	0.585	(0.032)	0.213	(0.008)	--		--	--	G3	20
G004401	0.897	(0.032)	-1.138	(0.049)	0.251	(0.018)	--		--	--	G3	21
G004402	0.934	(0.033)	0.716	(0.037)	0.250	(0.008)	--	--	--	--	G3	22
G004403	1.772	(0.037)	1.171	(0.048)	0.221	(0.006)	--	--	--	--	G3	23
G004404	1.386	(0.034)	1.428	(0.054)	0.168	(0.006)	--	--	--	--	G3	24
G004501	0.3[illegible]	(0.031)	2.117	(0.172)	0.270	(0.008)	--	--	--	--	G3	25
G004601	1.741	(0.087)	2.272	(0.194)	0.286	(0.006)	--	--	--	--	G3	26
G004701	1.166	(0.035)	0.003	(0.021)	0.186	(0.008)	--	--	--	--	G4	1
G004801	1.191	(0.042)	0.121	(0.027)	0.391	(0.008)	--	--	--	--	G4	2
G004901	1.041	(0.036)	0.156	(0.024)	0.339	(0.008)	--	--	--	--	G4	3
G005001	1.122	(0.035)	-0.228	(0.022)	0.187	(0.010)	--	--	--	--	G4	4
G005002	1.765	(0.053)	-0.778	(0.041)	0.298	(0.013)	--	--	--	--	G4	5
G005101	1.066	(0.032)	0.317	(0.024)	0.[illegible]47	(0.008)	--	--	--	--	G4	6
G005102	1.292	(0.045)	-0.216	(0.025)	0.275	(0.010)	--	--	--	--	G4	7
G005201	0.814	(0.028)	-0.100	(0.020)	0.224	(0.010)	--	--	--	--	G4	8
G005202	0.724	(0.024)	0.315	(0.021)	0.195	(0.008)	--	--	--	--	[illegible]	9
G005203	1.543	(0.033)	0.735	(0.032)	0.231	(0.006)	--	--	--	--	G4	10
G005301	1.446	(0.041)	0.034	(0.021)	0.141	(0.008)	--	--	--	--	G4	11
G005401	0.808	(0.027)	-0.27[illegible]	(0.020)	0.172	(0.010)	--	--	--	--	[illegible]	12

Table F-7 (continued)

1988 IRT Parameters, Geography Cross-sectional

NAEP ID	A	S.E.	B	S.E.	C	S.E.	GRADE 4/AGE 9 BLOCK	GRADE 4/AGE 9 ITEM	GRADE 8/AGE 13 BLOCK	GRADE 8/AGE 13 ITEM	GRADE 12/AGE 17 BLOCK	GRADE 12/AGE 17 ITEM
G005501	1.517	(0.046)	-0.031	(0.023)	0.191	(0.008)	--	--	--	--	G4	13
G005601	0.490	(0.022)	1.115	(0.054)	0.206	(0.007)	--	--	--	--	G4	14
G005701	0.949	(0.030)	0.087	(0.021)	0.252	(0.009)	--	--	--	--	G4	15
G005801	1.808	(0.037)	0.591	(0.030)	0.210	(0.006)	--	--	--	--	G4	16
G005901	1.476	(0.038)	0.407	(0.028)	0.275	(0.007)	--	--	--	--	G4	17
G006001	0.798	(0.027)	1.286	(0.051)	0.229	(0.006)	--	--	--	--	G4	18
G006101	2.311	(0.042)	0.688	(0.035)	0.250	(0.006)	--	--	--	--	G4	19
G006201	0.276	(0.025)	2.294	(0.213)	0.322	(0.007)	--	--	--	--	G4	20
G006301	0.344	(0.018)	0.356	(0.025)	0.185	(0.010)	--	--	--	--	G4	21
G006501	1.149	(0.027)	0.852	(0.031)	0.146	(0.006)	--	--	--	--	G4	22
G006601	0.079	(0.014)	6.425	(1.183)	0.192	(0.008)	--	--	--	--	G4	23
G006701	1.542	(0.048)	0.180	(0.028)	0.293	(0.008)	--	--	--	--	G4	24
G006801	0.985	(0.031)	1.746	(0.066)	0.175	(0.005)	--	--	--	--	G4	25
G006901	1.071	(0.027)	1.037	(0.037)	0.176	(0.006)	--	--	--	--	G4	26

Table F-8

1988 IRT Parameters, Mathematics Trend, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 9 BLOCK	AGE 9 ITEM
N250601	1.097	(0.078)	-0.231	(0.045)	0.212	(0.029)	M2	13
N250602	0.791	(0.053)	-0.584	(0.054)	0.189	(0.023)	M2	14
N250603	1.366	(0.071)	0.566	(0.056)	0.158	(0.013)	M2	15
N250701	0.743	(0.044)	-0.850	(0.059)	0.139	(0.022)	M1	7
N250702	1.001	(0.048)	0.841	(0.054)	0.117	(0.011)	M1	8
N250703	1.054	(0.064)	0.015	(0.033)	0.123	(0.016)	M1	9
N250901	0.599	[illegible]	-0.411	(0.040)	0.178	(0.019)	M2	17
N250902	1.101	[illegible]	1.181	(0.072)	0.157	(0.010)	M2	18
N250903	0.970	[illegible]	0.685	(0.050)	0.109	(0.012)	M2	19
N251401	0.654	(0.042)	-0.265	(0.038)	0.151	(0.021)	M2	16
N252001	1.244	(0.131)	2.670	(0.372)	0.196	(0.009)	M2	25
N252101	0.839	(0.060)	1.752	(0.143)	0.170	(0.012)	M1	25
N257201	1.233	(0.084)	-0.533	(0.055)	0.283	(0.020)	M1	11
N257801	0.588	(0.038)	-0.909	0.063)	0.240	(0.022)	M2	3
N258501	0.876	(0.066)	1.029	(0.092)	0.236	(0.012)	M3	19
N261401	0.509	(0.042)	-0.145	(0.037)	0.232	(0.020)	M2	12
N262201	0.441	(0.036)	-1.218	(0.105)	0.196	(0.024)	M1	10
N262401	0.594	(0.069)	0.928	(0.116)	0.300	(0.013)	M3	18
N262501	0.269	(0.031)	-0.688	(0.084)	0.227	(0.019)	M1	19
N262502	0.254	(0.062)	6.169	(1.519)	0.172	(0.008)	M1	20
N263401	0.888	(0.063)	-0.701	(0.063)	0.299	(0.022)	M2	4
N263402	1.010	(0.080)	-0.203	(0.043)	0.282	(0.018)	M2	5
N265401	1.582	(0.164)	2.224	(0.360)	0.340	(0.011)	M1	21
N266101	0.542	(0.052)	1.917	(0.192)	[illegible]	(0.011)	M1	22
N267001	0.597	(0.045)	-1.392	(0.110)	0.249	(0.026)	M3	16
N267601	1.268	(0.066)	-0.611	(0.049)	0.156	(0.020)	M1	3
N267602	1.103	(0.057)	-0.074	(0.031)	0.104	(0.014)	M1	18
N268201	1.248	(0.058)	1.026	(0.068)	0.201	(0.010)	M1	24
N269001	0.565	(0.087)	4.055	(0.634)	0.082	(0.007)	M2	26
N269101	0.540	(0.071)	2.970	(0.402)	0.238	(0.009)	M1	23
N270001	0.448	(0.030)	-0.727	(0.053)	0.000	(0.000)	M1	14
N270901	0.894	(0.037)	-2.165	(0.098)	0.000	(0.000)	M1	1
N271101	0.626	(0.034)	-0.305	(0.028)	0.000	(0.000)	M2	24
N272101	0.980	(0.096)	-0.533	(0.071)	0.286	(0.024)	M3	17
N272102	0.992	(0.062)	0.034	(0.039)	0.173	(0.018)	M1	15
N272301	0.946	(0.052)	-1.947	(0.123)	0.180	(0.040)	M2	1
N272801	0.576	(0.049)	-2.007	(0.176)	0.180	(0.036)	M3	15
N273501	0.744	(0.058)	-0.684	(0.068)	0.261	(0.026)	M2	6
N275401	0.985	(0.043)	-0.478	(0.033)	0.000	(0.000)	M2	7
N276001	0.879	(0.037)	-0.975	(0.049)	0.000	(0.000)	M2	21
N276002	0.778	(0.035)	1.507	(0.074)	0.000	(0.000)	M2	22
N276101	0.963	(0 040)	-0.758	(0.042)	0.000	(0.000)	M1	12
N276601	1.061	(0.062)	-1.010	(0.076)	0.170	(0.029)	M2	2
N276801	0.490	(0.045)	-3.763	(0.353)	0.000	(0.000)	M1	4
N276802	0.725	(0.038)	-1.591	(0.090)	0.000	(0.000)	M1	5
N276803	0.621	(0.035)	0.147	(0.027)	0.000	(0.000)	M1	6
N277401	1.026	(0.063)	-1.573	(0.114)	0.177	(0.038)	M1	2
N277501	0.842	(0.039)	-0.421	(0.031)	0.000	(0.000)	M2	8
N277601	1.438	(0.049)	-0.522	(0.037)	0.000	(0.000)	M2	9
N277602	1.267	(0.053)	0.172	(0.029)	0.000	(0.000)	M2	10
N277603	1.507	(0.063)	-0.011	(0.030)	0.000	(0.000)	M2	11
N284001	0.981	(0.050)	-0.383	(0.033)	0.000	(0.000)	M1	16
N284002	0.792	(0.037)	2.054	(0.103)	0.000	(0.000)	M1	17
N286101	0.814	(0.039)	-0.521	(0.035)	0.000	(0.000)	M1	13

Table F-9

1988 IRT Parameters, Mathematics Trend, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 13 BLOCK	ITEM
N250701	0.688	(0.035)	-2.717	(0.143)	0.106	(0.033)	M2	14
N250702	1.145	(0.051)	-0.797	(0.047)	0.102	(0.018)	M2	15
N250703	0.649	(0.031)	-2.110	(0.106)	0.110	(0.028)	M2	16
N250901	0.423	(0.029)	-2.565	(0.176)	0.152	(0.027)	M1	25
N250902	1.020	(0.049)	-0.349	(0.031)	0.075	(0.014)	M1	26
N250903	0.820	(0.039)	-1.510	(0.078)	0.096	(0.025)	M1	27
N252001	1.423	(0.064)	0.832	(0.062)	0.179	(0.010)	M2	40
N252101	0.933	(0.056)	0.623	(0.054)	0.240	(0.013)	M1	41
N252901	1.249	(0.072)	-0.036	(0.033)	0.109	(0.015)	M1	32
N253701	0.361	(0.031)	-0.504	(0.050)	0.271	(0.016)	M2	22
N254001	1.161	(0.084)	-0.479	(0.047)	0.118	(0.017)	M3	28
N254601	1.092	(0.054)	-1.553	(0.089)	0.284	(0.030)	M1	15
N254602	0.744	(0.045)	1.413	(0.095)	0.235	(0.009)	M1	46
N255701	1.317	(0.044)	1.268	(0.063)	0.139	(0.008)	M1	50
N256101	0.760	(0.033)	-1.0[illegible]6	(0.05[illegible])	0.000	(0.000)	M2	17
N256501	0.86[illegible]	(0.069)	0.581	(0.061)	0.318	(0.012)	M3	30
N256801	1.05[illegible]	(0.069)	0.841	(0.072)	0.312	(0.011)	M3	32
N257601	1.280	(0.055)	-0.538	(0.035)	0.000	(0.000)	M1	35
N258801	1.273	(0.055)	1.124	(0.076)	0.397	(0.010)	M1	38
N258802	1.619	(0.078)	0.484	(0.051)	0.254	(0.011)	M2	31
N258803	1.191	(0.044)	1.351	(0.068)	0.170	(0.007)	M2	41
N260101	1.299	(0.072)	0.415	(0.042)	0.160	(0.011)	M1	43
N261001	0.833	(0.049)	[illegible].011	(0.070)	0.219	(0.010)	M1	47
N261201	0.525	(0.052)	1.619	(0.166)	0.219	(0.012)	M2	38
N261301	0.700	(0.048)	0.768	(0.062)	0.113	(0.012)	M2	37
N261501	0.661	(0.056)	-0.545	(0.055)	0.141	(0.020)	M2	34
N261601	0.344	(0.043)	1.903	(0.239)	0.155	(0.012)	M2	36
N261801	0.679	(0.053)	0.044	(0.033)	0.223	(0.017)	M2	35
N262201	0.520	(0.037)	-1.789	(0.132)	0.361	(0.023)	M2	18
N262401	0.854	(0.054)	-0.556	(0.048)	0.323	(0.017)	M1	28
N262501	0.360	(0.033)	-0.237	(0.034)	0.348	(0.015)	M1	33
N262502	1.216	(0.068)	1.974	(0.151)	0.379	(0.008)	M1	34
N263101	0.527	(0.027)	-0.291	(0.024)	0.000	(0.000)	M1	39
N263401	0.675	(0.046)	-2.751	(0.196)	0.257	(0.040)	M2	12
N263402	0.635	(0.045)	-2.478	(0.181)	0.263	(0.036)	M2	13
N263501	1.389	(0.092)	0.187	(0.036)	0.115	(0.012)	M2	30
N264701	1.175	(0.056)	0.867	(0.059)	0.206	(0.010)	M2	33
N265201	0.810	(0.062)	-1.548	(0.127)	0.339	(0.032)	M1	36
N265202	0.643	(0.074)	-0.176	(0.041)	0.339	(0.018)	M1	30
N265901	0.933	(0.060)	0.930	(0.079)	0.333	(0.012)	M1	40
N265902	1.077	(0.073)	1.170	(0.103)	0.328	(0.011)	M3	31
N266101	0.849	(0.065)	-0.161	(0.033)	0.292	(0.014)	M3	27
N266801	0.559	(0.038)	-1.108	(0.080)	0.248	(0.021)	M1	31
N267201	0.776	(0.058)	-1.051	(0.087)	0.254	(0.026)	M1	23
N268001	1.012	(0.053)	0.382	(0.036)	0.152	(0.011)	M1	44
N269101	0.752	(0.046)	-0.384	(0.037)	0.213	(0.013)	M2	26
N269901	0.664	(0.049)	-0.274	(0.035)	0.288	(0.015)	M3	29
N270301	0.421	(0.031)	-1.596	(0.119)	0.126	(0.022)	M2	20
N270302	1.018	(0.047)	2.194	(0.118)	0.051	(0.005)	M2	21
N273901	1.786	(0.111)	0.258	(0.047)	0.184	(0.013)	M1	37
N274801	0.629	(0.051)	-0.192	(0.036)	0.269	(0.018)	M1	29
N275001	0.946	(0.040)	0.363	(0.027)	0.000	(0.000)	M1	42
N275301	0.372	(0.028)	-1.728	(0.132)	0.147	(0.022)	M3	25
N276801	0.433	(0.049)	-4.715	(0.542)	0.000	(0.000)	M1	17
N276802	0.493	(0.044)	-3.957	(0.359)	0.000	(0.000)	M1	18
N276803	0.435	(0.033)	-1.927	(0.148)	0.000	(0.000)	M1	19
N277401	0.778	(0.056)	-2.903	(0.220)	0.145	(0.042)	M2	8
N277601	0.856	(0.036)	-2.504	(0.113)	0.000	(0.000)	M1	20
N277602	0.624	(0.030)	-1.8[illegible]5	(0.095)	0.000	(0.000)	M1	21
N277603	0.617	(0.031)	-2.287	(0.117)	0.000	(0.000)	M1	22
N277901	0.591	(0.033)	-3.506	(0.199)	0.000	(0.000)	M2	9
N277902	0.688	(0.036)	-3.301	(0.178)	0.000	(0.000)	M2	10
N277903	0.573	(0.030)	-2.859	(0.154)	0.000	(0.000)	M2	11
N278901	1.559	(0.086)	0.415	(0.051)	0.212	(0.013)	M2	32

Table F-9 (continued)

1988 IRT Parameters, Mathematics Trend, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 13 BLOCK	AGE 13 ITEM
N278902	0.720	(0.051)	1.338	(0.107)	0.216	(0.012)	M2	29
N278903	1.338	(0.058)	1.066	(0.073)	0.169	(0.010)	M2	42
N278904	1.315	(0.057)	1.487	(0.097)	0.194	(0.010)	M1	49
N281401	0.728	(0.050)	1.711	(0.127)	0.106	(0.009)	M2	39
N281901	0.925	(0.040)	-2.181	(0.105)	0.146	(0.034)	M1	15
N282201	1.063	(0.058)	0.576	(0.051)	0.343	(0.011)	M2	28
N282202	0.936	(0.066)	-0.458	(0.045)	0.255	(0.017)	M3	26
N283101	1.578	(0.049)	1.554	(0.080)	0.148	(0.006)	M1	51
N286201	0.88[illegible]	(0.051)	-0.882	(0.061)	0.243	(0.021)	M1	24
N286301	1.189	(0.050)	0.660	(0.046)	0.205	(0.010)	M1	45
N286501	1.256	(0.042)	1.161	(0.058)	0.141	(0.008)	M1	48
N286502	1.671	(0.054)	1.171	(0.068)	0.160	(0.006)	M2	43
N286601	1.698	(0.059)	-0.194	(0.029)	0.000	(0.000)	M2	23
N286602	1.363	(0.051)	-0.247	(0.027)	0.000	(0.000)	M2	24
N286603	1.494	(0.050)	0.405	(0.030)	0.000	(0.000)	M2	25

Table F-10

1988 IRT Parameters, Mathematics Trend, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 17 BLOCK	AGE 17 ITEM
N251101	1.166	(0.035)	0.949	(0.041)	0.000	(0.000)	M1	
N251701	0.892	(0.056)	0.005	(0.029)	0.147	(0.015)	M2	41
N253901	1.647	(0.083)	0.011	(0.041)	0.259	(0.013)	M1	39
N253902	0.930	(0.057)	1.032	(0.0[illegible]	0.479	(0.011)	M1	40
N253903	1.168	(0.048)	0.915	(0.060)	0.322	(0.011)	M1	41
N253904	1.576	(0.062)	0.700	(0.058)	0.359	(0.011)	M1	42
N254001	0.923	(0.044)	-0.847	(0.050)	0.186	(0.020)	M2	21
N254301	1.035	(0.051)	0.084	(0.033)	0.258	(0.013)	M1	33
N254601	1.300	(0.049)	-1.815	(0.089)	0.237	(0.037)	M2	15
N254602	1.575	(0.070)	-0.024	(0.036)	0.2[illegible]1	(0.012)	M1	27
N255301	1.539	(0.052)	1.503	(0.086)	0.219	(0.009)	M2	46
N255501	0.808	(0.054)	0.668	(0.059)	0.232	(0.013)	M3	33
N255601	1.248	(0.059)	1.576	(0.107)	0.332	(0.011)	M2	45
N255701	1.451	(0.061)	-0.609	(0.045)	0.201	(0.018)	M1	32
N255801	0.679	(0.030)	1.668	(0.080)	0.000	(0.000)	M2	49
N256001	1.055	(0.068)	0.066	(0.027)	0.000	(0.000)	M3	34
N256101	1.003	(0.029)	-1.407	(0.051)	0.000	(0.000)	M1	15
N256801	1.300	(0.062)	-0.268	(0.038)	0.265	(0.015)	M1	36
N257101	0.579	(0.054)	1.853	(0.181)	0.254	(0.011)	M3	35
N258801	1.904	(0.110)	0.216	(0.048)	0.284	(0.012)	M2	38
N258802	1.728	(0.089)	-0.175	(0.042)	0.256	(0.014)	M1	26
N258803	0.992	(0.045)	0.250	(0.033)	0.222	(0.011)	M1	37
N258804	0.682	(0.037)	-1.852	(0.105)	0.254	(0.029)	M1	18
N259001	1.188	(0.045)	-0.218	(0.025)	0.000	(0.000)	M2	31
N259901	1.235	(0.066)	-0.225	(0.037)	0.289	(0.014)	M1	28
N260101	1.460	(0.055)	-0.973	(0.054)	0.195	(0.023)	M2	20
N260601	1.499	(0.035)	-1.136	(0.043)	0.000	(0.000)	M1	16
N260801	1.301	(0.044)	0.388	(0.030)	0.000	(0.000)	M2	43
N260901	2.210	(0.113)	0.086	(0.045)	0.157	(0.012)	M1	35
N261001	0.806	(0.045)	-0.734	(0.052)	0.216	(0.022)	M2	40
N261201	0.510	(0.031)	-1.518	(0.097)	0.215	(0.02[illegible])	M2	26
N261301	0.581	(0.031)	-1.299	(0.0[illegible]4)	0.153	(0.022)	M2	28
N261501	0.775	(0.036)	-2.237	(0.113)	0.166	(0.035)	M2	24
N261601	0.472	(0.032)	0.708	(0.055)	0.209	(0.012)	M2	27
N261801	0.589	(0.032)	-1.985	(0.114)	0.211	(0.029)	M2	25
N262301	0.517	(0.035)	-1.239	(0.089)	0.233	(0.023)	M2	17
N262401	0.920	(0.040)	-1.326	(0.068)	0.255	(0.025)	M1	17
N262501	0.878	(0.060)	0.217	(0.043)	0.477	(0.013)	M2	35
N262502	0.598	(0.045)	1.756	(0.141)	0.365	(0.010)	M2	36
N262601	0.756	(0.039)	0.432	(0.038)	0.233	(0.012)	M1	38
N263001	0.664	(0.027)	0.707	(0.035)	0.000	(0.000)	M1	43
N263101	0.754	(0.032)	-0.569	(0.033)	0.000	(0.000)	M2	37
N263201	0.973	(0.050)	-1.348	(0.080)	0.361	(0.026)	M2	18
N263202	0.659	(0.042)	-0.434	(0.041)	0.352	(0.016)	M2	19
N264301	0.800	(0.028)	0.988	(0.040)	0.000	(0.000)	M1	47
N264701	1.578	(0.082)	-0.033	(0.038)	0.216	(0.015)	M2	39
N266501	0.775	(0.060)	-0.326	(0.041)	0.244	(0.017)	M3	31
N268801	0.917	(0.039)	1.654	(0.085)	0.102	(0.009)	M2	48
N268901	1.691	(0.062)	0.639	(0.054)	0.184	(0.012)	M2	47
N269001	0.938	(0.046)	-0.398	(0.034)	0.169	(0.016)	M2	22
N270301	0.942	(0.036)	-1.403	(0.063)	0.140	(0.026)	M1	30
N270302	1.586	(0.059)	0.119	(0.031)	0.067	(0.009)	M1	31
N271301	1.374	(0.120)	0.185	(0.048)	0.261	(0.014)	M3	32
N278501	1.030	(0.033)	-0.759	(0.035)	0.000	(0.000)	M1	23
N278502	0.895	(0.032)	-0.559	(0.030)	0.000	(0.000)	M1	24
N278503	0.900	(0.030)	-0.831	(0.037)	0.000	(0.000)	M1	25
N278901	1.129	(0.056)	-0.229	(0.034)	0.232	(0.015)	M2	23
N278902	1.162	(0.065)	0.014	(0.036)	0.236	(0.016)	M2	42
N278903	1.921	(0.092)	0.365	(0.051)	0.227	(0.013)	M2	44
N278905	1.178	(0.046)	1.053	(0.063)	0.283	(0.010)	M1	44
N280401	0.550	(0.026)	-1.313	(0.067)	0.000	(0.000)	M2	30
N281401	0.685	(0.032)	-0.243	(0.027)	0.109	(0.015)	M2	29
N282801	1.806	(0.054)	1.310	(0.075)	0.206	(0.010)	M1	48
N286001	0.766	(0.035)	-0.944	(0.051)	0.169	(0.020)	M1	19

Table F-10 (continued)

1988 IRT Parameters, Mathematics Trend, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 17 BLOCK	AGE 17 ITEM
N286002	0.855	(0.032)	-1.658	(0.071)	0.121	(0.027)	M1	20
N286301	1.350	(0.071)	-0.450	(0.043)	0.221	(0.017)	M2	33
N286302	1.088	(0.056)	-0.439	(0.044)	0.289	(0.018)	M1	22
N286501	1.142	(0.049)	-0.847	(0.049)	0.149	(0.021)	M2	34
N286502	1.797	(0.097)	-0.123	(0.038)	0.131	(0.013)	M1	34
N287101	1.358	(0.060)	-0.382	(0.037)	0.202	(0.014)	M1	29
N287102	1.114	(0.050)	-0.556	(0.040)	0.172	(0.018)	M2	32
N287301	0.793	(0.030)	0.120	(0.022)	0.000	(0.000)	M1	45
N287302	0.984	(0.031)	1.226	(0.048)	0.000	(0.000)	M1	46

Table F-11

1988 IRT Parameters, Science Trend, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 9 BLOCK	AGE 9 ITEM
N400001	0.650	(0.056)	-1.173	(0.109)	0.237	(0.030)	S1	6
N400101	0.294	(0.069)	2.732	(0.643)	0.460	(0.016)	S1	15
N400102	0.455	(0.076)	1.909	(0.329)	0.424	(0.015)	S1	16
N400301	0.993	(0.113)	-0.130	(0.055)	0.340	(0.021)	S1	8
N400401	1.246	(0.092)	-1.214	(0.117)	0.417	(0.035)	S1	9
N400402	1.828	(0.126)	-0.733	(0.089)	0.280	(0.027)	S1	10
N400403	0.566	(0.063)	-1.941	(0.223)	0.422	(0.036)	S1	11
N400404	1.164	(0.098)	-0.651	(0.078)	0.322	(0.026)	S1	12
N400405	1.012	(0.095)	-0.748	(0.090)	0.390	(0.027)	S1	13
N400501	0.545	(0.063)	0.593	(0.083)	0.330	(0.018)	S1	14
N400601	0.648	(0.062)	-0.202	(0.044)	0.225	(0.021)	S1	17
N400701	0.741	(0.066)	0.070	(0.040)	0.202	(0.019)	S1	18
N400901	0.333	(0.049)	1.804	(0.268)	0.253	(0.015)	S1	19
N401001	0.542	(0.053)	0.729	(0.082)	0.210	(0.016)	S1	20
N401101	0.292	(0.048)	1.737	(0.288)	0.275	(0.016)	S1	21
N401201	0.851	(0.080)	2.036	(0.215)	0.243	(0.011)	S1	22
N401301	0.504	(0.060)	1.478	(0.183)	0.269	(0.014)	S1	23
N401501	0.260	(0.047)	0.249	(0.060)	0.347	(0.019)	S2	1
N401601	0.599	(0.058)	-1.492	(0.150)	0.207	(0.033)	S2	2
N401702	0.304	(0 059)	0.556	(0.118)	0.452	(0.018)	S2	4
N401703	0.299	(0.059)	1.035	(0.209)	0.443	(0.017)	S2	5
N401801	0.686	(0.109)	-0.035	(0.057)	0.447	(0.021)	S2	6
N401802	0.570	(0.082)	-0.962	(0.147)	0.432	(0.028)	S2	7
N401803	0.455	(0.075)	-0.279	(0.068)	0.440	(0.023)	S2	8
N401804	0.346	(0.068)	1.698	(0.338)	0.424	(0.016)	S2	9
N401901	0.469	(0.072)	1.855	(0.291)	0.318	(0.015)	S2	10
N402001	0.935	(0.091)	-1.045	(0.118)	0.381	(0.032)	S2	11
N402002	1.224	(0.106)	-1.036	(0.115)	0.386	(0.034)	S2	12
N402005	0.712	(0.103)	-0.510	(0.091)	0.411	(0.026)	S2	15
N402101	0.562	(0.061)	-0.332	(0.051)	0.206	(0.022)	S2	16
N402201	0.231	(0.039)	0.333	(0.067)	0.245	(0.019)	S2	17
N402401	0.253	(0.051)	2.764	(0.561)	0.235	(0.015)	S2	18
N402501	0.622	(0.090)	2.692	(0.407)	0.258	(0.011)	S2	19
N402602	0.401	(0.063)	-0.686	(0.117)	0.439	(0.022)	S2	21
N402701	0.453	(0.058)	1.980	(0.261)	0.199	(0.013)	S2	23
N402801	1.084	(0.083)	2.031	(0.189)	0.161	(0.009)	S2	24
N402901	0.373	(0.094)	4.734	(1.194)	0.185	(0.010)	S2	25
N403001	0.422	(0.062)	-5.043	(0.745)	0.228	(0.053)	S3	12
N403101	0.638	(0.062)	-3.422	(0.342)	0.232	(0.051)	S3	13
N403201	0.404	(0.048)	-3.042	(0.368)	0.212	(0.039)	S3	14
N403202	0.291	(0.038)	-1.195	(0.161)	0.238	(0.024)	S3	15
N403301	0.624	(0.056)	-1.079	(0.105)	0.218	(0.029)	S3	16
N403401	0.355	(0.048)	0.596	(0.089)	0.338	(0.017)	S3	17
N403501	0.563	(0 067)	0.257	(0.057)	0.400	(0.019)	S3	18
N403502	0.551	(0.059)	-1.918	(0.211)	0.404	(0.034)	S3	19
N403503	0.412	(0.060)	0.152	(0.054)	0.409	(0.020)	S3	20
N403601	0.811	(0.069)	0.534	(0.065)	0.254	(0.016)	S3	21
N403701	3.290	(0.390)	-0.287	(0.108)	0.312	(0.021)	S3	22
N403702	3.150	(0.247)	-0.496	(0.118)	0.374	(0.023)	S3	23
N403703	2.076	(0.204)	-0.326	(0.077)	0.302	(0.021)	S3	24
N403801	0.359	(0.057)	1.082	(0.180)	0.428	(0.017)	S3	25
N403803	0.497	(0.056)	-0.991	(0.119)	0.393	(0.026)	S3	27
N403804	0.484	(0.063)	-0.506	(0.080)	0.408	(0.023)	S3	28
N403901	0.653	(0.056)	-0.309	(0.046)	0.193	(0.023)	S3	29
N404001	0.203	(0.036)	1.764	(0.317)	0.223	(0.016)	S3	30
N404201	0.425	(0.050)	1.363	(0.165)	0.216	(0.015)	S3	31

Table F-12

1988 IRT Parameters, Science Trend, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 13 BLOCK	ITEM
N400201	0.464	(0.041)	-1.666	(0.151)	0.206	(0.029)	S1	16
N401201	0.544	(0.049)	0.415	(0.051)	0.249	(0.016)	S1	28
N404501	1.153	(0.055)	-2.021	(0.119)	0.164	(0.042)	S1	12
N404601	0.318	(0.039)	-0.641	(0.084)	0.228	(0.021)	S1	13
N404701	0.601	(0.043)	-1.538	(0.117)	0.194	(0.029)	S1	14
N404702	0.449	(0.041)	-0.140	(0.033)	0.201	(0.018)	S1	15
N404801	1.372	(0.085)	-1.624	(0.136)	0.422	(0.043)	S1	20
N404802	1.610	(0.140)	-0.514	(0.077)	0.360	(0.022)	S1	21
N404803	0.956	(0.078)	0.240	(0.049)	0.321	(0.016)	S1	22
N404901	0.691	(0.051)	-0.629	(0.057)	0.209	(0.022)	S1	17
N405001	0.349	(0.038)	0.200	(0.037)	0.214	(0.017)	S1	23
N405101	0.784	(0.052)	0.968	(0.077)	0.199	(0.012)	S1	24
N405201	0.515	(0.036)	-0.124	(0.033)	0.182	(0.019)	S1	25
N405301	0.623	(0.049)	1.251	(0.107)	0.199	(0.012)	S1	26
N405401	0.801	(0.053)	1.138	(0.087)	0.181	(0.011)	S1	27
N405501	0.628	(0.052)	-0.031	(0.035)	0.197	(0.019)	S1	29
N405601	0.233	(0.034)	1.041	(0.153)	0.198	(0.016)	S1	30
N405701	1.012	(0.067)	0.715	(0.065)	0.185	(0.013)	S1	31
N405801	0.493	(0.044)	1.324	(0.124)	0.166	(0.012)	S1	32
N405901	0.637	(0.049)	1.658	(0.137)	0.158	(0.011)	S1	33
N406001	0.455	(0.107)	4.846	(1.148)	0.174	(0.008)	S1	34
N406101	0.531	(0.120)	4.384	(1.008)	0.207	(0.008)	S1	35
N406201	0.360	(0.089)	5.620	(1.399)	0.099	(0.007)	S1	36
N406301	0.556	(0.052)	-1.563	(0.231)	0.430	(0.026)	S2	10
N406302	0.386	(0.051)	-0.408	(0.069)	0.428	(0.021)	S2	11
N406303	0.606	(0.063)	1.470	(0.166)	0.392	(0.013)	S2	12
N406304	0.471	(0.066)	1.354	(0.200)	0.419	(0.015)	S2	13
N406401	0.504	(0.066)	-0.157	(0.050)	0.461	(0.020)	S2	14
N406402	0.861	(0.090)	0.303	(0.062)	0.405	(0.018)	S2	15
N406403	0.753	(0.074)	-1.328	(0.142)	0.419	(0.031)	S2	16
N406404	0.910	(0.111)	-0.305	(0.067)	0.457	(0.022)	S2	17
N406405	0.628	(0.066)	-0.528	(0.075)	0.402	(0.025)	S2	18
N406501	0.495	(0.043)	0.62[illegible]	(0.064)	0.170	(0.016)	S2	19
N406601	0.491	(0.044)	-0.855	(0.082)	0.175	(0.023)	S2	20
N406701	0.576	(0.049)	0.093	(0.034)	0.240	(0.016)	S2	21
N406801	1.128	(0.074)	-1.417	(0.114)	0.386	(0.036)	S2	22
N406802	0.342	(0.047)	0.687	(0.104)	0.445	(0.015)	S2	23
N406803	0.816	(0.074)	-0.660	(0.074)	0.382	(0.022)	S2	24
N406804	1.057	(0.073)	-1.014	(0.086)	0.371	(0.027)	S2	25
N406805	1.037	(0.097)	1.523	(0.181)	0.550	(0.011)	S2	26
N406806	0.440	(0.053)	0.226	(0.050)	0.423	(0.017)	S2	27
N406901	0.613	(0.052)	0.019	(0.034)	0.231	(0.017)	S2	28
N407001	0.263	(0.035)	0.158	(0.038)	0.182	(0.019)	S2	29
N407101	0.817	(0.055)	2.218	(0.168)	0.126	(0.009)	S2	30
N407201	0.470	(0.041)	0.437	(0.050)	0.207	(0.015)	S2	31
N407301	0.319	(0.039)	1.672	(0.208)	0.234	(0.013)	S2	32
N407302	0.346	(0.046)	1.817	(0.245)	0.270	(0.014)	S2	33
N407601	0.453	(0.044)	1.743	(0.173)	0.180	(0.012)	S2	35
N407701	0.564	(0.044)	1.273	(0.107)	0.144	(0.012)	S2	37
N407801	0.666	(0.055)	2.158	(0.189)	0.199	(0.010)	S2	38
N407901	0.383	(0.037)	0.849	(0.089)	0.168	(0.015)	S2	39
N408001	0.846	(0.050)	1.268	(0.087)	0.176	(0.010)	S2	34
H408201	0.567	(0.070)	3.245	(0.415)	0.206	(0.009)	S2	40
N406301	0.788	(0.061)	0.970	(0.089)	0.298	(0.013)	S3	10
N408302	0.708	(0.065)	-1.545	(0.152)	0.408	(0.033)	S3	11
N408303	0.647	(0.060)	-1.697	(0.163)	0.415	(0.031)	S3	12
N408304	0.971	(0.079)	-1.384	(0.129)	0.414	(0.034)	S3	13
N408401	0.240	(0.032)	-1.476	(0.199)	0.223	(0.022)	S3	14
N408501	0.733	(0.056)	-0.896	(0.077)	0.205	(0.025)	S3	15
N408502	0.390	(0.040)	1.337	(0.140)	0.154	(0.013)	S3	16
N408601	0.388	(0.035)	-1.071	(0.102)	0.153	(0.022)	S3	17
N408701	0.346	(0.038)	-0.101	(0.031)	0.212	(0.018)	S3	18
N408801	0.174	(0.030)	0.655	(0.117)	0.234	(0.017)	S3	19
N408901	0.743	(0.079)	0.274	(0.0[illegible]5)	0.445	(0.015)	S3	20

Table F-12 (continued)

1988 IRT Parameters, Science Trend, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 13 BLOCK	AGE 13 ITEM
N408902	0.888	(0.069)	-1.740	(0.149)	0.410	(0.038)	S3	21
N408903	0.656	(0.066)	0.434	(0.062)	0.404	(0.015)	S3	22
N408904	0.540	(0.060)	0.877	(0.107)	0.411	(0.014)	S3	23
N409001	0.599	(0.045)	-0.364	(0.040)	0.163	(0.019)	S3	24
N409101	0.635	(0.045)	-1.494	(0.113)	0.239	(0.029)	S3	25
N409102	0.556	(0.047)	0.178	(0.036)	0.229	(0.016)	S3	26
N409103	0.518	(0.059)	2.017	(0.235)	0.306	(0.011)	S3	27
N409201	0.292	(0.039)	0.393	(0.061)	0.261	(0.017)	S3	28
N409301	0.706	(0.056)	-0.145	(0.035)	0.165	(0.018)	S3	29
N409501	0.607	(0.052)	2.148	(0.191)	0.134	(0.009)	S3	33
N409601	0.708	(0.061)	1.717	(0.162)	0.290	(0.011)	S3	34
N409701	0.633	(0.060)	2.485	(0.248)	0.165	(0.009)	S3	35

Table F-13

1988 IRT Parameters, Science Trend, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 17 BLOCK	AGE 17 ITEM
N400201	0.543	(0.116)	-1.669	(0.370)	0.196	(0.049)	S1	12
N401201	0.613	(0.124)	-0.226	(0.097)	0.229	(0.040)	S1	30
N404601	0.542	(0.118)	-0.565	(0.150)	0.197	(0.043)	S1	13
N405001	0.462	(0.098)	-0.305	(0.102)	0.198	(0.039)	S1	29
N405101	0.595	(0.111)	0.272	(0.103)	0.235	(0.035)	S3	14
N405201	0.444	(0.095)	-0.703	(0.168)	0.152	(0.040)	S1	31
N405401	0.619	(0.104)	0.631	(0.138)	0.145	(0.031)	S3	19
N405501	0.584	(0.121)	-0.295	(0.108)	0.196	(0.041)	S3	21
N406001	0.471	(0.115)	2.129	(0.536)	0.197	(0.026)	S1	33
N406101	0.494	(0.142)	2.885	(0.854)	0.214	(0.025)	S1	35
N406201	0.658	(0.130)	2.184	(0.457)	0.116	(0.022)	S1	37
N406301	0.334	(0.091)	-1.322	(0.371)	0.410	(0.047)	S1	21
N406302	0.420	(0.105)	-0.246	(0.118)	0.401	(0.042)	S1	22
N406303	0.506	(0.129)	0.383	(0.147)	0.397	(0.037)	S1	23
N406304	0.511	(0.138)	-0.276	(0.133)	0.395	(0.044)	S1	24
N406401	0.632	(0.158)	-0.678	(0.205)	0.395	(0.049)	S2	10
N406402	0.676	(0.189)	-0.075	(0.124)	0.391	(0.043)	S2	11
N406403	0.815	(0.189)	-1.522	(0.388)	0.395	(0.059)	S2	12
N406404	0.721	(0.182)	-1.204	(0.333)	0.393	(0.056)	S2	13
N406405	0.637	(0.171)	-0.963	(0.292)	0.397	(0.054)	S2	14
N406601	0.547	(0.115)	-0.915	(0.210)	0.151	(0.042)	S1	28
N406801	0.672	(0.157)	-1.921	(0.471)	0.396	(0.059)	S2	16
N406802	0.452	(0.121)	1.281	(0.367)	0.403	(0.033)	S2	17
N406803	0.575	(0.136)	-1.248	(0.314)	0.396	(0.051)	S2	18
N406804	0.709	(0.147)	-1.539	(0.344)	0.391	(0.056)	S2	19
N406805	0.458	(0.117)	0.473	(0.164)	0.408	(0.037)	S2	20
N406806	0.396	(0.105)	0.270	(0.128)	0.406	(0.040)	S2	21
N406901	0.500	(0.109)	-0.532	(0.142)	0.196	(0.042)	S2	27
N407001	0.333	(0.079)	-0.920	(0.232)	0.155	(0.042)	S2	33
N407101	0.614	(0.126)	1.878	(0.410)	0.150	(0.026)	S2	38
N407201	0.500	(0.106)	0.120	(0.084)	0.153	(0.035)	S2	32
N407301	0.346	(0.083)	0.510	(0.147)	0.204	(0.036)	S2	36
N407302	0.445	(0.110)	0.917	(0.249)	0.246	(0.035)	S2	37
N407401	0.652	(0.143)	0.348	(0.136)	0.375	(0.036)	S2	28
N407403	0.581	(0.151)	-0.258	(0.137)	0.393	(0.046)	S2	30
N407404	0.714	(0.166)	-1.370	(0.349)	0.395	(0.057)	S2	31
N407701	0.450	(0.098)	0.898	(0.214)	0.152	(0.032)	S2	35
N408101	0.625	(0.124)	1.626	(0.344)	0.142	(0.026)	S1	38
N408301	0.834	(0.186)	-0.241	(0.125)	0.381	(0.041)	S3	10
N408302	0.457	(0.119)	-1.685	(0.456)	0.401	(0.056)	S3	11
N408303	0.543	(0.122)	-2.012	(0.470)	0.398	(0.057)	S3	12
N408304	0.640	(0.162)	-1.585	(0.426)	0.396	(0.058)	S3	13
N408601	0.426	(0.091)	-1.329	(0.295)	0.164	(0.043)	S1	19
N408801	0.505	(0.101)	-0.340	(0.105)	0.198	(0.039)	S3	24
N408901	0.769	(0.168)	-1.203	(0.292)	0.393	(0.053)	S3	15
N408902	0.836	(0.165)	-1.022	(0.422)	0.395	(0.062)	S3	16
N408903	0.563	(0.127)	-0.172	(0.112)	0.392	(0.041)	S3	17
N408904	0.586	(0.135)	-0.374	(0.135)	0.398	(0.043)	S3	18
N409301	0.625	(0.122)	-1.324	(0.274)	0.149	(0.044)	S1	20
N409501	0.714	(0.129)	1.100	(0.225)	0.133	(0.028)	S1	34
N409901	0.867	(0.168)	-0.931	(0.209)	0.191	(0.046)	S1	18
N410003	0.509	(0.121)	-1.988	(0.486)	0.400	(0.055)	S1	16
N410004	0.499	(0.129)	-1.225	(0.334)	0.401	(0.050)	S1	17
N410101	0.626	(0.158)	-0.700	(0.207)	0.394	(0.046)	S1	25
N410102	0.433	(0.113)	-0.401	(0.144)	0.404	(0.043)	S1	26
N410103	0.566	(0.139)	-1.408	(0.365)	0.396	(0.053)	S1	27
N410201	0.491	(0.119)	1.890	(0.476)	0.1[illegible]9	(0.030)	S1	32
N410401	0.396	(0.093)	0.086	(0.088)	0.244	(0.039)	S2	15
N410501	0.415	(0.088)	-0.420	(0.118)	0.150	(0.039)	S2	22
N410601	1.057	(0.208)	2.077	(0.507)	0.229	(0.025)	S2	23
N410602	0.430	(0.122)	-2.476	(0.714)	0.405	(0.058)	S2	24
N410603	0.768	(0.170)	1.333	(0.336)	0.338	(0.029)	S2	25
N410604	0.414	(0.110)	-2.139	(0.577)	0.405	(0.055)	S2	26
N410701	0.542	(0.120)	0.833	(0.209)	0.201	(0.033)	S2	34

Table F-13 (continued)

1988 IRT Parameters, Science Trend, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	AGE 17 BLOCK	AGE 17 ITEM
N410801	0.542	(0.124)	1.554	(0.376)	0.193	(0.030)	S2	39
N410901	0.707	(0.134)	1.777	(0.367)	0.155	(0.024)	S2	40
N411001	0.545	(0.145)	2.730	(0.751)	0.193	(0.024)	S2	41
N411101	0.508	(0.096)	0.255	(0.093)	0.150	(0.035)	S3	22
N411201	0.566	(0.105)	0.490	(0.126)	0.195	(0.033)	S3	23
N411301	0.469	(0.139)	3.814	(1.163)	0.120	(0.023)	S3	20
N411401	0.846	(0.151)	0.534	(0.137)	0.152	(0.030)	S3	25
N411501	0.860	(0.125)	1.749	(0.300)	0.179	(0.024)	S3	26
N411502	0.619	(0.131)	-1.037	(0.240)	0.237	(0.048)	S3	27
N411601	0.609	(0.108)	1.227	(0.244)	0.184	(0.030)	S3	28
N411701	0.745	(0.119)	1.395	(0.256)	0.169	(0.027)	S3	29
N411801	1.069	(0.175)	0.650	(0.161)	0.167	(0.031)	S3	30
N411901	0.762	(0.122)	1.429	(0.261)	0.142	(0.025)	S3	31
N412001	0.572	(0.119)	2.048	(0.453)	0.187	(0.029)	S3	32

APPENDIX G

NAEP Reporting Subgroups

Composite and Derived Common Background Variables

Subject-specific Composite and Derived Reporting Variables

Appendix G

NAEP REPORTING SUBGROUPS

DSEX (Gender)

The variable SEX on the student file is the gender of the student being assessed, as taken from school records. For a few students, data for this variable was missing and was imputed by ETS after the assessment. The resulting variable DSEX on the student file contains a value for every student and is used for gender comparisons among students.

RACE (Observed Race/Ethnicity)

The variable RACE on the student file is the race/ethnicity of the student being assessed, as observed and recorded by the exercise administrator. Observed race/ethnicity was used in NAEP assessments before 1984. This variable is used for race/ethnicity subgroup comparisons to assessments conducted before 1984.

DRACE (Imputed Race/Ethnicity)

The variable DRACE on the student file is an imputed definition of race/ethnicity, derived from up to three sources of information. This variable is used for race/ethnicity subgroup comparisons within the 1988 assessment and among the 1988, 1986, and 1984[1] assessments.

Two common background items were used in the determination of race/ethnicity:

<u>Common Background Item Number 2:</u>

2. If you are Hispanic, what is your Hispanic background?

- ⬭ I am not Hispanic.
- ⬭ Mexican, Mexican American, or Chicano
- ⬭ Puerto Rican
- ⬭ Cuban
- ⬭ Other Spanish or Hispanic background

---

[1] The 1984 variable ETHNIC contains the same race/ethnicity categories as the 1988 variable DRACE, but the imputation procedure used was slightly different.

Students who responded to item number 2 by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the item, or provided information that was illegible or could not be classified, responses to item number 1 were examined in an effort to determine race/ethnicity. Item number 1 read as follows[2]:

Common Background Item Number 1:

1. Which best describes you?

- ⬭ White
- ⬭ Black
- ⬭ Hispanic
- ⬭ Asian or Pacific Islander
- ⬭ American Indian or Alaskan Native
- ⬭ Other (What?) ________

Students' race/ethnicity was then assigned to correspond with their selection. For students who filled in the sixth oval ("Other"), provided illegible information or information that could not be classified, or did not respond at all, observed race/ethnicity (RACE), if provided by the exercise administrator, was used.

Imputed race/ethnicity could not be determined for students who did not respond to background items 1 or 2 and for whom an observed race/ethnicity was not provided.

STOC (Size and type of community)

NAEP assigned each participating school to one of seven size and type of community categories designed to provide information about the communities in which the schools are located.

The STOC reporting categories consist of three "extreme" types of communities and four "residual" community sizes. Schools were placed into STOC categories on the basis of information about the type of community, the size of its population (as of the 1980 Census), and an occupational profile of residents provided by school principals before the assessment. The principals completed estimates of the percentage of students whose parents fit into each of six occupational categories.

Two versions of STOC were created: a weighted version used for cross-sectional analyses and an unweighted version, which was the version used historically. For both versions, schools were ranked in order based on principals' responses about the type of community, size of its population, and occupational profile of the students' parents. For the weighted version of STOC, schools were assigned to the extreme STOC categories (1, 2, and 3) so as

[2]In the bridge to 1984, the race/ethnicity response choices for this item were in a different order.

7

to give the result that 10 percent of sampled students (weighted) were enrolled in schools in each such category. For the unweighted version, the extreme 10 percent of the schools (unweighted) were assigned to each of the extreme STOC categories. For both versions of STOC, the remaining schools were classified according to one of the four residual STOC categories. The extreme STOC categories are as follows:

STOC 1 - Extreme Rural: This category was used for schools in rural areas where the principal reported that a high proportion of students' parents were farmers or farm workers and a low proportion were professional, managerial, or factory workers. At least some of the students in each of these schools were from open country or places with a population of fewer than 2,500 and no students were from a place with population in excess of 10,000.

STOC 2 - Low Metro: The low metro STOC category was used for schools in areas where the principal reported that a high proportion of students' parents were either not regularly employed or on welfare and a low proportion were employed in professional or managerial positions. The schools in STOC 2 were located in cities, or the urbanized area of cities, with a population greater than 200,000.

STOC 3 - High Metro: High metro schools were located in areas where the principal reported that a high proportion of students' parents were employed in professional or managerial positions and a low proportion were factory or farm workers, not regularly employed, or on welfare. STOC 3 schools were located in cities or the urbanized area of cities with populations greater than 200,000.

Schools that did not fall into STOC 1. 2, or 3 were classified according to four "residual" STOC categories, depending on the size of the community in which they were located. The four residual STOC reporting categories are as follows:

STOC 4 - Main Big City: STOC 4 schools were located within the limits of cities with populations greater than 200,000 but not classified as high or low metro.

STOC 5 - Urban Fringe: The schools assigned to STOC 5 were located in the urbanized area outside the limits of cities with populations greater than 200,000, but not classified as low or high metro.

STOC 6 - Medium City: STOC 6 schools were located in cities with populations between 25,000 and 200,000 that were not classified as belonging to the urbanized area of a big city.

STOC 7 - Small Place: The schools assigned to STOC 7 were located in communities with populations of fewer than 25,000. These communities were not located in the urbanized areas of big cities and were not classified as extreme rural.

PARED (Parental education)

The variable PARED on the student file is derived from responses to two common background items, B003501 and B003601.[3] Students were asked to indicate the extent of their father's education (item B003601) by choosing one of the following:

- He did not finish high school.
- He graduated from high school.
- He had some education after high school.
- He graduated from college.
- I don't know.

Students were asked to provide the same information about the extent of their mother's education (item B003501) by choosing one of the following:

- She did not finish high school.
- She graduated from high school.
- She had some education after high school.
- She graduated from college.
- I don't know.

The information was combined into one parental education reporting category as follows:

If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data. For students who did not know the level of education for both parents or did not know the level of education for one parent and did not respond for the other, the parental education level was classified as unknown. If the student did not respond for both parents, the student was recorded as having provided no response.

[3]In the bridge to 1984, the items used were B000601 and B000701. Instead of "had some education after high school," the third response choice read, "went to another school after graduating from high school."

REGION (Region of the country)

In addition to overall responses, NAEP computed data for four geographical regions in the United States. States were assigned to regions as follows:

Northeast	Southeast	Central	West
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia*	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia*			Utah
			Washington
			Wyoming

*That part of Virginia that is included in the Washington, DC, metropolitan statistical area is included in the Northeast region; the remainder of the state is included in the Southeast region.

DGRADE, MODGRD (Grade in school)

To enhance the usefulness of the data, in 1984 NAEP began sampling students by grade as well as by age. The ages sampled in assessments since 1984—9, 13, and 17—match the ages sampled in earlier assessments. However, some of the modal grades (the grade attended by most students of a particular age) for the ages sampled have varied in the last three assessments because of changes in how student age was determined and changes in the times of the year that students were tested.

In the 1988 main assessment, the respective modal grades for ages 9, 13, and 17 are 4, 8, and 12. Student age for all three cohorts was determined on a calendar-year basis; all students were tested at the same times of the year. The 1988 bridge samples, by definition designed to match previous assessment characteristics, sampled varying student cohorts.

The 1988 main sample included many students in each cohort who were both age-eligible (age 9, 13, or 17) and grade-eligible (attending respectively grade 4, 8, or 12). However, because NAEP collected data by grade <u>or</u> age, each cohort also included students who were age-eligible but not in the modal grade, and students who were grade-eligible but not of the modal age (the age of most students attending the particular grade).

For each 1988 sample, results for students in a particular grade can be selected using (1) the variable DGRADE, the student's actual grade at time of testing, on the student file, or (2) the student file variable MODGRD (setting MODGRD to a value of 2 will select those students who are in the modal grade).

DAGE, MODAGE (Student age)

For the 1988 main sample, student age for all three grade/age cohorts was calculated as of December 31, 1987. Because NAEP collected data by grade or age, each main sample student cohort includes students who were both age-eligible and grade-eligible, students who were age- eligible but not in the modal grade, and students who were grade-eligible but not of the modal age. The modal ages for each grade/age cohort in the main sample were defined by the following birth dates, based on a calendar year:

Age 9: born between January 1 and December 31, 1978
Age 13: born between January 1 and December 31, 1974
Age 17: born between January 1 and December 31, 1970

For the bridge to 1982, samples of age 13 and age 17 were selected based on the following birth dates:

Age 13: born between January 1 and December 31, 1974
Age 17: born between October 1, 1970 and September 30, 1971

The sample of grade 4/age 9, grade 8/age 13, and grade 11/age 17 students for the bridge to 1984 were selected according to the following birth dates:

Age 9: born between January 1 and December 31, 1978
Age 13: born between January 1 and December 31, 1974
Age 17: born between October 1, 1970 and September 30, 1971

The samples of ages 9 and 13 and grade 11/age 17 students for the bridge to 1986 were selected according to the following birth dates:

Age 9: born between January 1 and December 31, 1978
Age 13: born between January 1 and December 31, 1974
Age 17: born between October 1, 1970 and September 30, 1971

For all 1988 samples, results for students in a particular age can be selected using (1) the variable DAGE, the student's actual age at time of testing, on the student file or (2) the student file variable MODAGE (setting MODAGE to a value of 2 will select those students who are of the modal age).

COMPOSITE AND DERIVED COMMON BACKGROUND VARIABLES

Several NAEP variables are formed from the systematic combination of response values for one or more common background items (items given to every student concerning subjects such as materials in the home, languages spoken, hours spent watching television, and after-school activities).

The derived background variables maximize use of the data, incorporate a larger segment of the population, and save analysis costs by grouping items that measure similar characteristics into one variable. Some of these variables were used in the NAEP subject area reports. The derivation of each of these variables is explained in the following sections.

HOMEENV (Home Environment—Articles in the Home)

For all three age classes, the variable HOMEENV was created from the responses to background items B000901 through B000905 concerning articles found in the student's home (newspaper, dictionary, encyclopedia, more than 25 books, and magazines). The values for this variable were derived as follows.

1	0-3 ARTICLES	The student responded to at least three items and answered YES to three or fewer.
2	4 ARTICLES	The student answered YES to four items.
3	5 ARTICLES	The student answered YES to five items.
8	NO RESPONSE	The student answered fewer than three items.

HOMEEN2 (Home Environment—Articles [of 4] in the Home)

For all three age classes, the variable HOMEEN2 was created from the responses to background items B000901, B000903, B000904, and B000905 concerning articles found in the student's home (newspaper, encyclopedia, more than 25 books, and magazines). The values for this variable were derived as follows:

1	0-2 ARTICLES	The student responded to at least two items and answered YES to two or fewer.
2	3 ARTICLES	The student answered YES to three items.
3	4 ARTICLES	The student answered YES to four items.
8	NO RESPONSE	The student answered fewer than two items.

NCOMP (Number of Computer Science Courses Taken)

For age class 17, NCOMP was created from responses to items B005312 and B005313 concerning the student's coursework in computer science. The values for NCOMP were derived as follows:

1	0	The student answered HAVE NOT to both courses.
2	1	The student answered HAVE to one course.
3	2	The student answered YES to both courses.
8	NO RESPONSE	The student did not respond to one or both items.
9	MULT. & OUT-OF-RANGE	The student filled in more than one oval for both items.

NMATH (Highest Level of Mathematics Courses Taken)

For age class 17, NMATH was created from responses to items B005301 through B005307 concerning the student's coursework in mathematics. The values for NMATH were derived as follows:

1	GEN. MATH OR PRE-ALGEBRA	The student answered HAVE NOT to all items or HAVE to B005301 or B005302 and HAVE NOT to all others.
2	ALGEBRA	The student answered HAVE to B005303 and HAVE NOT to B005304, B005305, B005306, and B005307.
3	GEOMETRY	The student answered HAVE to both B005303 and B005305 and HAVE NOT to B005304, B005306, and B005307.
4	ALGEBRA 2	The student answered HAVE to B00530 or B005306 but HAVE NOT to B005307.
5	CALCULUS	The student answered HAVE to B005307.
6	SOMETHING ELSE	Any other response combination
8	NO RESPONSE	The student did not respond to any item.

NSCI (Highest Level of Science Courses Taken)

For age class 17, NSCI was created from responses to items B005308 through B005311, which concerned the student's coursework in science. The values for NSCI were derived as follows:

1	NO BIOLOGY	The student answered HAVE NOT to all items or HAVE to B005308 and other than HAVE to all other items.

2	BIOLOGY	The student answered HAVE to B005309 and other than HAVE to both B005310 and B005311.
3	CHEMISTRY	The student answered HAVE to both B005309 and B005310 and other than HAVE to B005311.
4	PHYSICS	The student answered HAVE to B005309, B005310, and B005311.
5	SOMETHING ELSE	Any other response combination
8	NO RESPONSE	The student answered none of the items.

NSIBS (Number of Siblings)

For all age classes, NSIBS was created from responses to item B005801 (How many brothers and sisters do you have?). The values for NSIBS were derived as follows:

1	NONE	The student answered NONE.
2	1	The student answered ONE.
3	2	The student answered TWO.
4	3	The student answered THREE.
5	4 OR MORE	The student answered FOUR, FIVE, or SIX OR MORE.
8	NO RESPONSE	The student did not respond.
9	MULT.	The student filled in more than one oval.

PARWK4, PARWK5 (Economic Support—Which Parents Work)

For age class 9, PARWK4 was created from responses to items B005901 and B006101, which asked if the student's mother (or stepmother) and father (or stepfather) worked for pay. The values for PARWK4 were derived as follows.

1	BOTH MOTHER & FATHER	The student answered YES to both B005901 and B006101.
2	FATHER ONLY	The student answered NO or DON'T LIVE WITH to B005901 and YES to B006101.
3	MOTHER ONLY	The student answered NO or DON'T LIVE WITH to B006101 and YES to B005901.
4	SOMETHING ELSE	Any other combination of responses
8	NO RESPONSE	The student did not respond to one or both items.

9 MULT. The student filled in more than one oval for both items

For age classes 13 and 17, PARWK5 was created from responses to items B006001 and B006201, which asked if the student's mother (or stepmother) and father or (stepfather) worked part-time or full-time for pay. The values for PARWK5 were derived as follows:

1 BOTH FULL TIME The student answered YES, FULL-TIME to both B006001 and B006201.

2 BOTH WORK SOME The student answered YES, FULL-TIME or YES, PART-TIME to BOO 201 and YES, FULL-TIME or YES, PART-TIME to B006001.

3 FATHER ONLY The student answered YES, FULL-TIME or YES, PART-TIME to B006201 and NO or DON'T LIVE WITH to B006001.

4 MOTHER ONLY The student answered YES, FULL-TIME or YES, PART-TIME to B006001 and NO or DON'T LIVE WITH to B006201.

5 SOMETHING ELSE Any other combination of responses

8 NO RESPONSE The student did not respond to one or both items.

9 MULT. The student filled in more than one oval for both items.

RACLANG (By Race, Language Other than English is Spoken by Student in Home)

For all age classes, the variable RACLANG was created from responses to item B003401 (Do you speak a language other than English in your home?) and the imputed race/ethnicity variable DRACE. The values for RACLANG were derived as follows:

1 HISPANIC The student answered YES to B003401, and DRACE was HISPANIC.

2 ASIAN The student answered YES to B003401, and DRACE was ASIAN.

3 OTHER RACE The student answered YES to B003401, and DRACE was anything other than HISPANIC or ASIAN.

4 NO The student answered NO to B003401.

8 NO RESPONSE The student did not respond to B003401.

9 MULT. The student filled in more than one oval for B003401.

RACOFTN (By Race, How Often Language Other than English is Spoken by Others in Home)

For age class 9, the variable RACOFTN was created from item B003201 (How often do the people in your home speak a language other than English?) and the imputed race/ethnicity variable DRACE. The values for RACOFTN for age class 9 were derived as follows:

1 HISPANIC	The student answered SOMETIMES or ALWAYS to B003201, and DRACE was HISPANIC.
2 ASIAN	The student answered SOMETIMES or ALWAYS to B003201, and DRACE was ASIAN.
3 OTHER RACE	The student answered SOMETIMES or ALWAYS to B003201, and DRACE was not HISPANIC or ASIAN.
4 NEVER	The student answered NEVER to B003201.
8 NO RESPONSE	The student did not respond to B003201.
9 MULT.	The student filled in more than one oval for B003201.

For age classes 13 and 17, the variable RACOFTN was created from item B003301 (How often do the people in your home speak a language other than English?) and the imputed race/ethnicity variable DRACE. The values for RACOFTN for these age classes were derived as follows:

1 HISPANIC	The student answered ABOUT HALF THE TIME, MOST OF THE TIME, or ALWAYS to B003301, and DRACE was HISPANIC.
2 ASIAN	The student answered ABOUT HALF THE TIME, MOST OF THE TIME, or ALWAYS to B003301, and DRACE was ASIAN.
3 OTHER RACE	The student answered ABOUT HALF THE TIME, MOST C THE TIME, or ALWAYS to B003301, and DRACE was not HISPANIC r ASIAN.
NEVER	The student answered NEVER or OCCASIONALLY to B003301.
8 NO RESPONSE	The student did not respond to B003301.
9 MULT.	The student filled in more than one oval for B003301.

SINGLEP (How Many Parents Live at Home)

For all three age classes, SINGLEP was created from items B005601 and B005701, which asked whether the student's mother (or stepmother) and father (or stepfather) lived at home with the student. The values for SINGLEP were derived as follows:

1 2 PARENTS AT HOME	The student answered YES to both items.

2	1 PARENT AT HOME	The student answered YES to B005601 and NO to B005701, or YES to B005701 and NO to B005601.
3	NEITHER AT HOME	The student answered NO to both items.
8	NO RESPONSE	The student did not respond to one or both items.
9	MULT.	The student filled in more than one oval for one or both items.

SINGPB (Which Parents Live at Home)

For all age classes, SINGPB was created from items B005601 and B005701 concerning whether the student's mother (or stepmother) and father (or stepfather) live at home with the student. The values for SINGPB were derived as follows:

1	MOTHER AND FATHER	The student answered YES to both items.
2	FATHER ONLY	The student answered YES to B005701 and NO to B005601.
3	MOTHER ONLY	The student answered YES to B005601 and NO to B005701.
4	NEITHER	The student answered NO to both items.
8	NO RESPONSE	The student did not respond to one or both items.
9	MULT.	The student filled in more than one oval for one or both items.

SINGFEM (Single Female Head of Household—Working)

SINGFEM was created from the variables SINGPB and PARWK5 (PARWK4 for age class 9) concerning, respectively, which parents live at home and which parents work. The values for SINGFEM were derived as follows:

1	YES	SINGPB was 3 and PARWK4 was 3 (age class 9); SINGPB was 3 and PARWK5 was 4 (other age classes).
2	NO	SINGPB was 3 and PARWK4 was 4 (age class 9); SINGPB was 3 and PARWK5 was 5 (other age classes).
3	SOMETHING ELSE	Any other response combination.
8	NO RESPONSE	No value was derived for one or both variables because of nonresponse.
9	MULT.	No value was derived for one or both variables because of multiple response.

TVWATCH (Amount of Television Viewing Each Day)

For all three age classes, the variable TVWATCH was created from item B001801 (How much television do you usually watch each day?). The values for TVWATCH were derived as follows:

1	0-2 HOURS	The student answered NONE, ONE HOUR OR LESS, or TWO HOURS to B001801.
2	3-5 HOURS	The student answered THREE HOURS, FOUR HOURS, or FIVE HOURS to B001801.
3	6 OR MORE HOURS	The student answered SIX HOURS OR MORE to B001801.
8	NO RESPONSE	The student did not respond.
9	MULT.	The student filled in more than one oval.

SUBJECT-SPECIFIC COMPOSITE AND DERIVED REPORTING VARIABLES

Reading Trend Reporting Variables

RTENGAG (Engagement in Reading-related Activities)

For ages 9, 13, and 17, RTENGAG was created from items S003502, S003503, S003504, and S003506, which asked how often students engaged in certain activities related to reading. The values for these items were recoded as follows:

Daily, weekly, monthly, yearly	-	Ever
Never	-	Never

The recoded responses for all four items were then combined as follows:

Never for all four items or ever for one item	-	0-1 activity
Ever for two items	-	2 activities
Ever for three items	-	3 activities
Ever for four items	-	4 activities

RTHOME (Extent of Reading in the Home)

For ages 13 and 17, RTHOME was created from items S004501, S004502, and S004503, which asked students how often the people they live with read newspapers, magazines, and books. The items were recoded as follows:

Daily	-	5
Weekly	-	4
Monthly	-	3
Yearly	-	2
Never	-	1

The average of the three recoded responses was then assigned one of the following labels:

1.0 - 3.5	-	Low (Never/Yearly/Monthly)
3.51 - 4.5	-	Medium (Weekly)
4.51 - 5.0	-	High (Daily)

RTPLEAS (Reading of Books, Newspapers, and Magazines)

For ages 9, 13, and 17, RTPLEAS was created from items S004301, S004304, and S004305, which asked students how often they read stories or novels, newspapers or magazines. The items were recoded as follows:

Daily	= 5
Weekly	= 4
Monthly	= 3
Yearly	= 2
Never	= 1

The average of the three recoded responses was then assigned one of the following labels:

1.0 - 3.5	= Low (Never/Yearly/Monthly)
3.51 - 4.5	= Medium (Weekly)
4.51 - 5.0	= High (Daily)

<u>Writing Cross-sectional Reporting Variables</u>

WENJOY (Reported Frequency of Enjoyment of Writing)

For grades 4, 8, and 12, item S001201 was recoded as:

Almost always	= 5
More than half the time	= 4
Half the time	= 3
Less than half the time	= 2
Never	= 1

WINST2 (Frequency of Writing Process Instruction)

For grades 8 and 12, items S000606, S000608, S000609, and S000610 were recoded as:

Almost always	= 5
More than half the time	= 4
Half the time	= 3
Less than half the time	= 2
Never	= 1

The average response to these four items was then rounded to the nearest integer and the same five categories were used to classify the variable.

WPLANNG (Reported Frequency of Planning)

For grade 12, items S000901 and S000902 were recoded as:

Almost always	- 5
More than half the time	- 4
Half the time	- 3
Less than half the time	- 2
Never	- 1

The average response to these two items was then rounded to the nearest integer and the same five categories were used to classify the variable.

WREVISE (Reported Frequency of Revising and Editing)

For grade 12, items S001302, S001303, S001304, S001308, and S001310 were recoded as:

Almost always	- 5
More than half the time	- 4
Half the time	- 3
Less than half the time	- 2
Never	- 1

The average response to these five items was then rounded to the nearest integer and the same five categories were used to classify the variable.

Writing Trend Reporting Variables

WVALUE (Overall Value Placed on Writing)

For grades 8 and 12, items S001501, S001502, S001503, S001504, S001401, and S001402 were recoded as follows:

Almost always	- 5
More than half the time	- 4
Half the time	- 3
Less than half the time	- 2
Never	- 1

The average response to these six items was then converted to three categories:

Low	- less than 3.5
Medium	- between 3.5 and 4.5
High	- greater than or equal to 4.5

WATTID (General Attitude Toward Writing Composites)

For grades 4, 8, and 11, items S001201, S001202, S001204, S001205, and S001207 were recoded as follows:

Almost always	= 5
More than half the time	= 4
Half the time	= 3
Less than half the time	= 2
Never	= 1

The average response to these five items was then rounded and converted into three categories:

Low	= less than 2.5
Medium	= between 2.5 and 3.5
High	= greater than or equal to 3.5

WPUSE (Personal and Social Uses of Writing)

For grades 8 and 11, items S001601, S001602, S001603, S001604, and S001609 were recoded as follows:

Daily	= 4
Weekly	= 3
Monthly	= 2
Never	= 1

The average response to these five items was then rounded and converted to three categories:

Low	= less than 1.5
Medium	= between 1.5 and 2.5
High	= greater than or equal to 2.5

WREVED (Revising Strategies)

For grades 4, 8, and 11, items S002002 and S002003 were first recoded as:

Yes	= 1
No	= 5

Then items S000904, S000906, and S000907 were recoded as:

Almost always = 5
More than half the time = 4
Half the time = 3
Less than half the time = 2
Never = 1

The average response to these five items was then rounded and converted to three categories:

Low = less than 2.5
Medium = between 2.5 and 3.5
High = greater than or equal to 3.5

WTFEED (Teachers' Feedback on Writing)

For grades 4, 8, and 11, items S002501, S002502, S002503, S002504, S002505, and S002506 were recoded as:

Almost always = 5
More than half the time = 4
Half the time = 3
Less than half the time = 2
Never = 1

The average response to these five items was then rounded and converted into three categories:

Low = less than 3.5
Medium = between 3.5 and 4.5
High = greater than or equal to 4.5

Civics Cross-sectional Reporting Variables

PTOPIC1 (Variety of Topics Studied)

For grade 4/age 9, items P800601, P800701, P800801, P800901, P801001, and P801101 asked students to indicate whether they had studied particular civics-related topics a lot, some, or not at all. Each of the six items was coded as follows:

A lot = 3
Some = 2
Not at all = 1
Missing = M

The average of the six recoded variables was then assigned the labels and codes:

2.5-3	= A lot	= 3
1.6-2.4	= Some	= 2
1-1.5	= None	= 1
M for all	= Missing	= 0 or M

The average was across all items that had a response. PTOPIC1 appears in the data sets on the public-use data tapes for students in the focused-BIB samples.

PSTUDY2 (Whether Civics Was Studied)

For grade 8/age 13, items P800101 to P800104 asked students whether they had studied American government or civics in grades 5 to 8, respectively. Each of the four items was coded as:

Yes	= 1
No	= 0
I don't know	= 0
Missing	= M

The sum of the four recoded variables was then assigned the following labels and codes:

1-4	= Studied at least 1 year	= 1
0	= Not studied	= 2
M for all	= Missing	= 0 or M

The sum was across all items that had a response. PSTUDY2 appears in the data sets on the public-use data tapes for students in both the focused-BIB and intercorrelation samples.

PTOPIC2 (Variety of Topics Studied - Mean of 10 Questions)

For grade 8/age 13, items P801201 to P801210 asked students to indicate whether they had studied particular civics-related topics a lot, some or not at all. Each of the ten items was coded as follows:

A lot	= 3
Some	= 2
Not at all	= 1
Missing	= M

The average of the ten recoded variables was then assigned the following labels and codes:

2.5-3	= A Lot	= 3
1.6-2.4	= Some	= 2
1-1.5	= None	= 1
M for all	= Missing	= 0 or M

The average was across all items that had a response. PTOPIC2 appears in the data sets on the public-use data tapes for students in the focused-BIB samples.

PSTUDY3 (Whether Civics Was Studied)

For grade 12/age 17, items P800201 to P800204 asked students whether they had studied American government or civics in grades 9 to 12, respectively. Each of the four items was coded as:

Yes	= 1
No	= 0
I don't know	= 0
Missing	= M

The sum of the four recoded variables was then assigned the following labels and codes:

1-4	= Studied at least 1 year	= 1
0	= Not studied	= 2
M for all	= Missing	= 0 or M

The sum was across all items that had a response. PSTUDY3 appears in the data sets on the public-use data tapes for students in both the focused-BIB and intercorrelation samples.

PTOPIC3 (Variety of Topics Studied)

For grade 12/age 17, items P801301 to P801310 asked students to indicate whether they had studied particular civics-related topics a lot, some, or not at all. Each of the ten items was coded as follows:

A lot	= 3
Some	= 2
Not at all	= 1
Missing	= M

The average of the ten recoded variables was then assigned the following labels and codes:

2.5-3	- A Lot	- 3
1.6-2.4	- Some	- 2
1-1.5	- None	- 1
M for all	- Missing	- 0 or M

The average was across all items that had a response. PTOPIC3 appears in the data sets on the public-use data tapes for students in the focused-BIB samples.

PHMWK (Civics Homework)

For grade 12/age 17, values for item P801901 were assigned the following labels and codes:

No assignments	- None	- 1
Don't do assignment	- Don't do	- 2
Less than 1 hour	- Hr/less	- 3
1 hour	- Hr/less	- 3
2 hours	- 2 hrs	- 4
3 hours	- 3/more	- 5
4 hours	- 3/more	- 5
5 hours or more	- 3/more	- 5
Missing	- Missing	- 0 or M

Civics Trend Reporting Variables

PPARED (Parent's Education - 2nd Set of Categories)

For ages 13 and 17, a parents' education variable was derived for 1982 data, so that comparisons with 1988 data be made for students having a parent who graduated from college. This variable is based on the 1982 variables PARED, FCLGRAD, and MCLGRAD (whether the father or mother graduated from college). The categories for PPARED are coded as:

1	< High school	if PARED - 1 (<HS)
2	High school graduate	if PARED - 2 (HS Grad)
3	Post high school	if PARED - 3 (post-HS), FCLGRAD - 2(No), and MCLGRAD - 2 (No)
4	College graduate	if PARED - 3 and either FCLGRAD - 1 (Yes) or MCLGRAD - 1 (Yes)
	Missing	if otherwise

U.S. History Cross-sectional Reporting Variables

HTOPICS (Topics studied)

For grade 4/age 9, responses to items H801101, H801201, H801301, H801401, H801501, H801601, and H801701 were recoded as follows:

A lot	= 3
Some	= 2
Not at all	= 1

The average of the seven recoded responses was then transformed to form the HTOPICS variable as follows:

≥ 2.5	A lot	= 3
> 1.5 but < 2.5	Some	= 2
≤ 1.5	Not at all	= 1
	Insufficient data	= 0

HTAKEN (United States history studied or expected to study)

For grade 8/age 13, responses to items H800701 to H800704 were collapsed as follows:

Yes to any item	= 1
All "I don't know"	= 2
No "Yes" but at least one "No"	= 3
Insufficient data	= 0

HTOPICS (Topics studied)

For grade 8/age 13, responses to items H801801 to H801811 were recoded as follows:

A lot	= 3
Some	= 2
Not at all	= 1

The average of the 11 recoded responses was then transformed to form the HTOPICS variable as follows:

≥ 2.5	A lot	= 3
> 1.5 but < 2.5	Some	= 2
≤ 1.5	Not at all	= 1
	Insufficient data	= 0

HTAKEN (United States history studied or expected to study)

For grade 12/age 17, responses to items H800801 to H800804 were collapsed as follows:

Any Yes	= 1
All "I don't know"	= 2
No "Yes" but at least one "No"	= 3
Insufficient data	= 0

HTOPICS (Topics studied)

For grade 12/age 17, responses to items H801901 to H801913 were recoded as follows:

A lot	= 3
Some	= 2
Not at all	= 1

The average of the 13 recoded responses was then transformed to form the HTOPICS variable as follows:

≥ 2.5	A lot	= 3
> 1.5 but < 2.5	Some	= 2
≤ 1.5	Not at all	= 1
	Insufficient data	= 0

<u>Geography Cross-sectional Reporting Variables</u> (Grade 12/Age 17)

GSKILLS (Using the skills and tools of geography)

Responses to items G800201, G800202, and G800203 were recoded as follows:

A lot	= 3
Some	= 2
Very little	= 1
Not at all	= 0

The average of the three recoded responses was then transformed to form the GSKILLS variable as follows:

≥ 2.5	A lot	= 3
1.5 to < 2.5	Some	= 2
< 1.5	Very little/Not at all	= 1
	Insufficient data	= 0

GLOCAT (Knowing locations)

Responses to items G800204, G800205, G800208, and G800209 were recoded as follows:

A lot	= 3
Some	= 2
Very little	= 1
Not at all	= 0

The average of the four recoded responses was then transformed to form the GLOCAT variable as follows:

≥ 2.5	A lot	= 3
1.5 to < 2.5	Some	= 2
< 1.5	Very little/Not at all	= 1
	Insufficient data	= 0

GPHYSIC (Understanding physical geography)

Responses to items G800206 and G800207 were recoded as follows:

A lot	= 3
Some	= 2
Very little	= 1
Not at all	= 0

The average of the two recoded responses was then transformed to form the GPHYSIC variable as follows:

≥ 2.5	A lot	= 3
1.5 to < 2.5	Some	= 2
< 1.5	Very little/Not at all	= 1
	Insufficient data	= 0

GCULT (Understanding cultural geography)

Responses to items G800210 to G800215 were recoded as follows:

A lot	= 3
Some	= 2
Very little	= 1
Not at all	= 0

The average of the six recoded responses was then transformed to form the GCULT variable as follows:

≥ 2.5	A lot	= 3
1.5 to < 2.5	Some	= 2
< 1.5	Very little/Not at all	= 1
	Insufficient data	= 0

GTAKEN (Geography courses taken or expected to take)

Responses to items G800301 to G800304 were collapsed as follows:

Yes to any item	= 1
No "Yes" but at least one "NO"	= 2
All I don't know	= 3
Insufficient data	= 0

GWORLD (World history and geography course work completed)

Responses to items G800101 and G800102 were recoded as follows:

Yes, in grade 9	= 1
Yes, in grade 10	= 1
Yes, in grade 11	= 1
Yes, in grade 12	= 1
No, I haven't taken	= 2

The two items were then collapsed as follows:

Yes to either item	= 1
No to both items	= 2
Insufficient data	= 0

GEOUS (U.S. history and geography course work completed)

Responses to items G800103, G800104, and G800105 were recoded as follows:

Yes, in grade 9	= 1
Yes, in grade 10	= 1
Yes, in grade 11	= 1
Yes, in grade 12	= 1
No, I haven't taken	= 2

The three items were then collapsed as follows:

Yes to either item	= 1
No to both items	= 2
Insufficient data	= 0

GPHY.E (Physical geography/earth science course work completed)

Responses to item G800106 were recoded as follows:

Yes, in grade 9	= 1
Yes, in grade 10	= 1
Yes, in grade 11	= 1
Yes, in grade 12	= 1
No, I haven't taken	= 2

The item was then collapsed as follows:

Yes to either item	= 1
No to both items	= 2
Insufficient data	= 0

GECONCM (Economic, political, human, cultural and urban course work completed)

Responses to items G800107, G800108, and G800109 were recoded as follows:

Yes, in grade 9	= 1
Yes, in grade 10	= 1
Yes, in grade 11	= 1
Yes, in grade 12	= 1
No, I haven't taken	= 2

The three items were then collapsed as follows:

Yes to either item	= 1
No to both items	= 2
Insufficient data	= 0

GTOPICS (Average of GSKILLS, GLOCAT, GPHYSIC and GCULT)

The average of the four variables GSKILLS, GLOCAT, GPHYSIC, and GCULT was transformed to form the GTOPICS variable as follows:

≥ 2.5	A lot of study	= 3
> 1.5 but < 2.5	Some study	= 2
0 - 1.5	Little or No study	= 1
	Insufficient data	= 0

GLOSSARY OF TERMS

GLOSSARY OF TERMS

anchoring. The process of characterizing score levels in terms of predicted observable behavior.

assessment session. The period of time during which a NAEP booklet is administered to one or more individuals.

average response method (ARM). A regression-based technique to predict for a respondent the conditional distribution of an average score on a set of items given responses to at least one of the items and other information.

background questionnaires. The instruments used to collect information about students'demographics and educational experiences.

bias. In statistics, the difference between the expected value of an estimator and the population parameter being estimated. If the average value of the estimator over all possible samples (the estimator's expected value) equals the parameter being estimated, the estimator is said to be unbiased; otherwise, the estimator is biased.

BIB (Balanced Incomplete Block) spiraling. A complex variant of multiple matrix sampling, in which items are administered in such a way that each pair of items is administered to a nationally representative sample of respondents.

BILOG. A computer program for estimating item parameters.

block. A group of assessment items created by dividing the item pool for an age/grade into subsets. Used in the implementation of the BIB spiral sample design.

booklet. The assessment instrument created by combining blocks of assessment items.

bridging. An administration of the same set of exercises under two different conditions or to two different populations to allow a statistical link (bridge) to be established between results under the different circumstances.

calibrate. To estimate the parameters of a set of items from responses of a sample of examinees.

clustering. The process of forming sampling units as groups of other units.

codebook. A formatted printout of NAEP data for each student, excluded student, teacher, and school in a particular grade/age.

coefficient of variation. The ratio of the standard deviation of an estimate to the value of the estimate.

combined ratio estimator. The ratio estimator resulting from first estimating the numerator and the denominator values and then using the quotient of these as the estimate of the ratio.

common block. A group of background items included in the beginning of every assessment booklet.

conditional probability. Probability of an event, given the occurrence of another event.

conditioning variables. Demographic and other background variables characterizing a respondent. Used in construction of plausible values.

cross-sectional assessment. An assessment that provides information about differences in educational performance across subgroups of students. It does not provide information about changes in students' educational performance across time. It may, however, provide baseline data for measuring future trends.

Current Population Survey. A household sample survey conducted monthly by the Bureau of the Census to provide estimates of employment, unemployment, and other characteristic of the general labor force, the population as a whole, and various subgroups of the population.

degrees of freedom. [of a variance estimator] The number of independent pieces of information used to generate a variance estimate.

derived variables. Subgroup data that were not obtained directly from assessment responses, but through procedures of interpretation, classification, or calculation.

design effects. The ratio of the variance for the sample design to the variance for a simple random sample of the same size.

distractor. An incorrect response choice included in a multiple-choice item.

excluded student questionnaire. An instrument completed for every student who was sampled but excluded from the assessment.

excluded students. Sampled students determined by the school to be unable to participate because they have limited English proficiency, are mildly mentally retarded (educable), or are functionally disabled.

expected value. The average of the sample estimates given by an estimator over all possible samples. If the estimator is unbiased, then its expected value will equal the population value being estimated.

field test. A pretest of items to obtain information regarding clarity, difficulty levels, timing, feasibility, and special administrative situations; performed before revising and selecting items to be used in the assessment.

focused-BIB spiraling. A variation of BIB spiraling in which items are administered in such a way that each pair of items *within a subject area* is administered to a nationally representative sample of respondents.

foils. The correct and incorrect response choices included in a multiple-choice item.

group effect. The difference between the mean for a group and the mean for the nation.

holistic scoring. A method of evaluating students' writing for overall fluency in responding to a task.

imputation. Prediction of a missing value according to some procedure, using a mathematical model in combination with available information. See **plausible values.**

imputed race/ethnicity. The race or ethnicity of an assessed student, as derived from his or her responses to particular common background items. A NAEP reporting subgroup.

item response theory (IRT). Test analysis procedures that assume a mathematical model for the probability that a given examinee will respond correctly to a given exercise.

jackknife. A procedure to estimate standard errors of percentages and other statistics. Particularly suited to complex sample designs.

machine-readable catalog. Computer processing control information, IRT parameters, foil codes, and labels in a computer readable format.

major strata. Used to stratify the primary sampling frame within each region. Involves stratification by size of community and degree of ruralization (SDOC).

Metropolitan statistical area (MSA). An area defined by the federal government for the purposes of presenting general-purpose statistics for metropolitan areas. Typically, an MSA contains a city with a population of at least 50,000 plus adjacent areas.

modal age. The age of the majority of a group of grade-eligible students.

modal grade. The grade attended by the majority of a group of age-eligible students.

mode of administration. The method by which students are administered assessment instruments. Both printed and tape-recorded administration methods are used.

multistage sample design. Indicates more than one stage of sampling. An example of three-stage sampling: 1) sample of counties (primary sampling units or PSUs); 2) sample of schools within each sample county;

3) sample of students within each sample school.

multiple matrix sampling. Sampling plan in which different samples of espondents take different samples of items.

NAEP scales. The anchored scales common across age/grade levels and assessment years used to report NAEP results.

nonresponse. The failure to obtain responses or measurements for al. ample elements.

nonsampling error. A general term applying to all sources of error except sampling error. Includes errors from defects in the sampling frame, response or measurement error, and mistakes in processing the data.

objective. A desirable education goal agreed upon by scholars in the field, educat rs, and concerned laypersons, and established through the consensus approach.

observed race/ethnicity. Race or ethnicity of an assessed student as perceived by the exercise administrator.

open-ended response item. A non-multiple-choice item that requires some ty written or oral response.

oversampling. Deliberately sampling a portion of the population at a higher rate than the remainder of the population.

paced tape. The audio recording that accompanies some booklets to assure uniformity in administration. Recorded instructions prevent reading difficulties from interfering with an individual's ability to respond.

parental education. The level of education of the mother and father of an assessed student as derived from the student's response to two assessment items. A NAEP reporting subgroup.

percent correct. The percent of a target population that would answering a particular exercise correctly.

plausible values. Proficiency values drawn at random from a conditional distribution of a NAEP respondent, given his or her response to cognitive exercises and a specified subset of background variables (conditioning variables). The selection of a plausible value is a form of imputation.

poststratification. Classification and weighting to correspond to external values of selected sampling units by a set of strata definitions after the sample has been selected.

primary sampling unit (PSU). The basic geographic sampling unit for NAEP. Either a single county or a set of contiguous counties.

primary trait scoring. A method of evaluating students' writing for effectiveness in accomplishing the specific goal or purpose of each writing task.

principal questionnaire. A data collection form given to school principals before assessments. The principals respond to questions concerning enrollment, size and occupational composition of the community, etc.

probability sample. A sample in which every element of the population has a known, nonzero probability of being selected.

pseudoreplicate. The value of a statistic based on an altered sample. Used by the jackknife variance estimator.

public-use data tapes. Computer tapes containing respondent-level cognitive, background and attitude, and demographic data. Available for use by researchers wishing to do secondary analyses of NAEP data.

QED. Quality Education Data, Inc. A supplier of lists of schools, school districts, and other school data.

random variable. A variable that takes on any value of a specified set with a particular probability.

region. One of four geographic areas used in gathering and reporting data: Northeast, Southeast, Central, and West (as defined by the Office of Business Economics, U.S. Department of Commerce). A NAEP reporting subgroup.

reporting subgroup Groups within the national population for which NAEP data are reported: for example, gender, race/ethnicity, grade, age, level of parental education, region, and size and type of community.

respondent. A person who is eligible for NAEP, is in the sample, and responds by completing one or more items in an assessment booklet.

response options. In a multiple-choice question, alternatives that can be selected by a respondent.

sample. A portion of a population, or a subset from a set of units, selected by some probability mechanism for the purpose of investigating the properties of the population. NAEP does not assess an entire population but rather selects a representative sample from the group to answer assessment items.

sampling error. The error in survey estimates that occurs because only a sample of the population is observed. Measured by sampling standard error.

sampling frame. The list of sampling units from which the sample is selected.

sampling weight. A multiplicative factor equal to the reciprocal of the probability of a respondent being selected for assessment with adjustment for nonresponse and perhaps also for poststratification. The sum of the weights provides an estimate of the number of persons in the population represented by a respondent in the sample.

school characteristics and policy questionnaire. A questionnaire completed for each school by the principal or other official; used to gather information concerning school administration, staffing patterns, curriculum, and student services.

selection probability. The chance that a particular sampling unit has of being selected in the sample.

simple random sample. Process for selecting n sampling units from a population of N sampling units so that each sampling unit has an equal chance of being in the sample and every combination of n sampling units has the same chance of being in the sample chosen.

size and type of community (STOC). One of the NAEP reporting subgroups, dividing the communities in the nation into seven groups on the basis of size and other characteristics.

standard error. A measure of sampling variability and measurement error for a statistic. Because of NAEP's complex sample design, sampling standard errors are estimated by jackknifing the samples from first-stage sample estimates. Standard errors may also include a component due to the error of measurement of individual scores estimated using plausible values.

stratification. The division of a population into parts, called strata.

stratified sample. A sample selected from a population that has been stratified, with a sample selected independently in each stratum. The strata are defined for the purpose of reducing sampling error.

student ID number. A unique identification number assigned to each respondent to preserve his or her anonymity. NAEP does not record the names of any respondents.

subject area. One of the areas assessed by National Assessment; for example, art, civics, computer competence, geography, literature, mathematics, music, reading, science, U.S. history, or writing.

systematic sample (systematic random sample). A sample selected by a systematic method; for example, when units are selected from a list at equally spaced intervals.

teacher questionnaire. A questionnaire completed by selected teachers of sample students; used to gather information concerning years of teaching experience, frequency of assignments, teaching materials used, and availability and use of computers.

trend assessment. An assessment based on replicating past procedures in order to report changes in educational achievement across time.

variance. The average of the squared deviations of a random variable from the expected value of the variable. The variance of an estimate is the squared standard error of the estimate.

REFERENCES CITED IN TEXT

REFERENCES CITED IN TEXT

Allen, R., Bettis, N., Kurfman, D., MacDonald, W., Mullis, I. V. S., & Salter, C. (1990). *The geography learning of high-school seniors.* (No. 19 G 01) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Anderson, L., Jenkins, L. B., Leming, J., MacDonald, W. B., Mullis, I. V. S., Turner, M. J., & Wooster, J. S. (1990). *The 1988 civics report card: Trends in achievement from 1976 to 1988 at ages 13 and 17 and achievement in 1988 at grades 4, 8, and 12.* Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Applebee, A. N., Langer, J. A., & Mullis, I. V. S. (1986). *The writing report card. Writing achievement in American schools.* (No. 15-W-02) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Applebee, A. N., Langer, J. A., & Mullis, I. V. S. (1988). *Who reads best? Factors related to reading achievement in grades 3, 7, and 11.* (No. 17-R-01) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Applebee, A. N., Langer, J. A., & Mullis, I. V. S. (1989). *Understanding direct writing assessments. Reflections on a South Carolina writing study.* Princeton, NJ: Educational Testing Service.

Applebee, A. N., Langer, J. A., Mullis, I. V. S., & Jenkins, L. B. (1990). *The writing report card, 1984-88. Findings from the Nation's Report Card.* (No. 19-W-01) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Applebee, A. N., Langer, J. A., Mullis, I. V. S., Jenkins, L. B., & Foertsch, M. A. (1990). *Learning to write in our nation's schools: Instruction and achievement in 1988 at grades 4, 8, and 12.* Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E. (1987a). The NAEP reading scale. In A. E. Beaton, *Implementing the new design. The NAEP 1983-84 technical report* (pp. 381-390). (No. 15-TR-20) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E. (1987b). *Implementing the new design. The NAEP 1983-84 technical report.* (No. 15 TR 20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E. (1988). *Expanding the new design: The NAEP 1985-86 technical report.* (No. 17-TR 20) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E., & Johnson, E. G. (1990). The average response method of scaling. *Journal of Educational Statistics, 15*, 9-38..

Beaton, A. E., & Johnson, E. G. (1987). The average response method (ARM) of scaling. In A. E. Beaton, *Implementing the new design. The NAEP 1983-84 technical report* (pp. 435-480). (No. 15-TR-20) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E., & Zwick, R. (1990). *The effect of changes in the National Assessment. Disentangling the NAEP 1985-86 reading anomaly.* (No. 17-TR-21) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Bock, R. D., Gibbons, R. D., & Muraki, E. (1985). *Full-information item factor analysis*. (MRC Report No. 85-1) Chicago: National Opinion Research Center.

Bock, R. D., Muraki, E., & Pfeiffenberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement, 25*, 275-285.

Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). *Assessing writing skill.* (Research Monograph Number 11) New York: College Entrance Examination Board.

Caldwell, N., Moore, D., & Slobasky, R. (1989). *Report on field operations and data collection activities - NAEP 1988*. Rockville, MD: Westat, Inc.

Campbell, A. (1987). Professional scoring. In A. E. Beaton, *Implementing the new design. The NAEP 1983-84 technical report* (pp. 175-184). (No. 15-TR-20) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis.* Boston: Duxbury Press.

Cochran, W. F. (1977). *Sampling techniques*. New York: John Wiley & Sons.

Educational Testing Service. (1987). *ETS standards for quality and fairness*. Princeton, NJ. Educational Testing Service.

Educational Testing Service. (1990). *NAEP. On the cutting edge of measurement since 1969*. Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Frankel, M. R. (1971). *Inference from survey samples*. Ann Arbor. Institute for Social Research, The University of Michigan.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd edition). New York. American Council on Education/Macmillan.

Hammack, D. C., Hartoonian, M., Howe, J., Jenkins, L. B., Levstik, L. S., MacDonald, W. B., Mullis, I. V. S., & Owen, E. (1990). *The U.S. history report card: The achievement of fourth-, eighth-, and twelfth-grade students in 1988 and trends from 1986 to 1988 in the factual knowledge of high school juniors*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, E. G. (1988). Mathematics data analysis. In A. E. Beaton, *Expanding the new design. The NAEP 1985-86 technical report* (pp. 215-240). (No. 17-TR-20) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Johnson, E. G. (1989). Considerations and techniques for the analysis of NAEP data. *Journal of Educational Statistics, 14*, 303-334.

Johnson, E. G. & King, B. F. (1987). Generalized variance functions for a complex sample survey. *Journal of Official Statistics, 3*, 235-250.

Johnson, J. R. (1987). Instrument and item information. In A. E. Beaton, *Implementing the new design. The NAEP 1983-84 technical report* (pp. 119-134). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

King, B. F., Bertrand, R. & Dupuis, F. A. (1989). *A world of differences. An international assessment of mathematics and science. Technical report.* Princeton, NJ: Educational Testing Service, International Assessment of Educational Progress.

Kirsch, I. S., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults, final report.* (No. 16-PL-01). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Kish, L. (1965). *Surve, sampling.* New York: John Wiley & Sons.

Kish, L. & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Series B, 36*, 1-22.

Krewski, D. & Rao. J. N. K. (1981). Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication. *Annals of Statistics, 9*, 1010-1019.

Langer, J. A., Applebee, A. N., Mullis, I. V. S., & Foertsch, M. A. (1990). *Learning to read in our nation's schools: Instruction and achievement in 1988 at grades 4, 8, and 12.* Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Little, R. J. A., & Rubin, D. B. (1983). On jointly estimating parameters and missing data. *American Statistician, 37*, 218-220.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ. Lawrence Erlbaum Associates.

Messick, S. J., Beaton, A. E., & Lord, F. M. (1983). *NAEP reconsidered: A new design for a new era.* (NAEP Report 83-1) Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*(3), 359-381.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association, 80*, 993-997.

Mislevy, R. J. (1988a). *Randomization-based inferences about latent variables from complex samples.* (ETS Research Report RR-88-54-ONR) Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (1988b). Scaling procedures. In A. E. Beaton, *Expanding the new design. The NAEP 1985-86 technical report* (pp. 177-204). (No. 17-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Mislevy, R. J. (in press). Randomization-based inferences about latent variables from complex samples. *Psychometrika.*

Mislevy, R. J., & Bock, R. D. (1982). *BILOG. Item analysis and test scoring with binary logistic models* [Computer program]. Mooresville, IN: Scientific Software.

Mislevy, R. J., & Bock, R. D. (1989). *PC-BILOG 3. Item analysis and test scoring with binary logistic models.* Mooresville, IN: Scientific Software.

Mislevy, R. J., & Sheehan, K. M. (1987). Marginal estimation procedures. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 293-360). (No. 15-TR 20) Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Mislevy, R. J., & Sheehan, K. M. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika, 54*, 661-679.

Mislevy, R. J., & Wu, P-K. (1988). *Inferring examinee ability when some item responses are missing.* (ETS Research Report RR-88-48-ONR) Princeton, NJ: Educational Testing Service.

Mullis, I. V. S., & Jenkins, L. B. (1990). *The reading report card, 1971-88. Trends from the Nation's Report Card.* (No. 19-R-01) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

National Assessment of Educational Progress. (1985). *The reading report card. Progress toward excellence in our schools.* (No. 15-R-01) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

National Assessment of Educational Progress. (1987a). *Reading objectives, 1986 and 1988 assessments.* (No. 17/19-R-10) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

National Assessment of Educational Progress. (1987b). *Civics objectives, 1988 assessment.* (No. 19-C-10) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M. (1987). Data entry system. In A. E. Beaton, *Implementing the new design. The NAEP 1983-84 technical report* (pp. 185-199). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Barone, J. L., & Kline, D. L. (1989). *A guide to the NAEP item information database.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Barone, J. L., & Kline, D. L. (1989). *A guide to the NAEP restricted-use data files.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Kline, D. L., Barone, J. L., Mychajlowycz, A. W., & Forer, D. C. (1989). *A primer for the NAEP item information database.* Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Kline, D. L., Johnson, E. G., Mislevy, R. M., & Rust, K. F. (1990). *National Assessment of Educational Progress 1988 public-use data tapes version 2.0 user guide.* Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Rogers, A. M., Kline, D. L., Johnson, E. G., & Rust, K. F. (1989). *National Assessment of Educational Progress 1988 public-use data tapes version 1.0 user guide.* Princeton, NJ. Educational Testing Service, National Assessment of Educational Progress.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York. John Wiley & Sons.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics, 1(4)*, 381-397.

Rust, K. F., Bethel, J., Burke, J., & Hansen, M. H. (1990). *The 1988 National Assessment of Educational Progress--Sampling and weighting procedures, final report.* Rockville, MD: Westat, Inc.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6*, 309-316.

Sheehan, K. M. (1985). *M-GROUP: Estimation of group effects in multivariate models* [Computer program]. Princeton, NJ: Educational Testing Service.

Sheehan, K. M. (1988). The IRT linking procedure used to place the 1986 intermediary scaling results onto the 1984 reading calibration scale. In A. E. Beaton, *Expanding the new design. The NAEP 1985-86 technical report* (pp. 555-565). (No. 17-TR 20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Stocking, M. L. (1986). *TBLT: Transforming b's using a least squares technique* [Computer program]. Princeton, NJ: Educational Testing Service.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201-210.

Wilson, D., Wood, R. L., & Gibbons, R. D. (1983). *TESTFACT: Test scoring and item factor analysis* [Computer program]. Chicago: Scientific Software.

Wingersky, M., Kaplan, B. A., & Beaton, A. E. (1987). Joint estimation procedures. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 285-292). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Wolter, K. M. (1985). *Introduction to variance estimation*. New York: John Wiley & Sons.

Yamamoto, K. (1988). Science data analysis. In A. E. Beaton, *Expanding the new design: The NAEP 1985-86 technical report* (pp. 243-255). (No. 17-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Yamamoto, K. (1990). Mathematics and science trend data analysis. In A. E. Beaton & R. Zwick, *The effect of changes in the National Assessment: Disentangling the NAEP 1985-86 reading anomaly* (pp. 111-144). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Zwick, R. (1987a). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement, 24*, 293-308.

Zwick, R. (1987b). Use of the NAEP almanacs. In A. E. Beaton, *Implementing the new design. The NAEP 1983-84 technical report* (pp. 517-522). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Zwick, R. (1988). *Professionally scored items*. Technical memorandum.

INDEX

INDEX

Absentee data
  adjustment of, for makeup sessions, 164
  data entry for, 114
  editing of, 155
Administration Schedule, 93, 95, 96
  materials processing, 112, 114–15
Age definition, 15, 16
  for all student samples, 16
  changes in, 7, 15
  for civics cross-sectional samples, 299
  for civics trend samples, 299
  effect of changes in, 189
  for geography cross-sectional samples, 326
  for mathematics trend samples, 330, 332
  for reading cross-sectional samples, 253
  for reading trend samples, 253
  for science trend samples, 330, 332
  for U.S. history cross-sectional samples, 318
  for U.S. history trend sample, 318
  for writing cross-sectional samples, 268
  for writing trend samples, 268
Anchoring, 6, 182–83
  of civics cross-sectional scale, 308–9
  of reading trend scale, 256
  of U.S history cross-sectional scale, 322, 323
Answer documents. *See also* Response methods/modes
  evolution of, 101
  key-entered booklets, 101
  and placement of open-ended item scores, 128
  scannable answer sheets, 101, 110
  scannable booklets, 101
ARM. *See* Average response method
Assessment instruments, 27–31, 69–83. *See also* Booklets; Item blocks; Excluded student questionnaire; School characteristics and policies questionnaire; Teacher questionnaire
Assessment items, summary statistics for, 344. *See also* Item blocks; Item development
Assessment Policy Committee, 33, 34, 43, 47
Assessment sessions
  summary statistics for, 345
  types of, 60–63
Assistant field directors
  responsibilities, 85
Average response method (ARM), 275–78
  application to the writing cross-sectional scale, 291–93
  application to the writing trend scale, 287–88
  composite, 275, 276, 277
  computation of plausible values for, 240–42
  estimable correlations for writing cross-sectional analysis, 293
  estimable correlations for writing trend analysis, 281
  estimation of cross-products matrix
    directly estimable terms, 278–82
    imputation of inestimable correlations, 279, 283–88
    terms estimable from other grades or years, 279, 282–83
  and scaling, 235
  use of, in writing data analysis, 182

Background and attitude question development, 26–27, 33–50
  for civics and U.S. history, 45
  for geography, 46
  for reading, 39
  for writing, 42
Balanced incomplete block (BIB) spiral design, 14–15, 27–29. *See also* Focused-BIB spiral design
  introduction of, in NAEP, 4
Bias
  consideration of, in secondary analyses, 245–247
  review of item text for, 34, 39, 42
BIB-spiral design. *See* Balanced incomplete block spiral design
BILOG computer program, 181, 261, 233, 239, 305, 321, 327, 337
Blocks. *See* Item blocks
Booklets
  arrangement of blocks in, for main samples, 70
  assignment to students, 29
  blocks used, for bridge booklets, 71, 78–82, for focused-BIB booklets, 69, 72–74; for intercorrelation booklets, 69, 72–74; for special study booklets, 70, 72–74
  civics, 298, 311, 312
  design of, for bridge samples, 29–30; for focused-BIB samples, 6, 27–28; for intercorrelation samples, 6, 28
  mathematics trend, 330
  number administered, for bridge samples, 78–80; for main samples, 72–74

Booklets (continued)
number created for main samples, 71
response methods, for bridge samples, 78–80; for main samples, 72, 73, 74; for reading, 254
science trend, 330
summary statistics, 344
used for all samples, 16
Bridge samples, 17–18, 179, 189–90. *See also* Trend samples
age definition for, 16
block contents for, 78–80
booklets used for, 16, 71, 78–82
for civics trend analysis, 311
introduction of, in NAEP, 5
for mathematics trend analysis, 329
modal grades for, 16
mode of administration for, 16
number of booklets administered in, 78–80
for reading anomaly analysis, 6, 255
for reading trend analysis, 252, 255
response methods for booklets in, 78–80
sample sizes for, 16
for science trend analysis, 329
student cohorts assessed for, 16
time of testing for, 16
for U.S history trend analysis, 317
for writing trend analysis, 268

Calculator items, 331
Chief State School Officers, 31, 87, 88
Civics assessment development
background questions, 45
items, 44
objectives, 43
Civics cross-sectional data analysis, 297–310
anchoring, 308
conditioning effects, 306, 574–85
conditioning variables, 306, 529–36; composite and derived, 306, 309, 620–28
dichotomization of open-ended items, 310
differential item functioning, 305
dimensionality, 301, 304–5
effect of block position, 301, 303
item analysis, 301
item calibration, 306
item classification, 310
item parameters, 305, 654–57
items excluded from scaling, 305
items used, 298, 300
KR-20 reliabilities for item blocks, 301, 302
not-reached percents for item blocks, 301, 302
parameter estimation, 305
plausible values, 305–6
populations assessed, 297
proficiency estimation, 306–7
reporting variables, composite and derived, 309, 699–702
sample sizes, 300
samples used, 298
scaling, 306–9
transformation of plausible values, 307
Civics cross-sectional samples
age definition for, 299
booklets used for, 299
modal grades for, 299
sample sizes for, 299
time of testing for, 299
Civics data analysis, 297–316. *See also* Civics cross-sectional data analysis; Civics trend data analysis
Civics trend data analysis, 311–316
conditioning effects, 314, 586–92
conditioning variables, 314, 537–44; composite and derived conditioning variables 315, 628
dichotomization of open-ended items, 316
equating, 314
item analysis, 313
item calibration, 313
item parameters, 314, 658–61
item response function, 313
items excluded from scaling, 311, 313
items used, 311, 312
KR-20 reliabilities for item blocks, 313
not-reached percents for item blocks, 313
parameter estimation, 313
plausible values, 314
populations assessed, 311
proficiency estimation, 315
reporting variables, composite and derived, 315, 702
sample sizes, 312
samples used, 311
transformation of plausible values, 315
Civics trend samples
age definition for, 299
booklets used for, 299
modal grades for, 299
mode of administration for, 311, 313
sample sizes for, 299
time of testing for, 299
Coefficient of variation
definition, 347
Common background questions
composite and derived, for reporting, 687–93
content of, 47
development of, 47
timing of, 47
Conditioning effects
for civics cross-sectional data, 306, 574–85
for civics trend data, 314, 586–92
for geography cross-sectional data, 327, 599–600
for mathematics trend data, 334, 335, 601–6

for reading cross-sectional data, 262, 559–68
for reading teacher data, 264, 569–70
for reading trend data, 256, 571–73
for science trend data, 339, 607–9
for U.S. history cross-sectional data, 322, 593–98
Conditioning variables
for 1988 scales, 248, 249
for civics cross-sectional data, 306, 529–36; composite and derived, 306, 309, 620–27
for civics trend data, 314, 537–44; composite and derived, 315, 628
common, 503–7
for geography cross-sectional data, 327, 549; composite and derived, 635–36
for mathematics trend data, 334, 335, 550–54; composite and derived, 637–38
and plausible values, 238, 240
for reading cross-sectional data, 262, 508–19; composite and derived, 613–14
for reading teacher data, 264
for reading trend data, 256, 520–22
and reduction of potential bias in secondary analyses, 246
for science trend data, 339, 555
for U.S. history cross-sectional data, 322, 545–48; composite and derived, 628–35
for writing cross-sectional data, 291, 523–25; composite and derived, 294–295, 615–18
for writing trend data, 275, 279, 526–28; composite and derived, 619–20
Consultants
for objectives and item development, 35, 471–80

Data analysis, 179–85. *See also* Civics data analysis, Geography data analysis; Mathematics data analysis; Reading data analysis; Science data analysis; Teacher data analysis; U.S. history data analysis; Writing data analysis
anchoring, 182–83
average response method, 182
dichotomization, 181
differential item functioning, 181
item analysis, 180
item response function, 181
item response theory, 181
parameter estimation, 181
plausible values, 181
proficiency estimation, 181
reliability for item blocks, 180
samples of students, 179–80
scaling, 181–83
speededness checks, 180
Data entry
for absentee data, 114
for excluded student questionnaire, 123
for school characteristics and policies questionnaire, 124
for school worksheet, 111–12
system, 152–53
for teacher questionnaire, 124
Data processing. *See also* Data transcription, Database creation; Editing data; Materials processing
flow, 102
scope of work, 100
systems, 100
Data transcription, 145–53
accuracy of, 100
entry and resolution, 152–53
for excluded student questionnaire, 102
loading, 150–52
quantities, 100
scanning, 145, 148–49
for school questionnaire, 102
for teacher questionnaire, 102
time period for, 100
Database. *See also* Database creation
overview of, 105
size, 107
structure, 106
Database creation, 105, 163–66
derivation of weights for, 163–64
master catalog for, 165–66
merging files for, 164–65
Database products, 167–75. *See also* Item information database; Public-use data tapes/files; Restricted-use data files
usefulness, 8, 106
Degrees of freedom
of jackknife variance estimator, 220, 222–28
Design and Analysis Committee, members of, 4
Design effects
for proportion-correct statistics, 212–17
for simple regression coefficients, 220, 221
for subgroup mean proficiency scores, 217–19
use of, in estimating sampling variability, 212–20
Design of assessment, 13–31. *See also* Assessment instruments; Background and attitude question development; Item development; Objectives development; Sample design
age definition, 15
assessment instruments, 27–31
background question development, 26–27
BIB spiral design, 14–15, 27–29
booklet design, 27–30
bridge samples, 17–18
considerations in, 14
focused-BIB samples, 15, 17
focused-BIB spiral design, 15, 28
half-samples, 17
improvements in, 14

753

Design of assessment (continued)
intercorrelation samples, 17
item block design, 27
item development, 26–27
modal grades, 15
objectives development, 26–27
overview of design, 14–15
populations assessed since 1969, 19–22
sample selection, 19, 23–26
special study samples, 17
subject areas assessed in 1988, 13; since 1969, 19–22
time of testing, 15
Dichotomization of open-ended items, 181
for civics, 310, 316
for reading, 257
for U.S. history, 319
Differential item functioning (DIF), 181
for civics cross-sectional data, 305
consideration of, in scaling, 233
for reading cross-sectional data, 261
Dimensionality
of civics cross-sectional items, 301, 304–5
consideration of, in scaling, 234
of reading trend items, 255
District supervisors
numbers of, 86
responsibilities, 86, 90, 91, 93–94
training, 86–87
Document literacy, 7, 180
booklet contents for, 70
number of open-ended items for, 127
preliminary item analysis for, 252
scoring of open-ended items for, 131

Editing data, 155–58
for absentees, 155
for assessed students, 156
for professionally scored items, 157
for questionnaires, 156–57
Education Commission of the States, 229, 242
Eligibility
distribution of students by, 345
for students in bridge samples, 23
for students in main samples, 19
Equating
for civics trend data, 314
for mathematics trend data, 330, 331, 334, 335
for reading cross-sectional data, 263
and scaling, 235
for science trend data, 330, 331, 337
Error analysis. *See* Quality control
"Estimate" items, 335
Excluded student questionnaire
content, 47, 48, 83
data transcription, 102
development, 47
editing data from, 156–57
and field administration, 95, 96
materials processing for, 122–25
number of items in, 344
quality control error analysis for, 161
Excluded students
distribution of, by age class and sample, 345
introduction of, in NAEP, 4
population estimates (weighted) for, 347
rates of exclusion, 66, 96, 98
and sample design, 64, 66
Exercise administrators, 93, 94
recruitment, 86, 91
responsibilities, 86, 93, 95, 96
training, 91

Field administration, 31–32, 85–98. *See also* Field administration forms; Field administration staff
arranging assessments, 94
distributing and collecting questionnaires, 95–96
district supervisors' responsibilities, 86, 90, 91, 93–94
makeup sessions, 93
obtaining cooperation of districts and schools, 87–91
organization of, 85–86
participation rates, for schools, 90, 92; for students in fall assessments, 96, 97; for students in winter/spring assessments, 97, 98
preparing reports and shipping materials, 96
sample sizes, for students in fall assessments, 96, 97; for students in winter/spring assessments, 97, 98
selection of students, 94–95
training of district supervisors, 86–87
Field administration forms. *See* Administration Schedule; Introductory Meeting Form; Results of Contact Form; Roster of Questionnaires; School Update Form, School Worksheet; Session Assignment Form; Student Listing Form; Teacher Survey Roster
Field administration staff. *See* Assistant field directors, District supervisors; Exercise administrators; Field director; Project director; School coordinators
Field director, 94
responsibilities, 85–87
Field testing
and item development process, 48–49
Focused-BIB samples, 15, 17, 180
age definition for, 16
block used, 72–74
booklet contents, 69
booklets used for, 16

754

for civics cross-sectional analysis, 297, 298
for geography analysis, 325
modal grades for, 16
mode of administration for, 16
number of booklets administered for, 72–74
for reading cross-sectional analysis, 252, 256
response methods used for, 72–74
sample sizes for, 16
student cohorts assessed for, 16
time of testing for, 16
for U.S. history cross-sectional analysis, 317
for writing cross-sectional analysis, 268, 289
Focused-BIB spiral design, 5–6, 15, 28
and reduction of potential bias in secondary analyses, 246

Geography assessment development
background questions, 46
items, 46
objectives, 45
Geography cross-sectional data analysis, 325–28
conditioning effects, 327, 599–600
conditioning variables, 327, 549; composite and derived, 635–36
item analysis, 325
item calibration, 327
item classification, 327–28
item parameter estimation, 327
item parameters, 666–67
item response theory, 327
items excluded from scaling, 325
KR-20 reliabilities for item blocks, 326
mean proportion correct for item blocks, 326
not-reached percents for item blocks, 325
plausible values, 327
proficiency estimation, 327
reporting variables, composite and derived, 704–7
samples used, 325
scaling, 327
Geography cross-sectional samples
age definition for, 326
booklets used for, 326
modal grades for, 326
sample sizes for, 326
time of testing for, 326
Geography data analysis. *See* Geography cross-sectional data analysis
Grade/age. *See* Student cohorts

Half-samples
purpose of, 17
student weights for, 202
Holistic data
analysis of, for writing trend, 288
correlation of scores with primary trait scores, 132
items used for, in writing trend, 289
sample sizes for writing trend, 289
scoring, 133–34
training of scorers, 138–39

Implementation of assessment. *See* Field administration; Materials processing; Professional scoring
Imputation. *See* Plausible values
Instruments, 27–31, 69–83. *See also* Booklets; Item blocks; Excluded student questionnaire; School characteristics and policies questionnaire; Teacher questionnaire
Intercorrelation samples, 17, 180
age definition for, 16
block used for, 72–74
booklet contents for, 69
booklets used for, 16
for civics cross-sectional analysis, 297
for geography cross-sectional analysis, 325
modal grades for, 16
mode of administration for, 16
number of booklets administered for, 72–74
purpose of, 6
for reading cross-sectional analysis, 252, 256
response methods used for, 72–74
sample sizes, 16
student cohorts assessed for, 16
time of testing for, 16
for U.S. history cross-sectional analysis, 317
Intraclass correlation coefficient, calculation of, 140n
Introductory Meeting Form, 88, 89, 90
IRT. *See* Item response theory
Item analysis, 180
consideration of, in scaling 234
for civics cross-sectional data 301
for civics trend data 313
for document literacy items 252
for geography cross-sectional data 325
for mathematics trend data 331
for reading cross-sectional data 258
for reading trend data 255
for science trend data 337
for U.S. history cross-sectional data 319
Item blocks
arrangement of, 70
assembly of, for field test, 39; for final assessment, 49
background items in, for main samples, 75–77
cognitive blocks: number created for main samples, 70; use in bridge sample booklets, 78–80; use in main sample booklets, 72–74
cognitive items in, for main samples, 75–77
common background, 69; used in bridge sample booklets, 78–80; used in main sample

Item blocks *(continued)*
booklets, 72–74
design of, 27
effect of position of, in booklets: for civics, 301, 303; for reading, 258; for writing, 294
for mathematics trend samples 330
for science trend samples 330
subject area background: use in bridge sample booklets, 78–80; use in main sample booklets, 72–74
timing of, for bridge samples, 71, 81, 82; for main samples, 70
total of, by subject area, 50
total open-ended items in, for main samples, 75–77
Item calibration
for civics cross-sectional data, 306
for civics trend data, 313
for geography cross-sectional data, 327
for reading cross-sectional data, 261
for reading trend data, 256
for U.S. history cross-sectional data, 321
Item classification
for civics cross-sectional data, 310
for geography cross-sectional data, 327–28
for U.S. history cross-sectional data, 323–24
Item development, 26–27, 33–50
for civics and U.S. history, 44
consensus process, 34
consultants for, 35, 471–80
general procedures for, 37–38
for geography, 46
major considerations in, 34
for reading, 39
schedule, 35
for writing, 41
Item information database, 8, 99, 167–68
creation procedures, 167
overview, 106
purpose of, 167
size, 100
structure, 167
Item response function, 181
for civics trend data, 313
for mathematics trend data, 335
for reading cross-sectional data, 261
for reading trend data, 256
Item response theory (IRT). *See also* Item response theory parameter estimation; Item response theory parameters
introduction of, in NAEP, 4
and scaling, 232
use in data analysis, 181
Item response theory parameters
for civics cross-sectional data, 305, 654–57
for civics trend data, 314, 658–61
for geography cross-sectional data, 666–67
for mathematics trend data, 334, 335, 668–72
for reading cross-sectional data, 262, 647–50
for reading trend data, 256, 651–53
for science trend data, 337, 673–77
for U.S. history cross-sectional data, 662–65
Item response theory parameter estimation, 181
for civics cross-sectional data, 305
for civics trend data, 313
for geography cross-sectional data, 327
for mathematics trend data, 331, 335
for reading cross-sectional data, 261
for reading trend data, 256
and scaling, 233
for science trend data, 337
for U.S. history cross-sectional data, 321
Items used in assessment, summary statistics for, 344. *See also* Item blocks; Item development
Items excluded from scaling
for civics cross-sectional analysis, 305
for civics trend analysis, 311, 313
for geography cross-sectional analysis, 325
for mathematics trend analysis, 331
for reading trend analysis, 256
for science trend analysis, 331
for U.S. history cross-sectional analysis, 321

Jackknife variance estimator. *See* Sampling variability, estimation of

Key entry
quantities, 100
KR-20 reliabilities
for civics cross-sectional item blocks, 301, 302
for civics trend item blocks, 313
for geography cross-sectional item blocks, 326
for reading cross-sectional item blocks, 258
for reading trend item blocks, 255
for U.S. history cross-sectional item blocks, 319, 320

Learning Area Committees
and item development, 34, 35, 37, 39–42, 44–46
and objectives development, 34, 35, 36, 40, 43–45
selection of, 35, 36

M-GROUP computer program, 181, 238, 305, 327, 334
Main samples, 15–17, 189. *See also* Focused-BIB samples; Intercorrelation samples; Special study samples
Makeup sessions, 93
number of, by age class and sample, 345
Materials processing, 102, 109–25
for administration schedules, 112, 114–15

data entry, 111, 113, 114, 123, 124
flow, 102
loading scanned data, 117–19, 123
for professionally scored responses, 115–16
quality control, 120, 125
for questionnaires, 122
receipt of materials, 109–11
resolution processing, 119–22, 123
scanning, 110, 116–17, 123
for school worksheets, 111–12, 113
for student instruments, 115–22
validation, 120, 124
verification, 120
Mathematics data analysis. *See* Mathematics trend data analysis
Mathematics trend data analysis, 329–337
conditioning effects, 334, 335, 601–6
conditioning variables, 334, 335, 550–54; composite and derived, 637-38
equating, 330, 331, 334-35
"estimate" items, 335
item analysis, 331
item parameter estimation, 331, 335
item parameters, 334, 335, 668–72
item response function, 335
items excluded from scaling, 331
main objective of, 330
mean proportion correct for item blocks, 333
plausible values, 334
proficiency estimation, 334
proficiency scores, 335
scaling, 331, 333–37
Mathematics trend samples
age definition for, 330, 332
booklets for, 330
modal grades for, 332
mode of administration for, 329, 330
sample sizes, 332
time of testing for, 330, 332
Mechanics data
analysis of, for writing trend, 288
items used for writing trend, 288
sample sizes for writing trend, 288
scoring, 133
training of scorers for, 138
Modal grades
for all student samples, 16
changes in, 15
for civics cross-sectional samples, 299
for civics trend samples, 299
for geography cross-sectional samples, 326
introduction of, in NAEP, 5
for mathematics trend samples, 332
for reading cross-sectional samples, 253
for reading trend samples, 253
for science trend samples, 332
for U.S. history cross-sectional samples, 318
for U.S. history trend sample, 318
for writing cross-sectional samples, 268
for writing trend samples, 268
Mode of administration
for all student samples, 16
for civics trend samples, 311, 313
effect of changes in, 189
for mathematics trend samples, 329, 330, 332
for science trend samples, 329, 330, 332

National Center for Education Statistics, 4
National Geographic Society, 33, 34, 45
Nonresponse, adjustments for. *See* Weighting procedures
Not-reached percents
for civics cross-sectional item blocks, 301, 302
for civics trend item blocks, 313
consideration of, in scaling, 234
for geography cross-sectional item blocks, 325
for reading cross-sectional item blocks, 258, 259
for reading trend items, 255
for U.S. history cross-sectional item blocks, 319

Objectives booklets
review and publication, 49
Objectives development, 26–27, 33–50
for civics, 43
consensus process, 34
consultants for, 35, 471–80
general procedures for, 35–36
for geography, 45
for reading, 38
major considerations in, 34
schedule, 35
for U.S. history, 43
for writing, 40
Office of Educational Research and Improvement (OERI) review of publications, 35
Office of Management and Budget (OMB) clearance
for all items, 35, 49, 50
for background questionnaires, 47
for civics and U.S. history items, 44
for geography items, 46
for reading items, 39
for writing items, 42
Open-ended items
and civics data analysis, 298, 300, 310, 316
numbers of, in bridge samples, 127; in main samples, 127
and reading data analysis, 256, 257, 258, 263
response times for, in bridge samples, 130; in main samples, 129
score ranges for, in bridge samples, 130; in main samples, 129

Open-ended items *(continued)*
scorer reliability for, in bridge samples, 143–44; in main samples, 141–42
total of, in each main sample block, 75–77
and U.S. history data analysis, 317, 319
Oversampling
of Black and Hispanic students, 5, 52, 57, 188, 190–191
of Black students, for writing mechanics analysis, 288

Paced tape. *See* Mode of administration
Parental notification or permission, 94
Participation rates
for schools, in bridge samples, 61, 90, 92; in main samples, 60, 90, 92; summary, 90, 92
for students, in fall assessments, 96, 97; in winter/spring assessments, 97, 98; by age class, 66
Percent correct
definition, 258n
for civics cross-sectional item blocks, 302
for civics trend item blocks, 313
for reading cross-sectional item blocks, 260
Plausible values, 236–248
for civics cross-sectional scale, 305–6
for civics trend scale, 314
computation of, in average response method scales, 240–42; in IRT-based scales, 237–40
and conditioning variables, 238, 240
and data analysis, 181
on database, 165
example of use in secondary analyses, 247–48
as intermediary computations only, 237
for mathematics trend samples, 334
methodology, for scaling, 235–37
for nonrespondents, 237
on public-use data tapes, 243
purpose of, 6, 236
teacher-based, for reading, 264
transformation to proficiency scale, 182; for civics cross-sectional, 307; for civics trend, 315; for geography cross-sectional, 327; for mathematics trend, 334, 337; for science trend, 337, 339; for U.S. history cross-sectional, 322
for U.S. history cross-sectional scale, 322
use of, in secondary analyses, 242–48
for writing cross-sectional scale, 291
for writing trend scale, 275–278, 288
Population estimates
excluded students, 347
students, 346
Populations assessed
in 1988, 188
since 1969, 19–22
Poststratification, 198–202
to adjust for oversampling in writing trend analysis, 288
changes in procedures, 199–201
effect of changes in procedures, 201-202, 203–205
for excluded student weights, 206
improvements in, 8
revisions for 1984 weights, 201, 641–42
for teacher-student weights, 207
Primary sampling units (PSUs), 23–24, 52–56
certainty and noncertainty, 53
designation by season for main assessments, 55
designation for bridge assessments, 55–56
drawn for multiple assessment years, 55
final sample of, 54
formation of, 52–54
stratification of, by geographic region, 53; by minority population, 54; by MSA or nonMSA, 53; by socioeconomic characteristics, 54
Primary trait data
agreement and reliability of scores, 139–144, 271–273
correlation of scores with holistic scores, 132
effect of across-year variation in scores, 270
analysis of, for writing cross-sectional, 289–91; for writing trend, 268, 270–75
sample sizes, for writing cross-sectional, 290; for writing trend, 274
scoring, 132–33
training for scoring, 137–38
Principal questionnaire, 60
Professional scoring, 127–44. *See also* Open-ended items; Primary trait data
answer documents and placement of scores, 128
editing data for, 157
guidelines for scoring of open-ended items, for civics, 135–36; for document literacy, 131; for mathematics, 134–35; for reading, 128, 131; for science, 135; for U.S. history, 136; for writing, 131–34
holistic scoring, 133–34
materials processing, 115–16
mechanics scoring, 133
number of items scored, 100
number of responses scored, 100
primary trait scoring, 132–33
reliability and resolution, 139–44
scoring supervisor, 137
scoring teams, 128, 137
training of scorers, 137–39
work schedule, 139
Proficiency estimates
of students, by subpopulation, 347
Proficiency estimation, 181
difference in, for reading trend, 255

for civics cross-sectional scale, 306–8
for civics trend scale, 315
for geography cross-sectional scale, 327
for mathematics trend samples, 334, 335
for reading cross-sectional scale, 262
for reading trend scale, 256
for science trend samples, 337
for U.S. history cross-sectional scale, 322
Project director (field administration)
responsibilities, 86, 87
Proportion correct
for geography cross-sectional item blocks, 326
for mathematics trend item blocks, 333
for science trend item blocks, 338
for U.S. history cross-sectional item blocks, 320
PSUs. See Primary sampling units
Public-use data tapes/files, 8, 99, 168–75
contents, 14
codebooks for, 173–74
control statement files for, 174–75
data definition for, 171–72
data file catalogs for, 173
data file layouts for, 172–73
file definition for, 169–70
machine-readable catalog files for, 175
overview, 106
plausible values on, 243
purpose of, 168, 343
size, 100
variables definition for, 170–71

Quality control
confidence limits for, 159, 161
of data entry, 159–62
error analysis, for questionnaire data entry, 161; for student data entry, 161
for materials processing, 120, 125
observed error rate for, 159, 160
quantities involved in, 100
Quality Education Data, Inc. (QED), 57, 59, 165
Questionnaires. See Excluded student questionnaire; Principal questionnaire; School characteristics and policies questionnaire; Teacher questionnaire

Reading assessment development
background questions, 39
items, 39
objectives, 38
Reading anomaly, 179, 235
data analysis for, 251
samples used for analysis of, 6, 255
Reading cross-sectional data analysis, 256–65
conditioning effects 262, 559–68; for teacher data, 264, 569–570
conditioning variables 262, 508–19; for teacher data, 264; composite and derived, 613–14
dichotomization of open-ended items, 257
differential item functioning, 261
effect of block position, 258
equating, 263
item analysis, 258
item calibration, 261
item parameters, 262, 647–50
item response function, 261
KR-20 reliabilities, 258
not-reached percents for item blocks, 258, 259
number of items administered, 257
parameter estimation, 261
percent correct for item blocks, 258, 260
proficiency estimation, 262
purpose of, 257
samples used, 252, 256
scaling, 262
scaling metric, 262–64
speededness, 258
teacher data analyses, 264–65
teacher-based plausible values, 264
Reading cross-sectional samples
age definition for, 253
booklets used for, 253
modal grades for, 253
response modes for, 254
sample sizes, 253
time of testing for, 253
Reading data analysis, 251–65. See also Reading cross-sectional data analysis; Reading trend data analysis
and reading anomaly, 251, 255
document literacy items, 252
main goals of, 251
sample characteristics, 253
Reading trend data analysis, 255–56
anchoring, 256
conditioning effects, 256, 571–73
conditioning variables, 256, 520–22
dimensionality, 255
item analysis, 255
item calibration, 256
item parameters, 256, 651–53
item response function, 256
items excluded from scaling, 256
KR-20 reliabilities for item blocks, 255
not-reached percents for items, 255
parameter estimation, 256
proficiency estimation, 256
purpose, 255
reporting variables, composite and derived, 695–96
samples used, 252, 255
scaling, 256

Reading trend samples
age definition for, 253
booklets used for, 253
modal grades for, 253
response modes for, 254
sample sizes, 253
time of testing for, 253
Reliability
internal consistency, for item blocks, 180; for civics cross-sectional blocks, 301, 302; for civics trend blocks, 313; for geography cross-sectional blocks, 325; for reading cross-sectional blocks, 258; for reading trend blocks, 255; for U.S. history cross-sectional blocks, 319
of professional scoring results, 139–44
rater, for writing trend scores, 270–273
Reporting subgroups, 681–86
age, 686
gender, 681
grade in school, 685–86
parental education, 684
race/ethnicity, imputed, 681–82
race/ethnicity, observed, 681
region of the country, 684
size and type of community (STOC), 682–83
Reporting variables, composite and derived
civics cross-sectional, 309, 699–702
civics trend, 315, 702
common background, 687–93
geography cross-sectional, 704–7
reading trend, 695–96
U.S. history cross-sectional, 703–4
writing cross-sectional, 696–97
writing trend, 697–99
Reports
list of, for 1988 subject areas, 3n
RESOLVE computer program, 263
Response methods/modes. *See also* Answer documents
for main sample booklets, 72–74
for reading booklets, 254
Restricted-use data files, 8, 99, 168
creation procedures, 168
overview, 106
purpose of, 168
structure, 168
Results of Contact form, 88–90
Roster of Questionnaires, 93

Sample design, 19, 23–26, 51–68, 188–190
assignment of sessions to schools, 60–63
excluded students, 64, 66
exclusion rates for students, 66
oversampling of minority students, 52, 57
participation rates for schools in bridge samples, 61; in main samples, 60
participation rates for students, 66
participation rates, overall, 68
populations sampled, 51
primary sampling units, 52–56
sample sizes for schools in bridge samples, 61; in main samples, 60
selection of schools for bridge samples, 60–63; for main samples, 57–60
selection of students, 63–64
selection of teachers, 67
session types for bridge samples, 60
size and type of community (STOC), 60, 682–83
stages of selection, 51
time of testing, 52
types of assessment, 52
Sample size(s)
for all student samples, 16
for civics cross-sectional samples, 299, 300
for civics trend samples, 299, 312
for geography cross-sectional samples, 326
for mathematics trend samples, 332
for reading cross-sectional samples, 253
for reading trend samples, 253
for schools in bridge samples, 61; in main samples, 60; by session type, 65
for science trend samples, 332
for students in fall assessments, 96, 97; in winter/spring assessments, 97, 98; by session type, 65
for U.S. history cross-sectional samples, 318
for U.S. history trend sample, 318
for writing cross-sectional primary trait analysis, 290
for writing trend holistic analysis, 289
for writing trend mechanics analysis, 288
for writing trend primary trait analysis, 274
summary statistics, 344–45
Sample weights. *See also* Weighting procedures
and database creation, 163–64
excluded student weights, 206
jackknife replicate weights, 207, 208, 210
student full-sample weight, 202, 206
student season-specific weight, 202, 206
teacher-student weights, 206
used for reading teacher analysis, 265
Samples of students, 15–23, 179–80. *See also* Bridge samples; Focused-BIB samples; Intercorrelation samples; Special study samples
Sampling variability, estimation of, 187–88, 207–28
degrees of freedom of jackknife variance estimator, 220, 222–28
effect of sample design on, 187, 208
jackknife replicate weights, 208, 210
using design effects, 212–20
using jackknife variance estimator, 208–11

Scaling, 181–83, 229–49. *See also* Plausible values
advantages and disadvantages, 230
average response method (ARM) model for, 235
biases in secondary analyses, 245–47
and changes in item context, 235
and changes in speededness conditions, 235
for civics cross-sectional data, 306–9
for civics trend data, 313–315
comparison of reporting scale and Consumer Price Index, 230–31
computing plausible values in ARM scales, 240–42
computing plausible values in IRT-based scales, 237–40
detection of conditional dependence in, 233
and differential item functioning, 233
and dimensionality, 234
and equating, 235
example of use in secondary analyses, 247–48
for geography cross-sectional data, 327
and item response theory, 232
and item analysis, 234
for mathematics trend data, 331, 333–37
methodology, 231–42
metric, for reading cross-sectional data, 262–64
and not-reached items, 234
overview of 1988 scales, 248–49
parameter estimation, 233
plausible values methodology for, 235–37
for reading cross-sectional data, 262
for reading trend data, 256
for science trend data, 337–39
in secondary analyses, 242–48
statistical tests for secondary analyses, 244–45
three-parameter logistic IRT model for, 232–35
for U.S. history cross-sectional data, 319, 321–22
for writing cross-sectional data, 291–93
for writing trend data, 275–88
Scanning, 110, 116–17, 123, 145, 148–49
quantities, 100
School characteristics and policies questionnaire 60
contents, 47, 48, 83
data transcription, 102
development, 47
editing data from, 156–57
field administration of, 95, 96
materials processing, 122–25
number of items in, 344
quality control error analysis for, 161
timing of, 48
School coordinators, 93, 94
assignment, 94
responsibilities, 93, 95, 96
School districts
obtaining cooperation of, 87–91
School Update Form, 88–90
School Worksheet, 93
materials processing, 111–12, 113
Schools
obtaining cooperation of, 87–91
participation rates, for bridge samples, 61, 90, 92; for main samples, 60, 90, 92; summary, 90, 92, 344
sample sizes, for bridge samples, 61; for main samples, 60; by session type, 65
selection of, for bridge samples, 60–63; for main samples, 57–60
summary statistics, 344
Science data analysis. *See* Science trend data analysis
Science trend data analysis, 329–31, 337–39
conditioning effects, 339, 607–9
conditioning variables, 339, 555
equating, 330, 331, 337
item analysis, 337
item parameter estimation, 337
item parameters, 337, 673–77
items excluded from scaling, 331
main objective of, 330
mean proportion correct for item blocks, 338
proficiency estimation, 337
scaling, 337–39
Science trend samples
age definition for, 330, 332
booklets used for, 330
modal grades for, 332
mode of administration for, 329, 330
sample size, 332
time of testing for, 330, 332
Session Assignment Form, 63, 91, 95
Sessions
summary statistics for, 345
types of, 60–63
Size and type of community (STOC), 60, 682–83
Special study samples, 7, 17, 180
age definition for, 16
booklets used, 16; contents, 70
modal grades for, 16
mode of administration for, 16
sample sizes, 16
student cohorts assessed for, 16
time of testing for, 16
Speededness
checks for, in item analysis, 180
considerations of, in scaling 235
and reading cross-sectional analysis, 258
State Advisory Committee, 43, 44
State education agencies
and assessment development, 42
State Testing Directors, 43
STOC. *See* Size and type of community
Stratification. *See* Primary sampling units

Student cohorts
for all student samples, 16
for civics cross-sectional samples, 299
for civics trend samples, 299
for geography cross-sectional samples, 326
for mathematics trend samples, 330, 332
for reading cross-sectional samples, 253
for reading trend samples, 253
for science trend samples, 330, 332
target populations, 19, 23
for U.S. history cross-sectional samples, 318
for U.S. history trend sample, 318
for writing cross-sectional samples, 268
for writing trend samples, 268
Student data
editing of, 156
materials processing, 115–22
quality control error analysis for, 160
Student Listing Form, 93
Students
cohorts assessed for each sample, 16
distribution of, by age class and sample, 345
population estimates (weighted) for, 346
proficiency estimates for, by subpopulation, 347
Subject areas
assessed in 1988, 3, 13
assessed since 1969, 19–22
list of reports for, 3
rationale for including, 33

TBLT computer program, 263, 314
Teacher data analysis
for reading, 264–65
for writing, 295
Teacher questionnaire
changes in, 8
contents, 47, 82
data editing, 156–57
data transcription, 102
development, 47
field administration of, 95, 96
materials processing, 122–25
number of items in, 344
quality control error analysis for, 161
summary statistics, 345
timing, 48
Teacher sample
selection of, for reading, 67; for writing, 67
Teacher Survey Roster, 93
Technical report
audience, 3
organization of, 9
purpose, 3
TESTFACT computer program, 301
Three-parameter logistic model
computing plausible values for, 237–40
use in data analysis, 181; for geography, 327; for reading, 261; for U.S. history, 321
use in scaling, 232–235
Time of testing, 15
for all student samples, 16
changes in, 15
for civics cross-sectional samples, 299
for civics trend samples, 299
effect of changes in, 189
for geography cross-sectional samples, 326
for mathematics trend samples, 330, 332
for reading cross-sectional samples, 253
for reading trend samples, 253
for science trend samples, 330, 332
by type of assessment, in sample design, 52
for U.S. history cross-sectional samples, 318
for U.S. history trend sample, 318
for writing cross-sectional samples, 268
for writing trend samples, 268
Trend samples. *See* Bridge samples; Civics trend samples; Mathematics trend samples; Reading trend samples; Science trend samples; U.S. history trend sample, Writing trend samples

U.S. history assessment development
background questions, 45
items, 44
objectives, 43
U.S. history cross-sectional data analysis, 317–24
anchoring, 322, 323
conditioning effects, 322, 593–98
conditioning variables, 322, 545–48; composite and derived, 628–35
dichotomization of open-ended items, 319
item analysis, 319
item calibration, 321
item classification, 323–24
item response theory, 321
items excluded from scaling, 321
item parameters, 662–65
items used, 317, 321
KR-20 reliabilities for item blocks, 319, 320
mean proportion correct for item blocks, 320
not-reached percents for item blocks, 319
parameter estimation, 321
plausible values, 322
proficiency estimation, 322
reporting variables, composite and derived, 703–4
scaling, 319, 321–22
U.S. history cross-sectional samples
age definition for, 318
booklets used for, 318
modal grades for, 318
sample sizes, 318

time of testing for, 318
U.S. history data analysis, 317–24. *See also* U.S. history cross-sectional data analysis; U.S. history trend data analysis
U.S. history trend data analysis, 324
item analysis, 324
item classification, 324
items used, 324
U.S. history trend sample
age definition for, 318
booklets used for, 318
modal grades for, 318
sample size, 318
time of testing for, 318

Weighting procedures, 190–207
adjustments for age-only eligibles nonresponse, 195; school nonresponse, 193; session nonresponse, 194; student nonresponse, 196
distribution of weight components for 1988 samples, 483–96
for excluded student weights, 20?
nonresponse adjustments for excluded student weights, 206; for student weights, 192–97; for teacher-student weights, 207
poststratification of excluded student weights, 206; of student weights, 198–202; of teacher-student weights, 207
for student base weight, 191–92
tape and spiral sample designations for, 190
for teacher-student weights, 206
trimming of excluded student weights, 206; of student weights, 197–98; of teacher-student weights, 207
Writing assessment development
background questions, 42
items, 41
longer response time for items, 7, 41, 50; booklets used to measure, 70
objectives, 40
Writing cross-sectional data analysis, 289–96
average response method, 291–93
components of, 289
conditioning variables, 291, 523–25; composite and derived, 294, 2?5, 615–18
effect of block position on performance, 294
effect of longer response time on performance, 295–96
items used, 290
plausible values, 291
primary trait analysis, 289
process and instruction analysis, 294–95
purpose of, 267
reporting variables, composite and derived, 696–97
sample sizes, 290
samples used, 268
scaling, 291–93
teacher data analysis, 295
Writing cross-sectional samples
age definition for, 268
modal grades for, 268
time of testing for, 268
Writing data analysis, 267-96. *See also* Writing cross-sectional data analysis; Writing trend data analysis
objectives of, 267
Writing trend data analysis, 268–89
agreement and reliability for primary trait scores, 271–273
average response method (ARM), 275
conditioning variables, 275, 279, 526–28; composite and derived, 619–20
effect of across-year variation in primary trait scoring, 270–75
holistic data analysis, 288, 289
items used, 268, 269, 288, 289
plausible values, 275–278, 288
primary trait analysis and scoring, 268
purpose of, 267
reporting variables, composite and derived, 697–99
rescoring of 1984 writing papers, 270
sample sizes, for holistic analysis, 289; for mechanics analysis, 288; for primary trait analysis, 274
samples used, 268
scaling, 287–88
Writing trend samples
age definition for, 268
modal grades for, 268
time of testing for, 268

Appendix 16

END

U.S. Dept. of Education

Office of Education Research and Improvement (OERI)

ERIC

Date Filmed

March 29, 1991