DOCUMENT RESUME

ED 135 864 TM 006 095

AUTHOR Shann, Mary; And Others
TITLE Student Effects of an Interdisciplinary Curriculum for Real Problem Solving: The 1974-75 USMES Evaluation. Final Report.
INSTITUTION Boston Univ., Mass.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE Dec 75
GRANT PES-74-00542-A01
NOTE 407p.; For related documents, see TM 005 913 and TM 006 092

EDRS PRICE MF-$0.83 HC-$22.09 Plus Postage.
DESCRIPTORS Achievement Gains; Basic Skills; Curriculum Development; *Curriculum Evaluation; Elementary Education; *Elementary School Mathematics; *Elementary School Science; *Interdisciplinary Approach; Mathematics Curriculum; *Problem Solving; Process Education; Program Attitudes; Program Evaluation; Science Curriculum; *Student Reaction
IDENTIFIERS *Unified Science Mathematics for Elementary Schools

ABSTRACT

This document contains the report of the Unified Science and Mathematics for Elementary Schools (USMES) evaluation team on the 1974-75 investigation of the cognitive and affective responses of students to USMES, an interdisciplinary, process curriculum designed to develop the problem solving abilities of elementary school students. Included in this evaluation of USMES are the results of a pre-post control group design to assess the curriculum's effect on students' basic skill development, their attitude change, and their progress in complex problem solving as measured by small group performance on real-life relevant, simulated problem tasks. The results of interviews with USMES teachers and students, unstructured observations during site visitations and field staff's documentation of USMES usage are also included in this report. (Author/MV)

FINAL REPORT

Grant Number PES 74-00542 A01

STUDENT EFFECTS OF AN INTERDISCIPLINARY CURRICULUM FOR REAL PROBLEM SOLVING: THE 1974-75 USMES EVALUATION

by

The USMES Evaluation Staff

Mary H. Shann, Project Director
Norma C. Reali, Associate Director
Hilary Bender, Editorial Consultant

and

Thomas Aiello, Data Analyst
Linda Hench, Research Assistant

Boston University

December, 1975

National Science Foundation
Materials and Instruction Development Section

FINAL REPORT

Grant Number PES 74-00542 A01

STUDENT EFFECTS OF AN INTERDISCIPLINARY CURRICULUM FOR REAL PROBLEM SOLVING: THE 1974-75 USMES EVALUATION

by

The USMES Evaluation Staff

Mary H. Shann, Project Director
Norma C. Reali, Associate Director
Hilary Bender, Editorial Consultant

and

Thomas Aiello, Data Analyst
Linda Hench, Research Assistant

Boston University

December, 1975

National Science Foundation
Materials and Instruction Development Section

Acknowledgements

This final report is submitted by the evaluation team, but it represents the cooperative efforts of a great many more people.

First, without the generous cooperation of the students, teachers, and administrators of the elementary schools in our sample, there would have been only limited opportunity to evaluate the USMES curriculum. Since these people gave of their time for extensive data collection, with little incentive other than the hope of contributing to better educational programs, we feel they deserve our sincerest appreciation. We thank them also for their graciousness and kindness to us during our interview visits.

Next, much of the credit for obtaining the cooperation of the schools belongs to our field staff. Without our observers' insights and their tact in dealing with the school personnel, we might never have secured, and retained, our sample.

We would be remiss if we did not acknowledge the unfailing cooperation of the USMES Central Staff throughout the conduct of this evaluation. They facilitated our entree into USMES schools for data collection, and they honored with dispatch our requests for various kinds of information and materials. Curriculum developers and evaluators are natural adversaries, but the USMES Central Staff exhibited an open and honest interest in timely feedback, including critical findings, from the evaluation team.

As Project Manager for both the USMES development project and the USMES evaluation project, Dr. Raymond J. Hannapel of the National Science Foundation facilitated among the participants the exchange of ideas and information so crucial to a responsive evaluation study.

Our advisory board helped us to sharpen our data gathering tools for the study and to articulate new techniques for assessing problem solving. They pointed out errors and ambiguities in the text of this report, some of which we have been able to correct. Professors Jeremy Kilpatrick, Fletcher Watson, and Wayne Welsh deserve our sincere thanks for their many valuable and constructive suggestions for the improvement of this evaluation and for their continued encouragement.

To our administrative and secretarial assistant, Miss Anita Paci, we express a great share of our appreciation. Her occasional absences demonstrated how sorely we depended on her help. Her humor and efficiency in the conduct of many thankless tasks deserve to be acknowledged with heartful applause.

Finally, our gratitude goes to our families, who, during our work on this study, put up with many inconveniences and a few long absences. Their expressions of encouragement and their support did not go unnoticed.

December, 1975
Boston, Massachusetts

Mary H. Shann
Project Director

TABLE OF CONTENTS

Chapter		Page
	Table of Contents	i
	List of Tables	v
	List of Figures	xi
I.	AN OVERVIEW OF THE USMES PROGRAM AND THE USMES EVALUATION PROJECT	1
	Introduction	1
	The Nature of USMES: Its Philosophy and Goals	2
	Issues Investigated During the 1973-74 USMES Evaluation	3
	Focus of the 1974-75 Evaluation Project	6
	Organization of the Evaluation Report	7
	The Report on New Instrument Development	8
II.	METHODS AND PROCEDURES	9
	Observer Training	9
	Sample Selection and Procurement	10
	Characteristics of Sample Schools and Classes	13
	A. Geographical Distribution	13
	B. Population Densities and Socioeconomic Levels	14
	C. Grade Levels	15
	D. Class Size	16
	Teacher Characteristics	16
	A. Teaching Experience	16
	B. Teachers' Specialized Training in Mathematics and Science	16
	Observation of USMES and Control Classes	17
	Assessment of Student Effects of USMES	18
	Analysis of Student Performance Data	20
	Interviews and Other Data Collection During Site Visitations	22
	Logistics for Data Collection	22
	Shared Emphasis on New Instrument Development	23
	Summary	24
III.	CHARACTERIZING ACTUAL TREATMENTS	26
	The Special Need for Contrasting Actual Treatments for USMES and Control Groups	26
	Length and Intensity of USMES Usage Experienced by Sample USMES Classes	27
	A. Teacher Experience with USMES	28
	B. Student Experience with USMES	31
	C. Intensity of USMES Usage During 1974-75	32
	Design Lab Facilities in the USMES Schools	32
	Exposure to USMES in Sample USMES Schools	35

Table of Contents (Continued)

Chapter		Page
	Apportionment of Time Across School Subjects and Activities by USMES and Control Classes	37
	Non-USMES Curricula in USMES and Control Classes	40
	Activities and Interactions in USMES and Control Classes	42
	A. Expected Differences Between USMES and Control Groups	42
	B. Procedures for the Observation of Student Behaviors	42
	C. Results of Classroom Activity Analysis	45
	D. Comparisons with the 1973-74 Classroom Activity Analysis Results	47
	Chapter Summary	51
IV.	INTERVIEWS WITH USMES TEACHERS AND THEIR STUDENTS	53
	The Interview Respondents	53
	Arrangements for the Interviews	54
	Focus of the Interviews	55
	Interviews with USMES Teachers	57
	A. The Strategy of Teacher Interviews	57
	B. Responses from USMES Teachers	58
	C. Summary of Teacher Responses	68
	Interviews with USMES Students	68
	A. The Strategy of Student Interviews	69
	B. Responses from USMES Students	70
	C. Summary of Student Responses	78
	Conversations with Principals	79
	A Synthesis of Responses	81
V.	THE EFFECTS OF USMES ON STUDENTS' BASIC SKILL DEVELOPMENT	85
	Background for This Assessment	85
	Use of the Stanford Achievement Tests	86
	Procedures for Test Administration and Scoring	89
	Data Analysis	91
	A. Reading Comprehension	94
	B. Mathematics Computation	95
	C. Mathematics Application	98
	D. Mathematical Concepts	103
	E. Science	106
	F. Social Science	109
	Additional Problems for the Analysis and Interpretation of SAT Data	110
	Discussion and Summary	117

Table of Contents (Continued)

Chapter		Page
VI.	STUDENT PERFORMANCE ON SIMULATED REAL-LIFE PROBLEM TASKS: RESULTS OF THE PLAYGROUND AND PICNIC PROBLEMS	127
	Conceptual Basis for the Playground and Picnic Problems	128
	Test Administration	130
	Scoring Procedures and Scorer Reliability	131
	Other Reliability and Validity Evidence	134
	Results	140
	A. Behavioral Aspects	140
	B. Cognitive Aspects	143
	1. Repeated Measures Analyses of Cognitive Scores	146
	2. Covariance Analyses of Cognitive Scores ...	164
	3. Specific Factors Which Students Considered in Their Solutions to the Playground and Picnic Problems	167
	C. Product Aspects	175
	Chapter Summary	175
VII.	SURVEYING ATTITUDE CHANGE IN USMES AND CONTROL STUDENTS ...	179
	The Importance of Attitude Assessment	179
	Attitudes and USMES	180
	What Attitudes to Measure and How	181
	Pre-testing	185
	Pilot Testing	186
	Post-testing	187
	Determining the Factor Structure of the Instrument	187
	Scoring and Analysis	195
	Results for Part I of the Attitude Scale	196
	A. Factor 1: Science Appreciation	200
	B. Factor 2: Academic Insecurity	204
	C. Factor 3: Non-mastery	208
	D. Factor 4: Preference for Group Learning Activities	212
	E. Factor 5: Arithmetic Enjoyment	212
	F. Factor 6: Self Directed Learning	219
	G. Factor 7: Arithmetic Value	223
	Results from Part II of the Attitude Scale	223
	A. Factor 1: Need to Please--Acquiesce	229
	B. Factor 2: Divergent Thinking	232
	C. Factor 3: Academic Competition	237
	D. Factor 4: Convergent Thinking	241
	Summary and Discussion	241
VIII.	MORALE PROBLEM: POSSIBLE FOCUS FOR FUTURE EVALUATIONS	251

Table of Contents (Continued)

Chapter | Page

IX. REPORT SUMMARY 254

Focus for the 1974-75 Evaluation Report 254
Sample Selection, Data Collection, and Method of Analysis 255
Characteristics of Experimental and Control Classes 256
Interview Results: Teachers, Students, Administrators 258
Basic Skill Development 260
Proof of Concept Assessment 264
Attitude Changes in USMES and Control Students 265
Recommendation for a Future Study 268
Conclusion 269

REFERENCES 270

APPENDICES 273

A. School Information Form - 1974-75 273
B. Class Information Form - 1974-75 275
C. USMES Classroom Activity Analysis - 1974-75 281
D. Administrator's Manual for the Playground Problem. A Measure of Problem Solving Ability for Use in the Evaluation of USMES. Prepared by the USMES Evaluation Staff, Boston University 284
E. Administrator's Manual for the Picnic Problem. A Measure of Problem Solving Ability for Use in the Evaluation of USMES. Prepared by the USMES Evaluation Staff, Boston University 294
F. Scoring Protocol for the Playground Problem: A Manual for Rating and Coding Students' Performance on a Test of Complex Problem Solving. Prepared by Mary H. Shann, Ph.D., USMES Evaluation Project Director 304
G. Scoring Protocol for the Picnic Problem: A Manual for Rating and Coding Students' Performance on a Test of Complex Problem Solving. Prepared by Mary H. Shann, Ph.D., USMES Evaluation Project Director 321
H. Teacher Interview - Spring, 1975 340
I. Child Interview - Spring, 1975 347
J. Pilot-test Attitude Scale - Form 0 355
K. Pre-test Attitude Scale - Form 1 361
L. Pre-test Attitude Scale - Form 2 366
M. Post-test Attitude Scale - Form 3 371

LIST OF TABLES

Table		Page
3. 1	Time Spent on USMES Activities Including Design Lab Work by 37 Sample USMES Classes During 1974-75	33
3. 2	Average Number of Hours Reportedly Spent on Selected School Subjects by USMES and Control Classes	38
3. 3	Results of 1974-75 Classroom Activity Analysis: Percentages of Observers' Tallies in 29 Student Behavior Categories for USMES and Control Groups	44
5. 1	Number of Sample Classes with Pre-test and Post-test Scores Used in the Analysis of Stanford Achievement Test Data	88
5. 2	Form of SAT for Pre-test and Post-test by Single Grade and Combinations of Grades	90
5. 3	Repeated Measures Analysis of Variance for SAT Reading Comprehension	96
5. 4	Means and Standard Deviations for SAT Reading Comprehension	97
5. 5	Repeated Measures Analysis of Variance for SAT Mathematics Computation	99
5. 6	Means and Standard Deviations for SAT Mathematics Computation	100
5. 7	Repeated Measures Analysis of Variance for SAT Mathematics Application	101
5. 8	Means and Standard Deviations for SAT Mathematics Application	102
5. 9	Repeated Measures Analysis of Variance for SAT Mathematics Concepts	104
5.10	Means and Standard Deviations for SAT Mathematics Concepts	105
5.11	Repeated Measures Analysis of Variance for SAT Science	107
5.12	Means and Standard Deviations for SAT Science	108
5.13	Repeated Measures Analysis of Variance for SAT Social Science	111
5.14	Means and Standard Deviations for SAT Social Science	112

List of Tables (Continued)

Table		Page
5.15	Standard Deviations of Scores on 6 SAT Subtests	114
5.16	Forms of the SAT Administered Within Groups of Grades	116
6. 1	Number of Sample Classes with Pre-test and Post-test Scores Used in the Analysis of the Playground Problem and the Picnic Problem Grade Level by Treatment	132
6. 2	Correlations Between Picnic and Playground Problem Scores on Pre-test and Post-test for USMES and Control Classes, and for Total Sample	136
6. 3	Correlations Between Pre-test and Post-test on Playground and Picnic Problems for USMES and Control Classes, and Total Sample	138
6. 4	Distribution of Pre-test and Post-test Ratings of Reliability/Validity Problems for USMES and Control Classes on the Playground Problem	141
6. 5	Distribution of Pre-test and Post-test Ratings of Reliability/Validity Problems for USMES and Control Classes on the Picnic Problem	142
6. 6	Distribution of Pre-test and Post-test Ratings on Behavioral Aspects of Students' Performance on the Playground Problem	144
6. 7	Distributions of Pre-test and Post-test Ratings on Behavioral Aspects of Students' Performance on the Picnic Problem	145
6. 8	Repeated Measures Analysis of Variance for the Playground Problem Variable, Identification	147
6. 9	Repeated Measures Analysis of Variance for the Playground Problem Variable, Measuring	148
6.10	Repeated Measures Analysis of Variance for the Playground Problem Variable, Calculates	149
6.11	Repeated Measures Analysis of Variance for the Playground Problem Variable, Records	150
6.12	Repeated Measures Analysis of Variance for the Picnic Problem Variable, Identification	151

List of Tables (Continued)

Table		Page
6.13	Repeated Measures Analysis of Variance for the Picnic Problem Variable, Measuring	152
6.14	Repeated Measures Analysis of Variance for the Picnic Problem Variable, Calculates	153
6.15	Repeated Measures Analysis of Variance for the Picnic Problem Variable, Records	154
6.16	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Identification for the Playground Problem	156
6.17	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Measuring for the Playground Problem	157
6.18	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Calculations for the Playground Problem	158
6.19	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Records for the Playground Problem	159
6.20	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Identification for the Picnic Problem	160
6.21	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Measuring for the Picnic Problem	161
6.22	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Calculations for the Picnic Problem	162
6.23	Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Records for the Picnic Problem	163
6.24	Two Factor Analyses of Covariance for Playground Problem Scores	165
6.25	Two Factor Analyses of Covariance for Picnic Problem Scores	166

Table		Page
6.26	Distribution of Pre-tes[illegible] Ratings of Identif[illegible]cation of Selected [illegible] [illegible]MES and Control Classes on the Play[illegible] Pro[illegible]em	16[illegible]
6.27	Distribution of Pre-test and Post-test Ratings of Measuring of Selected Variables for USMES and Control Classes on the Playground Problem	169
6.28	Distribution of Pre-test and Post-te[illegible] [illegible]ngs of Calculations on Selected Variables for [illegible] d Control Classes on the Playground Problem	170
6.29	Distribution of Pre-test and Post-test Ratings of Identification of Selected Variables for USMES and Control Classes on the Picnic Problem	172
6.30	Distributions of Pre-test and Post-test Ratings of Measuring of Selected Variables for USMES and Control Classes on the Picnic Problem	173
6.31	Distribution of Pre-test and Post-test Ratings of Calculations of Selected Variables for USMES and Control Classes on the Picnic Problem	174
6.32	Distributions of Pre-test and Post-test Ratings for Students' Drawings of the Playground Designs	176
7. 1	Results of Factor Analysis and Item Analysis of Part I of the Attitude Scale	190
7. 2	Results of Factor Analysis and Item Analysis of Part II of the Attitude Scale	193
7. 3	Pre-test and Post-test Means and Standard Deviations for Items on Part I of the Attitude Scale Given by Factor for USMES and Control Groups	197
7. 4	Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 1: Science Appreciation	201
7. 5	Analysis of Covariance for Attitude Scores on Part I, Factor 1: Science Appreciation	201
7. 6	Means and Standard Deviations for Attitude Scores on Part I, Factor 1: Science Appreciation	202
7. 7	Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 2: Academic Insecurity	205

List of Tables (Continued)

Table		Page
7. 8	Analysis of Covariance for Attitude Scores on Part I, Factor 2: Academic Insecurity	205
7. 9	Means and Standard Deviations for Attitude Scores on Part I, Factor 2: Academic Insecurity	206
7.10	Analysis of Covariance for Attitude Scores on Part I, Factor 3: Non-mastery	209
7.11	Means and Standard Deviations for Attitude Scores on Part I, Factor 3: Non-mastery	210
7.12	Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 4: Preference for Group Learning Activities	213
7.13	Analysis of Covariance for Attitude Scores on Part I, Factor 4: Preference for Group Learning Activities	213
7.14	Means and Standard Deviations for Attitude Scores on Part I, Factor 4: Preference for Group Learning Activities	214
7.15	Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 5: Arithmetic Enjoyment	216
7.16	Analysis of Covariance for Attitude Scores on Part I, Factor 5: Arithmetic Enjoyment	216
7.17	Means and Standard Deviations for Attitude Scores on Part I, Factor 5: Arithmetic Enjoyment	217
7.18	Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 6: Self-Directed Learning	220
7.19	Analysis of Covariance for Attitude Scores on Part I, Factor 6: Self-Directed Learning	220
7.20	Means and Standard Deviations for Attitude Scores on Part I, Factor 6: Self-Directed Learning	221
7.21	Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 7: Arithmetic Value	224
7.22	Analysis of Covariance for Attitude Scores on Part I, Factor 7: Arithmetic Value	224

List of Tables (Continued)

Table		Page
7.23	Means and Standard Deviations for Attitude Scores on Part I, Factor 7: Arithmetic Value	225
7.24	Pre-test and Post-test Means and Standard Deivations for Items on Par of the Attitude Scale Given by Factor for USMES [illegible] Groups	227
7.25	Repeated Meas[illegible]alysis of Variance for Attitude Scores on Part II, Factor 1: Need to Please--Acquiesce	230
7.26	Analysis of Covariance for Attitude Scores on Part II, Factor 1: Need to Please--Acquiesce	230
7.27	Means and Standard Deviations for Attitude Scores on Part II, Factor 1: Need to Please--Acquiesce	231
7.28	Repeated Measures Analysis of Variance for Attitude Scores on Part II, Factor 2: Divergent Thinking	234
7.29	Analysis of Covariance for Attitude Scores on Part II, Factor 2: Divergent Thinking	234
7.30	Means and Standard Deviations for Attitude Scores on Part II, Factor 2: Divergent Thinking	235
7.31	Repeated Measures Analysis of Variance for Attitude Scores on Part II, Factor 3: Academic Competition	238
7.32	Analysis of Covariance for Attitude Scores on Part II, Factor 3: Academic Competition	238
7.33	Means and Standard Deviations for Attitude Scores on Part II, Factor 3: Academic Competition	239
7.34	Repeated Measures Analysis of Variance for Attitude Scores on Part II, Factor 4: Convergent Thinking	242
7.35	Analysis of Covariance for Attitude Scores on Part II, Factor 4: Convergent Thinking	242
7.36	Means and Standard Deviations for Attitude Scores on Part II, Factor 4: Convergent Thinking	243
7.37	Summary of Significant results of Repeated Measures Analyses of Variance and of Analyses of Covariance on Eleven Attitude Factors	246

LIST OF FIGURES

Figure		Page
5. 1	Graph of Means on SAT Reading Comprehension Subtest Showing Pre-to-post Differences	118
5. 2	Graph of Means on SAT Reading Comprehension Subtest Showing Grade Level Differences	118
5. 3	Graph of Means on SAT Mathematics Computation Subtest Showing Pre-to-post Differences	119
5. 4	Graph of Means on SAT Mathematics Computation Subtest Showing Grade Level Differences	119
5. 5	Graph of Means on SAT Mathematics Application Subtest Showing Pre-to-post Differences	120
5. 6	Graph of Means on SAT Mathematics Application Subtest Showing Grade Level Differences	120
5. 7	Graph of Means on SAT Mathematics Concepts Subtest Showing Pre-to-post Differences	121
5. 8	Graph of Means on SAT Mathematics Concepts Subtest Showing Grade Level Differences	121
5. 9	Graph of Means on SAT Science Subtest Showing Pre-to-post Differences	122
5.10	Graph of Means on SAT Science Subtest Showing Grade Level Differences	122
5.11	Graph of Means on SAT Social Science Subtest Showing Pre-to-post Differences	123
5.12	Graph of Means on SAT Social Science Subtest Showing Grade Level Differences	123
7. 1	Means on Part I, Factor 1: Science Appreciation; Scale Range: 6 to 30; Scale Midpoint: 18	203
7. 2	Means on Part I, Factor 2: Academic Insecurity; Scale Range: 4 to 20; Scale Midpoint: 12	207
7. 3	Means on Part I, Factor 3: Non-mastery; Scale Range: 2 to 10; Scale Midpoint: 6	211
7. 4	Means on Part I, Factor 4: Preference for Group Learning; Scale Range: 6 to 30; Scale Midpoint: 18	215

List of Figures (Continued)

Figure		Page
7.5	Means on Part I, Factor 5: Arithmetic Enjoyment; Scale Range: 3 to 15; Scale Midpoint: 9	218
7.6	Means on Part I, Factor 6: Self Directed Learning; Scale Range: 5 to 25; Scale Midpoint: 15	222
7.7	Means on Part I, Factor 7: Arithmetic Value; Scale Range: 3 to 15; Scale Midpoint: 9	226
7.8	Means on Part II, Factor 1: Need to Please--Acquiesce; Scale Range: 6 to 30; Scale Midpoint: 18	233
7.9	Means on Part II, Factor 2: Divergent Thinking; Scale Range: 5 to 25; Scale Midpoint: 15	236
7.10	Means on Part II, Factor 3: Academic Competition; Scale Range: 2 to 10; Scale Midpoint: 6	240
7.11	Means on Part II, Factor 4: Convergent Thinking; Scale Range: 2 to 11; Scale Midpoint: 6	244

CHAPTER I

AN OVERVIEW OF THE USMES PROGRAM AND THE USMES EVALUATION PROJECT

Introduction

This document contains the report of the USMES evaluation team on our 1974-75 investigation of the cognitive and affective responses of students to USMES, an interdisciplinary, process curriculum designed to develop the problem solving abilities of elementary school students. Included in this evaluation of Unified Science and Mathematics for Elementary Schools are the results of a pre-post control group design to assess the curriculum's effects on students' basic skill development, their attitude change, and their progress in complex problem solving as measured by small group performance on real-life relevant, simulated problem tasks. The results of our interviews with USMES teachers and students, our unstructured observations during site visitations, and our field staff's documentation of USMES usage are also included in this report.

A companion document contains the results of our review of existing tests and other measurement techniques for appraising the performance of children in real complex problem solving. In the absence of any appropriate measures, the evaluation team pursued a two-fold thrust: (1) evaluation of the USMES program, and (2) new instrument development. It is the second report which documents our work on the creation of new techniques for the assessment of student's progress in complex problem solving.

The Nature of USMES: Its Philosophy and Goals

The following statement of the purposes and intentions of the USMES program is drawn from two descriptive documents prepared by their Central Staff: The USMES Guide (May, 1974) and the USMES Systems Approach to Development, Widespread Implementation and Maintenance of a Real Problem Solving Program in Elementary Schools (March, 1974).

The goal of the USMES program is the development of thirty-two interdisciplinary units engaging the student in long-range investigations of real and practical problems taken from his or her school or community environment. By responding to these problems, called "challenges," the student develops his problem-solving abilities, and does so in a manner that gives him an experiential understanding (learning-by-doing) of the problem-solving process, as well as the acquisition of its basic skills and concepts.

USMES intends to teach the cognitive skills and strategies of problem solving as a new area of learning, and not merely as a new method or a new content within an already defined area. Furthermore, this program sees itself as interdisciplinary in nature, in that its presence in the curriculum would support and facilitate the existing disciplines--mathematics, reading etc.

> [USMES] will not fulfill every cognitive and affective need;....other, more structured programs may be needed to teach the more formal aspects of the disciplines which are within the cognitive range of children in grades 1-8. (The USMES Guide, p. 9.)

USMES developers further believe that to learn the process of problem solving, the student himself must analyze the problem, choose the variables to be investigated, search out the facts, and judge the correctness of the hypotheses and conclusions. The teacher acts only as a coordinator and collaborator. This, they acknowledge, requires a new, more indirect style of teaching.

Progress toward a solution to a problem requires the combined efforts of a group of students, not just an individual student working alone. While some work may be done individually, the USMES construct provides for a division of labor and an exchange of ideas--a total group effort.

A final essential characteristic of this program is the relevance of the task. The "challenges" undertaken by the students must be both real, i.e., embody some valid aspect of school or community life rather than an invented problem imposed prepackaged by the curriculum, and practical, i.e., the student's solution may lead to the actual improvement of that situation being investigated. The problem leads to an experience of useful accomplishment in the student's life.

Issues Investigated During the 1973-74 USMES Evaluation

The 1973-74 academic year was the first period of responsibility for the evaluation of the USMES program by the evaluation Project Director. Some internal formative evaluation activities by the USMES development staff and limited assistance from evaluation consultants to the developers had preceded the award of a grant from the National Science Foundation for the independent evaluation of USMES. However, the informational needs of a variety of audiences concerned about USMES had been largely unserved to that point.

The evaluation team conferred with the National Science Foundation, which sponsors both the program development and its independent evaluation, with the developers of the USMES program and the members of their Planning Committee, with on-site users of USMES and trainers of users, and with prospective USMES users and trainers. Our preliminary conversations with these interest groups led to the identification of several critical areas for investigation during the first year of a comprehensive evaluation of the program.

Their principal concerns were addressed by the following evaluation activities:

a. documenting actual USMES usage and differentiating the student-to-student and student-to-teacher interaction patterns fostered under the USMES environment from those found in control classes;

b. investigating "proof of concept," i.e., whether or not the program is increasing students' abilities in complex problem solving;

c. comparing USMES and control students' performance on standardized tests of reading and mathematics achievement, to answer the concern of USMES teachers and principals, and of prospective USMES users, that they remain successfully accountable for the communication of basic skills;

e. studying the effectiveness of USMES teacher training efforts at the local and national levels;

f. probing to discover indirect effects of the USMES program... on students' attitudes... on teaching styles... on non-USMES teachers and students within the school... on the adminstrators of selected schools... on school scheduling... on school practices.

The 1973-74 evaluation project remained without funding until March 22, 1974. As a result, the evaluation team was unable to assume early control of the evaluation activities to train observers, designate samples, advise observers on data collection problems, and monitor the data collection process--all necessary to insure complete, usable returns. These problems resulted in a serious loss of usable data. The data losses were most damaging to an adequate assessment of student performance in basic skills and in problem solving, because the measures for these traits were time consuming or otherwise difficult to administer, and they were disruptive to the school day.

On the other hand, the schedules for site visitations, interviews, and mailing teacher questionnaires were little affected by the eight-month delay in the grant award. Consequently, we salvaged as much as possible of the originally planned student performance test data, but we also turned to and relied upon the perceptions of the teachers trained and actually involved the USMES project. The issues for investigation were broadened and, despite the funder's concern for "proof of concept," the 1973-74 project assumed the demeanor of a formative evaluation for the continuing development of the USMES program, more than the thrust of a summative evaluation of the curriculum's effects on students.

It was the judgement of the evaluation team that the information from program monitoring, the interview data, and our unstructured observations during the site visits provided very comprehensive, helpful, and illuminating information about the USMES program. We relied heavily on these kinds of data in our assessment of USMES development, implementation, and dissemination during the 1973-74 school year. The detailed evaluation report for 1973-74 is given by Shann, August, 1975. A synopsis of this report is found in Bender, August, 1975. These reports are available from the Project Director.

Focus of the 1974-75 Evaluation Project

The original proposal to the National Science Foundation for the continued evaluation of USMES during 1974-75 was broader in scope than the plan which was funded. As amended, the 1974-75 USMES Evaluation focused on student effects of the program: their progress in problem solving; their basic skill development; their attitudes toward mathematics, science, problem solving, and toward various learning activities embodied by the USMES philosophy. Teacher training, support networks for USMES users, formative program monitoring, material resource usage, and program dissemination patterns were deleted as areas for investigation under the revised evaluation plan for 1974-75.

The Project Director submitted the ammended proposal to the National Science Foundation after she had conferred simultaneously with several representatives of the Foundation's education directorate. Clearly, it was the Foundation's overriding concern for an evaluation of USMES to pursue the investigation of "proof of concept." The assessment of other student effects of the program and the description and documentation of actual program usage were also of interest to NSF.

Once the issues for investigation were specified, the evaluators selected the following indicators through which we would gather some of the required data:

a. the Picnic Problem and the Playground Problem, tests especially designed for assessing the performance of small groups on real-life relevant, simulated tasks of complex problem solving;

b. selected subtests of the Stanford Achievement Test to measure students' basic skill development in the areas of reading, mathematics, science, and social studies;

c. the Classroom Activity Analysis scales especially developed for observing student-to-student and student-to-teacher interaction patterns in USMES and control classes.

d. on-site visits for unstructured observations and interviews to determine the actual patterns of program adoption or adaptation.

f. interview schedules for on-site interviews with USMES teachers about their perceptions of the program's effects on students and on their own teaching styles;

g. interview schedules for direct interviews with USMES students about how well they liked USMES and what they thought their USMES units have taught them.

Additionally, we had hoped to locate an existing test which would have been appropriate for measuring the problem solving abilities of individual children. An exhaustive review of the problem solving literature failed to produce such a test. Having reaffirmed our thinking that new tests would have to be developed to assess younger students' performance in complex problem solving, we directed our efforts toward new instrument development in addition to the evaluation tasks.

Organization of the Evaluation Report

Following this initial overview, Chapter II includes a discussion of the methods and procedures used in this evaluation design. Actual treatments for both USMES and control sample classes are characterized in Chapter III. Interviews with USMES teachers and their students, and conversations with principals provide the bases for Chapter IV.

In Chapters V, VI, and VII, students' basic skill development, their performance in real-life relevant problem situations, and their attitude changes are reported and discussed. Our perceptions about serious problems with USMES program maintenance and program dissemination are included in Chapter VIII. The concluding chapter summarizes this report, offers specific conclusions, and makes final recommendations regarding development, implementation, maintenance, and widespread dissemination of the USMES program.

The Report on New Instrument Development

It is the judgement of the evaluation team that our work on new instrument development for problem solving can have meaningful application beyond the evaluation of the USMES program. We address that report to a wider audience whose concerns may embrace the evaluation of other curricula for elementary schools, research on child development, or theoretical development of models of problem solving. Furthermore, the evaluation work and the synthesis of new measures of problem solving are conceptually very different endeavors.

For these reasons, and because the 1974-75 USMES Evaluation report itself is already a lengthy document, we have decided to issue a separate report of our work on new instrument development for the assessment of complex problem solving by elementary school children.

CHAPTER II

METHODS AND PROCEDURES

The 1974-75 USMES Evaluation was directed toward a comprehensive investigation of proof of concept of the USMES curriculum. In addition to assessing the effects of USMES on students' performance in real, complex problem solving, the evaluation was designed to examine other student effects of the program--students' basic skill development and changes in their attitudes toward mathematics, science, and various learning strategies. Another goal of this evaluation was to document how USMES was actually being used in elementary school classes.

Wide application of a variety of data collection techniques was required to achieve these ends. The purpose of this chapter is to describe the sample selections, the instruments of data collection, and the methods of analysis which were employed for t e 1974-75 USMES evaluation.

Observer Training

One essential component of this evaluation was the training of competent, responsible field staff personnel who would serve as on-site evaluators for the test administration and for the observation of class activities in USMES and control sample classes. Observers were sought in those geographical areas where there were sufficiently large numbers of USMES-trained teachers and thus the promise of USMES activity. We prevailed upon USMES contacts (principals, teachers, district level administrators) to recommend responsible persons. In general, these "observers" were college graduates who exhibited a sincere interest in elementary school education. Some had advanced degrees; some had been teachers themselves. They were employed on a part-time basis by the evaluation project for data collection from sample classes in their areas.

The observer training took place in Boston during a three-day period in

August, 1974. Conducted by two members of the evaluation staff, the observer of personnel and their responsibilities, discussion of the time lines for data collection, review of instruments which would be used, and practice in administration or application of those forms which required special training. Seventeen observers from 15 locations were trained to work on the evaluation.

<u>Sample Selection and Procurement</u>

Some of the limitations to sampling for the previous year's evaluation persisted for the 1974-75 evaluation. Random sampling was not feasible; selections had to be limited to classes in those areas from which we had trained observers. Moreover, the evaluation staff could not have visited a great many widely scattered sites which could have resulted from random sampling. Given these constraints, purposive sampling was done to insure that USMES class selections represented a cross section of (a) grade levels, (b) USMES unit challenges, (c) socioeconomic levels, and (d) teacher experience with USMES--in a manageable number of geographic areas. The number of sample USMES classes in each of the geographic areas was proportionate to the intensity of USMES involvement expected by the program developers.

To achieve representation of a cross-section of teacher experience with USMES in the sample of USMES classes, we selected teachers from each of the following designations:

a. <u>First-trial Implementation Teachers</u>: Those who were new to USMES who were attending their first national level workshop during the Summer of 1974. There they received training and resource materials for newly developed units. Their implementation of these new units was supposed to enable the developers to assess the adequacy of the workshop training and the utility of the resource materials. By June, 1975, these teachers should have had one year's experience with USMES.

b. Former First-trial Implementation Teachers: Those who had received USMES training at a national workshop during the Summer of 1973, or during the Summer of 1972. These teachers had not been invited back to subsequent summer workshops by the USMES Central Staff. During the 1974-75 academic year, these teachers should have been experiencing their second or third year with the USMES program.

c. Development Teachers: Those who had been invited by the USMES Central Staff to attend subsequent workshops after their initial experience with USMES training. These teachers were expected to formulate and try out ideas for new USMES challenges in their classrooms. By June, 1975, they had had from two to four years' experience with USMES.

Control classes were selected to match the USMES sample classes, one for one, on the bases of (a) grade level, (b) socioeconomic level, (c) geographic areas, and (d) general character of the school program--"traditional," "non-graded," "open," etc.

While matching provided no assurance of equivalence of USMES and control groups on all characteristics but the treatment effect, the criteria used for "matching" were most salient to the indices of program effectiveness. Furthermore, random sampling was not possible.

In previous years' evaluations, control classes were chosen from non-USMES classes in the same schools as the sample USMES classes. This practice had the advantages of minimizing extraneous variance and of reducing the complexities of data collection from a larger number of schools. However, this practice also had the undesired effect of reducing the treatment difference between USMES and control classes. In some cases, much contamination of the non-USMES classes resulted from the influence of USMES teachers, students, and materials in the same building (Shann, August, 1975, pp. 37-38). For the 1974-75 evaluation, the control classes were selected from schools neighboring the sample USMES schools.

The evaluation design called for a sample of 40 USMES and 40 control classes which satisfied the criteria outlined above. In addition to sampling more heavily from among those sites which were to have the most USMES usage, we decided to select more classes from the middle school grades, because most USMES usage is observed at these grade levels.

For both USMES and control classes, selection for the evaluation sample must be distinguished from willingness on the part of teachers, principals, and other administrators to have the classes participate in the evaluation activities, and in turn, from implicit or explicit parental permission for testing of students as required by a school district or, in one case, by state law. The assurances for USMES classes were attained very readily in most cases; an USMES school, in the person of its principal or district level administrator, had agreed to participate in any evaluation activities, if sampled, before the school could send teachers to a national USMES workshop.

Procurement of appropriate control classes proved to be much more difficult, and virtually impossible in a few cases. The senior members of the evaluation team devoted considerable effort toward identifying appropriate sample classes and securing permissions for testing. Without principals, district administrators, and our field staff evaluators, our success in procuring sample control classes would have been greatly diminished.

Control classes could not be obtained for three sample USMES classes in Minnesota. Nevertheless, we did collect data from the three USMES classes. Observers who failed to meet their commitments and serious difficulties in securing permissions for testing accounted for the loss of all but the interview data from one USMES class in New Jersey and two USMES classes in Washington, D.C. It was pointless then to pursue collecting control data in the latter two locations.

Thus, toward the goal of 40 USMES and 40 control classes, we were able to collect usable pre- and post-test data on one or more measures from a total of 37 USMES classes and 34 control classes. There were no effects of sample attrition on the interview data collected by the senior evaluation staff members from all 40 USMES teachers and three students of each of those teachers. All other data were based on a maximum of 37 USMES and 34 control classes.

Characteristics of Sample Schools and Classes

The descriptions presented below are based on data from the School and Class Information Forms shown in Appendices A and B. Because we assured participating schools and teachers of anonymity, their names and any especially distinguishing characteristics are not identified.

A. Geographical Distribution

"USMES schools" are scattered throughout the country. That is, evidence of USMES usage and the presence of USMES-trained teachers can be found in many states. However, program dissemination has been most apparent in college and university towns where USMES Planning Committee Members or other USMES contacts reside, and in metropolitan Boston towns located near the offices of the Program Developers. The developers' continued efforts to disseminate the program in less affluent urban settings have met with some success in small cities, but "inner city" schools are not using USMES. The two states which have witnessed the most widespread use of USMES are California and Michigan.

The geographic distribution of sample schools and classes for the 1974-75 USMES evaluation reflected this national pattern of USMES use. Locations included in the actual sample were: San Jose, Los Gatos, Marina, Monterey, Bakersfield, [illegible], and Los Angeles, California; Boulder, Colorado; Washington, D.C. Athens, Georgia; Iowa City, Iowa; Plainfield, New Jersey; Arlington, Waltham and Watertown, Massachusetts; Lansing, East Lansing, and Sterling Heights, Michigan; Burnsville, Minnesota; and Portland, Oregon. As stated earlier in this report, the number of sample USMES classes from each location was proportionate to the intensity of USMES usage in that area, and control classes were selected from neighboring non-USMES schools in the same communities.

B. Population Densities and Socioeconomic Levels

The distributions of population densities and socioeconomic levels of the communities in which sample schools and classes were located also reflected the national picture of USMES usage. The community settings of the sample schools ranged from lightly populated but essentially suburban regions to more densely populated "suburban" districts adjacent to large cities and themselves part of the cities' metropolitan area. Some small and middle-sized cities also used the program. In fact, most of the school systems interested in trying USMES were suburban systems. Very densely populated, large urban systems had too many other needs and immediate problems which prempted attention to USMES. Nor could very sparsely populated, rural systems provide the resources and personnel to support the program.

The socioeconomic levels of these communities in which the sample schools were located also showed a constricted range. Most were middle class, albeit from lower-middle, working class areas to upper-middle, fairly wealthy suburban areas and university towns. However, no classes from inner city schools or other impoverished areas were included in the 1974-75 sample. And no truly upper class schools were included. Again, schools at these extremes did not use USMES.

C. Grade Levels

The grade levels taught by the sample USMES teachers and their controls were distributed as follows:

Grade Levels	USMES Teachers At That Grade	Control Teachers At That Grade
2	1	1
2-3	1	-
3	3	3
3-4	2	3
4	7	7
5	6	6
5-6	3	1
6	7	5
7	2	3
7-8	2	-
8	3	5
	37	34

Most of the lower grade classes in both USMES and control groups were largely or completely self-contained. Departmentalized programs for both groups were observed more frequently in grades 5 and 6, and almost exclusively in grades 7 and 8.

D. Class Size

The sample USMES classes ranged from 19 to 50 students with a mean of 29 students per class. Essentially the same variability and average for class size were found for the control group: 15 to 50 students with a mean of 28.

Teacher Characteristics

A The information below on teacher characteristics was culled from the Class Information Form shown in Appendix B.

A. Teaching Experience

The number of years of teaching experience for USMES and control teachers in the sample can be summarized as follows:

Statistic	USMES Teachers	Control Teachers
$\bar{X}$	9.0	10.6
Mdn	7.3	9.0
Mode	7	5
Range	2-27	1-39
N	37	34

Overall, the control teachers were slightly more experienced. One control teacher had nine years more teaching experience than any other teacher in the sample. With that exception, the variability in the number of years of teaching experience was similar for both groups.

B. Teachers' Specialized Training in Mathematics and Science

The USMES teachers, as a group, had virtually the same amount of specialized, formal educational training in math and science as the control teachers. Only a small minority of either group--7/37 USMES teachers, and 5/34 control teachers--had undergraduate majors or minors in math or science. Fewer still had graduate degrees with majors or minors in math or science. Both groups reported more educational background in the social sciences: eight USMES teachers majored in a social science and five more minored in one of these areas; for the control teachers, the count was eight majors and four minors in the social sciences. For most of the teachers in the sample, their training in mathematics and science was limited to undergraduate methods courses, inservice courses, and graduate educational methods courses.

Observation of USMES and Control Classes

The Classroom Activity Analysis form shown in Appendix C was developed to enable an objective accounting of the kinds of activities which USMES and control children pursued in class and the patterns of child-to-child and child-to-teacher interactions found in those classes. Observers were trained in its use, and they were directed to apply the form in the sample USMES classes three times: at the beginning, middle, and end of their USMES units. The controls for each USMES class were to be observed three times with the form, and the times were to parallel the times selected for observing the sample USMES classes. This observation technique and the results of its applications are reported in Chapter III.

These observations and other information acquired through our field staff were necessary to document the differences between the treatments which the USMES group and the control group were receiving. Without this information, one could not determine what was being evaluated. By design, the USMES approach could have resulted in as many different treatment groups as there were classes using USMES. What commonalities can be abstracted from actual program use? Classes in the control group were expected to be homogeneous only with respect to their non-use of USMES. Yet, how can one characterize the mixtures of "more traditional" math and science programs which control classes recieved? These are the issues addressed in Chapter III.

Assessment of Student Effects of USMES

The USMES project claims that, by responding to real-life, meaningful challenges taken from the local school/community environment, students will be involved in all aspects of problem solving: "definition of the problem, determination of the important factors in the problem, observation, data collection and analysis, measurement, discussion and group work, formulation and trial of suggested solutions, clarification of values, decision making, and communication of findings to others" (USMES News, October, 1975, p. 2). The program seeks to enhance these problem solving abilities of elementary school students without impairing their basic skill development. In fact, the project claims that while investigating real problems, students learn many mathematics, science, social science, and language arts skills.

The desired affective context for this approach to solving practical problems by scientific methods involves many interesting claims too: students should be better motivated to acquire needed skills when they see that their work can lead to some useful accomplishment; furthermore, "real problem solving also demands social skills like working cooperatively in small groups and accepting constructive criticism from peers without becoming upset" (USMES News, October, 1975, p. 2). The ultimate goals of USMES are directed toward the preparation of young people "to care about the world they live in, to believe they can make a difference."

Thus, our investigation of proof of concept of the USMES curriculum included objective measurement of student effects of the program in two cognitive areas and several affective areas. These were: (a) basic skill development; (b) performance on real-life relevant tasks of complex problem solving; and (c) changes in attitudes toward math, science, problem solving, group learning, self-directed learning, and other referents salient to the USMES philosophy and approach.

Measurement of basic skills was accomplished with six subtests of the Stanford Achievement Test battery, 1973 edition. Since the administration of these tests was time-consuming, disruptive to the normal school schedule, and therefore resented by many school personnel, only two of the six subtests were given to every student in the sample. For the remaining four subtest measures half of the classes received pre- and post-administrations of two of those tests; the balance of the classes received the other two tests.

The Picnic Problem and the Playground Problem, specially designed tests of small group performance on simulated life-like problem tasks, were administered to each sample class to assess their development in complex problem solving. The manuals for administration of these tests are shown in Appendices D and E; the scoring manuals are illustrated in Appendices F and G. Analysis of the observers' notes and children's audio tapes from the administration of the Picnic and Playground Problems yielded four interval level cognitive scores and four ordinal ratings of affective performance in these simulated real-life test situations. (Neither the developers nor the evaluators were entirely satisfied with the validity of these simulated problem tasks for assessing the effects of USMES, but no other existing measures even approach their utility for the evaluation. Thus, a second major thrust--new instrument development for complex problem solving--was pursued along with the 1974-75 evaluation activities.)

Assessment of students' attitude change in selected areas relevant to USMES required the development, pilot testing, and factor analysis of a two-part Likert scale of attitudes. Item sampling techniques were used during the more exploratory, instrument development pretesting. All students received the same 52-item version for the post-test.

Analysis of Student Performance Data

A pre-test, post-test control group design was pursued for the collection of student performance data in each of these three areas: basic skills, problem solving, and attitudes. The factorial design for the data analyses included two independent variables: (1) treatment, to examine differential performance of USMES versus non-USMES, or "control," students; and (2) grade level, to investigate whether any program effects are more pronounced at certain grade levels or whether the student performance data show maturational/developmental trends regardless of the treatments which the students received. The dependent variables were the measures of student performance.

In general, the data were submitted to two kinds of univariate analyses. First, two-factor repeated measures analyses of variance were conducted to determine if the treatment groups at each grade level had realized statistically significant gains on any of the measures of performance. Second, covariance analyses were used to test the hypotheses that there were no statistically significant differences between the treatment groups on the student performance measures once adjustments had been made for pre-test differences.

The evaluators are aware of the controversy which surrounds the application of analysis of covariance when the desired practice of random assignment of subjects to treatments cannot be followed. Some noted researchers, like Aiken (1970) and Cronbach and Furby (1970), state categorically that the procedures of "matching" and "statistical control of concomitant variables" should not be viewed as substitutes for random assignment of subjects to treatment conditions in the analysis of covariance, and that barring random assignment, ANCOVA should not be used. Yet, others, like Ferguson (1971, p. 288-298), do advocate its use as a procedure for the statistical control of one or more variables "uncontrolled because of practical limitations associated with the conduct of the experiment." Similarly, another authority on the nature of use of analysis of covariance suggests its application "to remove the effects of disturbing variables in fields of research in which randomized experiments are not feasible" (Cochran, 1957, p. 264). We subscribe to the position articulated by Kerlinger (1971, p. 373) on this issue: the use of analysis of covariance with intact groups is a poor alternative to random assignment, but the procedure can be somewhat helpful in the face of one of the major difficulties of educational and sociological research--the inability to set up experimental groups at will.

For some areas of the student performance data, other assumption for the analysis of covariance, and even more general requirements for the analysis of variance, could not be justified. The Stanford Achievement Test data showed evidence of treatment-slope interaction and heterogeneity of variance, and so for this data we do not report the analysis of covariance results because their validity is suspect. The ANCOVA results are included for the interval scaled problem-solving scores and for the attitude factor scores.

Further specification and discussion of the instrumentation, data collection, scoring analysis, and results for the measures of student performance are discussed in later chapters of this report. Assessment of basic skill development is treated in Chapter V; the examination of student performance in problem solving is reported in Chapter VI; and the survey of attitudinal changes is discussed in Chapter VIII.

Interviews and Other Data Collection During Site Visitations

The two senior members of the evaluation project staff visited each of the field sites included in the sample. These visits encompassed 19 USMES schools and 15 control schools in nine states and the District of Columbia.

The primary purpose of the site visitations was to interview USMES teachers and their students about their application of the USMES approach and their perceptions of its effects on students. Structured interview schedules were developed to direct our conversations with teachers and students. These forms are shown in Appendices H and I. Further discussion of the interviewing techniques and the results of our interviews with the sample of 40 USMES teachers and 120 USMES students can be found in Chapter IV.

Another purpose of our site visits was to thank the principals of both USMES and control schools in the sample for their cooperation in the evaluation project and to determine if we could correct any difficulties with the testing arrangements. We also used the occasion of these courtesy visits to solicit any comments about USMES which the principals wanted to offer. Their comments about the philosophy and substance of the USMES program are reported in Chapter IV, but their views on problems with the implementation and dissemination of the program are discussed separately in Chapter VIII.

Logistics for Data Collection

Much of the data for this evaluation were collected by the trained observers who constituted our field staff. Upon receipt of the guidelines, timelines, manuals, forms and test blanks from the evaluation office, they conducted the September-October pre-testing and April-May post-testing of basic skills, problem solving, and attitudes. They observed the sample USMES and control classes three times during the year with the Classroom Activity Analysis Forms. And, it was they who made the arrangements for the site visitations by the senior staff members. Periodic telephone communication and regular mailed correspondence between the project office and the field staff enabled us to correct many problems which otherwise may have resulted in serious losses of complete usable data.

After the pre-testing, and again after the post-testing, observers mailed completed forms and answer sheets to the evaluation office for coding, scoring, and analysis. All scoring and computers analyses were accompliched at the Boston University Computing Center.

Computer output showing the pre- and post-test scores of their students on the Stanford Achievement Test were returned to the principals of sample schools along with manuals and guidelines for the appropriate interpretation and use of these scores.

Shared Emphasis on New Instrument Development

An exhaustive literature search conducted by the evaluation staff produced no appropriate, valid measure of the problem solving skills of elementary school students. Dissatisfied with several limitations of the simulated complex problem situations offered by the Picnic and Playground Problem tests, the evaluators directed substantial efforts toward the development of two new measures of problem solving. One is a paper-and-pencil test of skills in the components of the problem solving process. It contains some items from the Sequential Test of Educational Progress-Science, Level 4 and several new contextual multiple-choice items set in USMES-like, problem scenerios. The other measure is considerably more complex in its administration, scoring, and conceptual framework. It requires carefully trained examiners to apply interview and observation techniques with individual children who are working on USMES challenges. The method yields information about the nature and sequencing of activities which children engage in during their attempts to solve "real," complex problems.

Results of the literature search and our work on the development, pilot testing and refinement of these new instruments to measure problem solving are contained in a document separate from this evaluation report.

Summary

This chapter detailed the evaluation design for this project, particularly its bases of sample selection, the characteristics of participating sample schools and classes, the methods of data collection, and the techniques of analysis applied to the resulting data.

Purposive sampling of new and experienced USMES teachers from 15 geographic areas was used to achieve a sample of USMES classes further representing a cross section of grade levels, socio-economic levels, and unit challenges. Control classes came from non-USMES schools which were located in the same or neighboring communities as the USMES schools. These control classes were selected to match the USMES sample classes, one-for-one, on the bases of grade level, socioeconomic level, geographic area, and general features of the schools' program.

Interviews were completed with all 40 USMES teachers and 120 students in the evaluation sample. However, the maximum sample size achieved for other areas of the data collection was 37 USMES classes and 34 control classes.

The sample attrition from the proposed complement of 40 USMES and 40 control teachers can be attributed to problems with two observers who did not meet their commitments to data collection, and to the very stringent requirements of one state's law for permissions for pupil testing.

The principal focus of the 1974-75 evaluation project was proof of concept of the USMES program, and several indicators besides the interview technique, were used to acquire data on the program's effects on student performance. The pre-test, post-test control group design governed data collection on students' basic skill development, their performance in problem solving, and changes in their attitudes toward math, science, problem solving and various learning activities. Six subtests from the Stanford Achievement Test battery were selected to measure basic skills. Problem solving ability was assessed with the Picnic Problem and the Playground Problem, two tests of small group performance in simulated, real-life problem situations. A Likert-type attitude scale was developed to investigate attitude change.

The student performance data were submitted to two-factor repeated measures analyses of variance to determine if the treatment groups at each grade level had realized statistically significant gains from pre-test to post-test administration on any of the measures of performance. For some of the data, covariance analyses were also used to test the hypotheses that there were no significant differences between the treatment groups' adjusted post-test performances.

A second major thrust pursued by the evaluation staff did not involve the evaluation directly. These efforts on the development of new measures of complex problem solving by elementary school children are documented in a separate report which may engage a wider audience than the present one on the USMES evaluation.

CHAPTER III

CHARACTERIZING ACTUAL TREATMENTS

In planning and excuting their research designs, many evaluators overlook a crucial component of sound practice in evaluation studies--that of describing the actual differences between the programs in the experimental and control situations, or even of certifying that the treatments do differ (Charters & Jones, 1973, p. 5).

The Special Need for Contrasting Actual Treatments for USMES and Control Groups

Meaningful evaluation of an innovative curriculum project like USMES must include a description of the way the new program is actually being used in field settings, as well as a description of the more traditional treatment groups, or "control" groups, against whose performance the success of USMES-taught classes is measured. This descriptive component continues to be especially important for the evaluation of USMES.

Unlike more structured curricula which might prescribe relatively uniform student and/or teacher activities through texts, workbooks, teacher guides, programmed instruction, etc., USMES is purported to be "an important new style of education" (USMES Central Staff, March, 1973, p. 1) designed to involve students in real problem solving. While a series of challenge units and tangible resource materials have been developed by USMES, this program, according to its developers, is more accurately portrayed as a philosophy of education than as a collection of materials. Each USMES challenge unit should evolve from the children's identification of, and action on, a problem which is real and important to them. And so by design, the USMES approach could result in as many different treatment groups as there are classes using USMES. Furthermore, we could assume that classes in the "control" group were homogeneous only with respect to their non-use of USMES. Treatments and dosages could not be manipulated or controlled by the evaluators.

The purposes of this chapter are to describe how USMES classes actually used the program, to differentiate the treatment USMES classes received from the math/science programs used in control classes, and to distinguish between the classroom activity patterns of USMES classes and those of control classes. The primary bases for these descriptions are data from the following forms: (a) School Information Forms; (b) Class Information Forms; and (c) Classroom Activity Analysis Forms. These forms are exhibited in Appendices A, B. and C, respectively. Also included is pertinent information gathered from interviews with teachers and students. The interview guides for teachers and for students are shown in Appendices H and I.

Length and Intensity of USMES Usage Experienced by Sample USMES Classes

Both empirical and logical analyses point to length and intensity of USMES usage as among the most critical factors for determining the impact of USMES on students. In response to the Program Monitoring Form used in the 1973-74 USMES Evaluation, teachers emphasized that the most successful, satisfying USMES units for both teachers and students were those pursued on an intensive basis. Many teachers commented on the necessity of not spreading an USMES unit over a long period of time. They recommended that once a unit is started, it should be done often, not merely once a week (Shann, August, 1975, p. 44ff). Our interview data corroborated this viewpoint.

The 1973-74 teacher interview data also revealed that only a very small minority of USMES teachers considered USMES a replacement for the subjects of math, science, and social science in the school's curriculum, (Shann, August, 1975, p. 40-41). The vast majority viewed the USMES program as a reinforcing supplement to regular class work in math and in social studies. (More commonly, USMES was considered a replacement for science. In the absence of USMES, several classes would not have received any science.) Naturally, teachers' optional use of USMES as a supplement to the regular curriculum results in variable application of the program. But it stands to reason that a brief and/or weak application of a treatment should not be expected to influence student performance.

The program's developers, users, and evaluators alike agree on the importance of length and intensity of USMES usage for determining the effects the program may have. Operational definitions of these factors are agreed upon and applied less readily. Within the limitations of this cross-sectional investigation of USMES, funded for a one-year interval, we have tried to characterize, for the 1974-75 sample, teacher and student experience with the program in prior years and during the 1974-75 evaluation year.

A. Teacher Experience with USMES

Those sample USMES teachers who responded to the Class Information Form reportedly had one to four years' experience with USMES as follows:

Years' Experience With USMES	Number of Teachers Reporting
1	14
2	8
3	10
4	5
Total	37

Those having only one year's experience were first trial implementation teachers trained during the Summer of 1974. The 2- and 3-year USMES teachers were either development teachers or formerly trained implementation teachers.

Those five teachers who had used USMES for all four years it has been used in schools were development teachers who received the repeated invitations from the USMES Central Staff to return to workshops and pursue ideas for new unit development in their classes.

Of course, the sample of USMES teachers represents a selection bias--their own or the developers' if not the evaluators'. One would expect that USMES teachers must be predisposed to the program's philosophy or to other incentives attached to USMES usage, since they are volunteers to the program. This orientation is probably more intense for the more experienced program users.

The total number of USMES unit challenges every used by these 37 teachers is distributed as follows:

Number of Units Ever Used	Number of Teachers Reporting That Experience
1	11
2	7
3	7
4	6
5	3
6	1
7	2
Total	37

The list of units which they used virtually exhausts the units available for implementation in Fall, 1974, and the new development units--those in preparation for implementation in Fall, 1975. The unit challenges used by the 37 sample USMES teachers during 1974-75 and in prior years are listed below:

	Number of Sample Teachers Who Completed Work on the Unit During 74-75	Number of Sample Teachers Who Used the Unit in the Past
dvertising	0	1
icycle Transportation	2	1
urglar Alarm	1	4
lassroom Design	2	0
ommunity Service	0	1
onsumer Research	1	5
escribing People	1	5
esigning for Human Proportions	0	4
esign Lab Design	1	1
ice Design	0	4
ating in School	1	1
nergy Conservation	0	2
ames for Indoor Recess	1	0
etting in Shape	1	0
etting There	2	1
rowing Plants	5	1
unch Lines	1	3
anufacturing	3	7
ass Communication	2	1
ature Trails	1	0
rientation	2	0
Pedestrian Crossings	0	2
Play Area Design	3	1
School Rules	2	0
School Supplies	3	0
School Zoo	2	4
Small Group Dynamics	0	1
Soft Drink Design	3	8
Sound in Environment	0	1
Using Fee Time After School	0	1
Ways to Learn	3	4
Weather Prediction	9	3

Some of the units listed above may have been known by various names when they were in their developmental stages. Units which were available for implementation by 1974, that were not used by USMES teachers in the evaluation sample were: Classroom Management, Community Gardening, Making Schools Safer, Planning Special Occasions, and Traffic Flow. However, these units may have been used during 1974-75 by teachers who were not in the sample.

Only one of the 37 sample USMES teachers for whom we had Class Information Forms responded "no" when asked if he/she had an interest in using USMES the following year. (Three of the 37 did not respond to that question.) The reason the one teacher gave for not wanting to become involved in USMES the following year was that she felt the logs (teacher accounts of the USMES class activities required by the developers) took "too much time to write up," and she felt "that there are times when it would not be appropriate to tell the the developers certain things."

B. Student Experience with USMES

We had intended to investigate a practical, reliable way to document the number of years in which students in the evaluation sample had experienced at least one USMES unit prior to 1974-75, and then try to characterize the intensity of that prior exposure to USMES. However, we learned from interviews with their USMES teachers during 1974-75 that for the vast majority of the students, 1974-75 provided their first experience with the USMES program. Furthermore, in a cross-sectional study using the class means as the units of analysis, more refined data on students' prior experience with USMES was of little value.

C. Intensity of USMES Usage During 1974-75

Intensity of USMES usage by the sample during the 1974-75 evaluation year is described in several ways.

The total number of units used during that school year is distributed as follows:

Number of Units Used	Number of Teachers Using That Number of Units
1	23
2	11
3	2
4	1
Total	37

Further investigation of the intensity of actual program usage during 1974-75 revealed a variable pattern of USMES application. Teachers spent anywhere from 3/4 to 5 hours per day, during 1 to 5 days per week, over 1 to 32 weeks per school year on USMES. Table 3.1 shows the means, medians, and standard deviations for measures of time spent by the sample USMES teachers on up to three USMES units during 1974-75.

Most of the USMES teachers pursued only one unit challenge during the school year. On the average, they spent an hour and a half a day, for three days each week, over 12 to 15 weeks (or one semester) during the school year on their USMES unit. But some individual classes used USMES intensively over a short period of time, while other classes received smaller doses of USMES over a great many weeks. Several other combinations of values for intensity and duration of program use were also observed.

Design Lab Facilities in the USMES Schools

The 1973-74 interviews of sample USMES teachers, their principals, and the team leaders for local USMES teacher training indicated that

TABLE 3.1

Time Spent on USMES Activities Including Design Lab Work by 37 Sample USMES Classes During 1974-75

Time Spent On		Unit 1	Unit 2	Unit 3
Weeks/Year	$\bar{X}$	12.33	11.23	15.00
	Median	15.38	9.80	11.38
	S	8.96	9.02	8.53
	N for calculation	29	13	3
Days/Week	$\bar{X}$	2.97	2.54	2.40
	Median	3.12	2.5	2.17
	S	1.24	.89	.49
	N for calculation	32	13	3
Hours/Day	$\bar{X}$	1.50	1.08	1.25
	Median	1.39	1.65	1.33
	S	.62	1.10	.43
	N for calculation	31	14	3
Total Number of Classes Which Did That USMES Unit		37	14	3

Note--The N for a given calculation is lower than the total number of classes doing a given number of USMES units because some teachers indicated a "?" or no response to the item.

many of these persons regarded the Design Lab as an essential part of the USMES program. Several administrators would not encourage, or even allow, their USMES-trained teachers to begin a unit before a Design Lab had been set up in their schools (Shann, August, 1975, pp. 166-167). Some of our respondents who observed this problem attributed it to the USMES developers' oversell of the Design Lab and its use as "bait" to schools for becoming involved with USMES. Whatever the source of this misconception about the place of the Design Lab in USMES, the program's developers have since made a concerted effort to correct the problem by clarifying the role of the Design Lab in their newer written materials and through staff members' explanations at workshops.

Also, from the 1973-74 interview of USMES teachers we learned of growing frustration and resentment among these teachers about the program because they were unable to use the highly touted Design Lab in the ways the developers described. If the facility existed in a school, typically it was not staffed on a regular basis by someone other than the USMES classroom teachers. In general, teachers were most reluctant to bring their classes to unsupervised Design Labs. They felt that by themselves they were incapable of overseeing safe activity for many small groups of children.

Questions regarding Design Lab facilities in the schools were included on the 1974-75 School Information Form to determine the existence of such facilities and the patterns of staffing them among the 19 schools composing the 1974-75 USMES sample. (No control classes came from schools where USMES was being, or had been used.) Only three USMES sample schools did not have some Design Lab facility. (In two schools the "Design Labs" were primarily industrial arts workshops and only secondarily Design Labs, but the dual function was an asset to USMES use; as industrial arts workshops the facilities were very well equipped, and they were readily available for USMES use.)

Of the 16 schools with Design Labs, 13 had a manager for these facilities. In most cases (8/13), it was a teacher who served as the Design Lab manager. Paid teacher aides were the others who filled this position. At one sight, bright, mature high school students came as volunteers to the elementary school to assist the teacher who managed the Design Lab. That teacher and his principal noted that the arrangement was very satisfactory, and very satisfying to the student assistants.

Only six of the 13 Labs with managers were staffed on a regular basis, but irregular staffing did not mean little staffing. One lab was staffed "as needed," typically for about 20 hours each week. The number of hours each week the 13 labs were staffed ranged from ½ to 30, with a mean of 19.95 and a median of 12.

The pervasive concern expressed by teachers in 1973-74 about not being able to use the Design Lab did not appear to be the case in 1974-75. Two reasons could account for this observation: the needs for Design Lab use were met sufficiently since most schools had Labs, and they were staffed by a manager; and/or the classes were doing units which by nature or by application required lesser use of the facilities. Even in one school with many classes using USMES, there was reportedly no complaint among teachers about difficulties with Design Lab use.

Exposure to USMES in Sample USMES Schools

The extent of USMES usage by other classes within the schools where the 1974-75 sample classes were located is another factor salient to characterizing the actual treatment received by the experimental group. Did the classes receive the innovative program in contrast to other classes in the schools, or did many classes in their schools use USMES? Furthermore, what was the length of the commitments which the sample schools had been making to the USMES program?

From the School Information Form, we obtained data on the number of classes in the sample USMES schools which completed at least one USMES unit during 1974-75, and the number of classes for the previous year. These numbers were distributed as follows:

School	Number of Classes Doing At Least One USMES Unit During 1973-74	During 1974-75
A	3	3
B	2	2
C	4	4
D	14	10
E	6	5
F	6	5
G	3	2
H	4	3
I	0	11
J	8	8
K	5	4
L	6	6
M	1	1
N	3	2
O	0	4
P	0	5
Q	1	1
R	0	6
S	No response	No response

Schools lettered "I, O, P, and R" were new to USMES in 1974-75. "S" was ascertained to be a three-year USMES school, and all of the remaining schools in the sample had had USMES in the school curricula for two to four years.

One point which these data bear upon which should be a concern to the USMES developers is the lack of growth of USMES program users among the teachers in the "old" USMES schools. And in several schools, the list on page shows that the problem is not one of program maintenance but of program decline since 1973-74. This occurred despite renewed investments by the developers to train new USMES teachers from many of these "old" USMES schools. More is said about these problems in Chapter VIII of this report.

Apportionment of Time Across School Subjects and Activities by USMES and Control Classes

One means of differentiating the treatments experienced by the USMES group versus the control group was to investigate how much time was spent, on the average, for various school subjects and activities by the two groups, when the experimental group used USMES and did not use USMES. The three-way comparison could also enable investigation of another important question: "Where does the time come from for USMES?"

The data for these compairsons are summarized in Table 3.2. The medians, means, and standard deviations for these time allotments are based only on the responses provided by the sample teachers on page 1 on the Class Information Form. Oftentimes teachers in self-contained classrooms are hardpressed to say how much time they spend on various subjects because they need not follow a prescribed schedule, and their plans are flexible, depending on student interests. However, more precise estimates of time expenditures for

TABLE 3.2

Average Number of Hours Each Week Reportedly Spent on Selected School Subjects by USMES and Control Classes

School Subjects	USMES Classes						Control Classes		
	When USMES Done			When USMES Not Done					
	Median	$\bar{X}$	S	Median	$\bar{X}$	S	Median	$\bar{X}$	S
Mathematics	4.75	3.77	2.12	4.82	4.51	1.64	4.82	4.27	1.46
Science	1.75	1.99	1.99	2.65	2.67	1.68	2.33	2.56	1.68
Social Science	2.50	2.42	2.08	3.17	3.33	1.90	3.52	3.35	1.49
Language Arts	4.08	4.50	3.51	4.77	4.95	3.28	4.21	4.87	3.29
Music	.71	.80	.96	.97	1.13	1.19	.51	.75	.83
Art	.92	1.01	1.15	1.01	1.35	1.26	1.00	1.23	1.28
Physical Ed.	1.71	1.97	1.46	2.36	2.41	1.29	1.65	1.84	1.35
Special Projects	0.0	1.08	1.58	0.0	1.30	.32	0.0	.86	1.55
Other	0.0	.38	1.24	0.0	.60	1.46	0.0	1.01	2.52

various subjects made by independent observers or through other monitoring were beyond the scope of this evaluation project. Thus, interpretation of the statistics in Table 3.2 must be made with caution.

As to where time comes from for USMES, all other indicators--teacher and student interviews and information from observers--pointed clearly to science time as the primary source. Teachers reported taking some additional time for USMES from selected other subjects, from recess, from special project time, by condensing instruction in all subjects, or by "just taking the time" without realizing what other area of instruction was reduced. But most frequently, USMES was viewed as a replacement for science and a supplement to other subjects covered by the interdisciplinary program.

This primary, widespread use of science time for USMES was obscured by the results shown in Table 3.2. Expecting to find a greater discrepancy between average time spent on science by the experimental group when USMES was done and when USMES was not done, we returned to the raw data for these averages from page one of the Class Information Forms themselves. It became clear that some respondents probably misunderstood the point of our question. Teachers' estimates of the time they spent per week on science when USMES was done should not have included USMES time. Yet some teachers, for whose classes we knew USMES was the only "s nce," reported that USMES time as science time in Column A, when USMES was done, and reported "0" science time in column B, when USMES was not done.

Other comparisons between time spent when USMES was done and when USMES was not done may be distorted by similar misunderstandings among respondents of by lack of clarity in the question itself. The averages suggest only tentative conclusions. Time for USMES appears to have come primarily from science, language arts, and "physical education." (We determined that some teachers in self-contained classrooms who reported borrowing time from "physical education" were in fact referring to recess time.) Interestingly enough, little time, on the average, appears to have been borrowed from regular mathematics instruction for USMES. Yet, in our interviews, most of the teachers cited a number of math skills which were developed or reinforced by their USMES units.

Differences between average times spent on various subjects by USMES classes when USMES was not done and by control classes were minimal. For a given subject or activity, the difference in average times between treatment groups was rarely more than 20 minutes per week. The variability within each treatment group was considerably larger.

The evaluators repeat the caution, however, that only tentative conclusions can be drawn from the data in Table 3.2. Inconsistencies and impossibilities were present in the data. For example, one full-time teacher of a self-contained class reported only an 8-hour week. Another's time estimates totaled a 50 hour week. Also, the summed means over subjects in Table 3.2 yield 18 hours, 22 hours, and 20½ hours for weekly instruction time on the various subjects spent respectively by USMES classes when USMES was done, by USMES classes when USMES was not done, and by control classes.

Non-USMES Curricula in USMES and Control Classes

When asked to list the names of the science, social studies, math and language arts programs and texts used by their classes, the 1973-74 sample teachers responded on the Class Information Form with a lengthy set of replies defying anything but gross categorization. To avoid the difficulty of analyzing such unstructured responses again, we posed the question about curricula in the framework of a table calling simply for check marks in appropriate cells of the table. (See question #1 on page 1 of the Class Information Form in Appendix B.)

The 1974-75 results were analyzed readily. Yet, not surprisingly, the results again yielded tremendously variable patterns of curriculum usage within each treatment group. The variation within treatment groups was as large as the variation between each group, except with respect to USMES. No control teachers were trained in USMES use and none used its materials, though (on page 3 of the Class Information Form), some control teachers reported efforts to use the USMES philosophy in their classes. (These control teachers in non-USMES schools had heard about USMES at district faculty meetings or through local USMES trainers' dissemination efforts. Other control teachers asked the observers "What is the USMES philosophy?" After a description, some responded, "Oh, I teach like that.")

The data on the use of non-USMES curricula supported a few other generalizations:

a. Only half of the USMES teachers reported using a single text for math, or science, or social science, whereas three-quarters of the control teachers did so.

b. None of the sample teachers in either treatment group reported using SAPA; 12% in each group said they used SCIS; but 30% of the USMES teachers also incorporated parts of ESS into their science program whereas only 9% of the control teachers were using ESS.

c. One third of the USMES teachers used individualized math programs for their students, whereas only 17% of the control teachers did so.

Collectively, these observations suggest that USMES teachers may have been more disposed to trying what most educators would regard as more innovative curricula.

Activities and Interactions in USMES and Control Classes

A. Expected Differences Between USMES and Control Groups

One of the premises of the USMES program is that teachers and students using USMES engage in very different teaching/learning patterns from those found in non-USMES classrooms. In the USMES mode of learning, the teacher takes on a new role--that of coordinator/collaborator--rather than the director's role typically adopted by classroom teachers. Thus, USMES students are expected to engage in active, hands-on, "learning by doing."

The "real problem" which the students tackle is supposed to provide a focus for various student activities: collecting real data; constructing measuring instruments, scale models, and test equipment; trying out suggested improvements; preparing reports or summaries of their work; presenting their findings to the proper audiences. Furthermore, the developers contend that progress toward a solution to an USMES problem requires the efforts of groups of students, not just that of an individual student working alone. By comparison, children in control classes would be expected to exhibit more passive, structured, teacher-directed, and teacher-dominated behaviors.

B. Procedures for the Observation of Student Behaviors

The Classroom Activity Analysis form shown in Appendix C was developed by Susan Rogers, a former member of the USMES Evaluation Team, to enable assessment of differences in the patterns of activities for USMES versus control classes. The categories on the form represent classes of student behaviors which could be observed in an elementary school classroom. The form underwent successive revisions and pilot-testing over a period of two years in USMES and non-USMES classrooms.

Observers were trained for their proper use of the form. Upon entering the classroom, the observer conducted seven rounds of observations. Each round could take anywhere from a few seconds, if all the children were doing the same activity, to a maximum of five minutes. To insure a uniform time sampling procedure, the time period between the start of each round was set at five minutes. During each round, the observer was to look at each child as if taking a snapshot, then tally for each student that behavior category on the form which best described what the child was doing. Lists of observable student behaviors in each category accompany the Classroom Activity Form in Appendix C.

The sampling unit for our use of this observational method of data collection was the individual classroom, either USMES or control. Our trained observers were to visit each USMES classroom in the sample three times during the course of the 1974-75 school year--at the beginning, middle, and end of an USMES unit, with the specific dates to be worked out by the individual observers and their participating teachers. Observers were urged to arrange comparable time slots within a day and within the week for observing corresponding control classes for the USMES sample. USMES classes were to be observed while USMES was going on; control classes were to be observed during math or science class periods.

There was attrition in the number of sample classes observed at each successive observation point from the beginning, to the middle, to the end of the USMES unit. Table 3.3 shows that the number of USMES classes observed during the three periods declined from 33 to 26 to 22 classes; 37 should have

TABLE 3.3

Results of 1974-75 Classroom Activity Analysis: Percentages of Observers' Tallies in 29 Student Behavior Categories for USMES and Control Groups

Observation Period	Beginning of Unit		Middle of Unit		End of Unit	
Treatment Group	USMES	Control	USMES	Control	USMES	Control
Number of Classes	(33)	(27)	(26)	(24)	(22)	(24)
Category of Student Behavior	Percentages of Tallies in Each Category					
1. Measures	1.6	0.1	1.6	0.0	1.9	0.2
2. Counts	0.1	1.1	0.5	0.1	0.3	0.4
3. Constructs/ ssembles	6.0	0.6	9.4	0.0	6.9	0.0
4. Graphs	2.2	1.0	1.1	0.0	0.9	0.0
5. Tests/ xperiments	2.3	0.6	2.7	1.5	1.0	2.5
6. Calculates	1.9	17.2	0.4	13.6	0.3	15.1
7. Records Data	1.4	1.0	2.7	0.0	2.3	1.7
8. Writes composition/illustrates	5.2	2.9	7.4	2.2	0.5	0.9
9. Writes (prestructured)	2.7	11.7	2.5	10.6	5.2	9.2
10. Reads How-to-cards; plays tapes	0.3	0.0	0.0	0.0	0.0	0.0
11. Reads (prestructured)	2.7	6.0	1.1	13.4	3.1	5.8
12. Free Reading	0.7	1.1	1.5	1.0	1.4	0.1
13. Talks to another-task	6.6	3.4	5.7	3.7	5.7	3.7
14. Talks to another-social	4.4	5.2	3.1	3.4	4.6	3.6
15. Takes part in small group discussion-task	7.6	1.5	6.1	1.3	4.6	1.8
16. Takes part in small group discussion-social	1.8	0.6	0.8	1.4	2.2	1.6
17. Gives prestructured information to teacher	0.9	2.8	0.8	1.5	1.1	1.8
18. Gives original information to teacher	3.1	1.2	3.0	0.8	3.8	1.1
19. Seeks information from teacher	2.8	3.7	1.7	2.1	2.8	2.4
20. Talks to teacher-social	0.4	0.1	0.3	0.1	0.7	0.0
21. Takes part in class discussion or presentation	5.4	1.4	5.9	1.8	3.1	2.4
22. Listen/look at child	4.3	4.6	5.5	6.2	7.5	9.5
23. Listen/look at small group	2.5	0.9	1.8	0.7	1.7	1.0
24. Listen/look at class	1.8	0.3	4.1	1.3	1.9	3.2
25. Listen/look at teacher	19.4	16.5	17.7	20.4	20.1	19.4
26. Listen/look at film or AV materials	0.1	2.4	0.0	3.5	0.2	2.1
27. Collecting materials/maintenance	2.6	2.2	3.6	1.9	3.9	1.7
28. Resting/waiting	6.4	7.0	5.5	4.6	7.9	5.4
29. Fooling around	3.0	2.9	3.3	2.7	4.6	3.2
TOTAL PERCENTAGES	100.2	100.0	99.8	99.9	100.2	99.8

been observed at each point. The trained observers reported some difficulty in making arrangements with the teachers to conduct the observations. Additionally, they noted that USMES teachers who had agreed to have their classes observed at an appointed time using USMES decided "at the last minute" to change their plans and not use USMES. When the observation of an USMES class could not be achieved, the control match was not scheduled for observation.

For those classes which were observed, the resulting data from the Classroom Activity Analysis forms were organized as follows:

a. The frequencies of the behaviors for each round in each class were tabulated and keypunched.

b. Next, for each record set containing the tallies of an observer for one visit to one classroom, the frequencies were averaged across the seven rounds.

c. For each group (USMES and control) during each observation period, the behavior frequencies averaged across rounds were then averaged for groups within periods.

d. These average frequencies were expressed as percentages of total frequencies for each group in each observation period.

C. Results of Classroom Activity Analysis

The relative frequencies of student behaviors for USMES and control groups during the three observation periods are presented in Table 3.3. Several trends and comparisons are noteworthy:

a. Teachers continued to dominate class activities. In both USMES and control groups students spent the largest percentages of class time listening to and/or looking at the teacher, as indicated by line 25 of Table 3.3. These figures from the 1974-75 analysis are similar for USMES and control classes. (In comparison to the 1973-74 classroom activity analysis, there was an increase in teacher dominated activity in the first two observation periods for USMES Groups, but the amount of observed time control classes spent looking at and listening to the teacher declined from the 1973-74 analysis to the 1974-75 analysis.)

b. Especially noteworthy are the significantly larger amounts of time which control classes spent on structured kinds of activities: calculating (line 6); writing--prestructured (line 9); and reading--prestructured (line 11). Indeed, there was a sustained emphasis in control classes on these three activities. They consumed 35% of their observed class time during the first observation period, 38% during the second, and 30% during the third. The corresponding percentages (sums for lines 6, 9 and 11) for USMES classes were only 7%, 4%, and 9% respectively.

c. USMES students spent slightly more time than control students on the following activities: measuring (line 1); graphing (line 4); testing/experimenting, in the first and second periods only (line 5); and recording data (line 7). The differences are small, but they are persistent and collectively noteworthy. They also spent substantially more time constructing and assembling--more physical, hands-on, activity (line 3). It would appear from these combined data (lines 1, 3, 4, 5, and 7) that USMES students do engage more frequently in some component activities of problem solving.

d. Even though USMES teachers were more directive than expected, for the balance of the time USMES students engaged in more creative and more self-directed activity than the control students. During the first and second observation periods they more often pursued composition writing or illustrating (line 8). Throughout, USMES students engaged more frequently in talking to another about task-related matters (line 13), in taking part in small group discussion about task-related matters (line 15), and in giving original information to the teacher (line 18). During the first observation period, USMES students spent slightly less time than control students giving pre-structured information to the teacher (line 17), and seeking information from the teacher (line 19).

These observational data tend to support the USMES developers' claims about the roles which teachers and students play in the USMES curriculum:

> To learn the process of real problem solving, the students themselves, not the teacher, determine the route they will take. The children analyze the problem, choose the variables that should be investigated, search out the facts, and judge the correctness of the hypotheses and conclusions. The teacher is an invaluable resource, a coordinator, a collaborator (USMES News, October, 1975, p. 2).

D. Comparisons with the 1973-74 Classroom Activity Analysis Results

The 1973-74 Class Activity Analysis results are reproduced in Table 3.4 of this 1974-75 evaluation year report to facilitate comparisons between the two years' data. The form and observational procedures were essentially the same for both years. However, there was an important difference to which the reader is alerted, which is essential for meaningful comparisons.

TABLE 3.4

Results of the 1973-74 Classroom Activity Analysis: Percentages of Observers' Tallies in 28 Student Behavior Categories During Fall, Winter, and Spring Observation Periods for USMES Control Classes

Observation Period	FALL		WINTER			SPRING	
Treatment Group	USMES Imp.	Control	USMES Dev.	USMES Imp.	Control	USMES Imp.	Control
Number of Classes	(10)	(10)	(14)	(7)	(6)	(5)	(3)
Category of Student Behavior	Percentages of Tallies in Each Category						
1. Measures	1.2	0.0	1.8	2.0	0.0	2.5	0.0
2. Counts	0.0	0.4	1.2	0.9	0.0	0.9	0.0
3. Constructs	7.9	0.4	11.6	0.0	0.0	2.8	0.0
4. Assembles	4.5	0.0	3.0	0.0	0.0	1.8	0.0
5. Tests/Experiments	18.4	0.8	1.3	2.1	5.7	4.0	0.0
6. Calculates	8.2	11.7	0.9	1.9	10.6	0.6	12.7
7. Records Data	6.2	1.8	1.1	6.5	0.2	2.3	0.0
8. Writes/Illustrates	0.2	0.2	5.6	2.5	4.1	3.9	0.0
9. Writes (pre-structured)	0.4	5.7	2.1	0.0	6.1	0.0	6.3
10. Reads How-To-Cards	0.2	0.0	0.8	0.0	0.0	0.1	0.0
11. Reads-Task	0.1	1.9	2.8	4.7	0.8	4.3	0.0
12. Free Reading, Writing, Drawing	0.5	0.9	3.4	2.2	1.2	1.2	0.6
13. Messes Around with Materials	1.1	1.2	1.3	0.1	0.9	1.4	0.0
14. Talks to Another-Task	2.2	1.0	2.8	3.2	3.5	4.6	3.9
15. Talks to Another-Social	3.8	4.7	1.5	3.2	7.2	4.1	8.0
16. Small Group-Task	2.3	0.3	12.4	0.0	1.4	1.4	0.0
17. Small Group-Social	1.1	1.9	0.1	0.0	0.0	0.0	0.0
18. Gives Pre-structured Info to Teacher	0.6	1.5	0.5	1.0	2.3	1.8	9.4
19. Gives Original Info to Teacher	3.3	0.8	1.3	2.2	1.7	6.1	0.0
20. Seeks Info from Teacher	2.9	3.5	2.0	1.7	2.7	2.6	3.2
21. Talks to Teacher, Social	0.2	0.4	0.3	2.3	0.0	0.1	0.0
22. Takes Part in Class Discussion, Presentation	4.9	11.2	6.7	17.1	4.4	8.7	10.7
23. Listen/Look at Child	7.3	1.9	4.8	13.2	4.2	2.7	2.3
24. Listen/Look at Small Group	1.4	7.2	1.2	2.1	1.2	0.3	0.0
25. Listen/Look at Class	2.3	2.6	7.2	3.2	0.4	0.5	1.9
26. Listen/Look at Teacher	14.1	31.0	13.2	21.0	30.1	28.4	26.3
27. Collecting Material/Maintenance	2.6	4.6	4.2	0.7	2.0	2.8	3.5
28. Resting/Waiting/Fooling Around	2.2	2.6	4.9	6.2	9.1	10.3	11.1
Total Percentages	100.1	100.2	100.0	100.0	99.8	100.2	99.9

In 1973-74 the observation periods for USMES classes were Fall, Winter, and Spring, and this seasonal distribution of observations was expected to correspond roughly with the beginning, middle, and end of USMES units. But while the developers had urged that USMES units be pursued over the school year, more frequently units were pursued for shorter periods. The variable lengths of time which the 1973-74 sample classes pursued USMES account in part for large attrition in the number of sample classes observed over the 1973-74 school year, as shown in Table 3.4. (The reader is referred to the 1973-74 evaluation report for additional information about the data base for 1973-74 activity analysis--Shann, August, 1975, Chapter III.)

The developers no longer suggest that USMES challenges be pursued for an entire school year, and any changes in the patterns of students' activities from beginning, to middle, to end of their work on the solution to an USMES challenge was of interest. Therefore, the three observation times for the 1974-75 activity analysis were changed from Fall-Winter-Spring, to beginning-middle-end of USMES unit work. In both years, observers were urged to visit each control class at approximately the same time of day and day of week as the USMES "match" for that class.

One major difference between the two years' results is that in the 1973-74 data the expected shifts in activity emphases over the course of the unit (i.e. the 1973-74 school year) were found. Students were more active and self-directed at the beginning of their units than toward the end of the year, when the USMES teacher dominated more class time. In the 1974-75 data, little variation can be seen in the patterns of activities from one observation period to another. More careful time sampling would have to be achieved to answer the questions which arise from this comparison of 1973-74 and 1974-75 activity analysis data.

In both years, virtually no time was spent by USMES students using How-to-cards (line 10 of Tables 3.3 and 3.4). Many sources of evidence support the conclusion that students are simply not using these resources.

The 1973-74 student activity data show that in the Fall, USMES students were spending a great deal of the observed class time (18.4%) testing and experimenting. However, from the 1974-75 results, this figure was only 2.3%, and it remained low throughout the unit work. This difference may be attributed to reduced use of the Design Lab in 1974-75.

The 1973-74 USMES evaluation report documents that during that school year, many persons thought the Design Lab was essential to USMES and that no one should attempt an USMES unit without access to a Design Lab (Shann, August, 1975, p. 167). The developers made a deliberate, concerted effort to correct teachers' misconceptions about the role of the Design Lab in USMES, and the developers offered new units for implementation in 1974-75 which, by nature, did not stress Design Lab use. If not done in the Design Lab, testing and experimenting may not have been going on, or these activities may not have been so obvious to observers. We offer this explanation because in 1973-74 teachers were clamoring to use the Design Lab and they voiced frustration and disappointment when they could not do so. However, teachers rarely offered comments about the Design Lab in our 1974-75 interviews with them. Furthermore, we learned directly from the 1974-75 student interviews that they used their labs infrequently, if at all.

In both years of classroom observation, the control classes spent substantially larger amounts of time than USMES classes on calculating and on prestructured reading and writing. But this difference between USMES and control groups was even more pronounced in 1974-75. It may have been that the 1974-75 control teachers were more reluctant to have their classes observed since they were from non-USMES schools and perhaps had a lesser sense that we were interested in program evaluation, and not teacher evaluation. In response, perhaps they more frequently directed their students to workbook exercises during observation periods.

In 1974-75 the USMES teachers were more directive at the beginning of the units than the 1973-74 sample USMES teachers. Our 1974-75 interview data suggest an explanation for this curious result. Many teachers reported that their instructors at Summer 1973 USMES workshops were urging them to refrain almost totally from directing their students' work on an USMES unit. (This report may not have been factual, and indeed the developers acknowledge the importance of some teacher direction with USMES. Nevertheless, those teachers' perceptions were real.) With experience in teaching USMES, those teachers expressed more and more confidence about the importance of teacher direction at certain critical points in their USMES units.

Chapter Summary

One essential component of sound evaluation practice is a description of what the experimental program is like, not just in concept but in practice, and of how it differs from the treatment applied to the control group. Without this documentation, one could not make meaningful comparisons between the performance scores of students in the experimental program and the scores of students who did not receive the innovative curriculum.

For this USMES evaluation, the description and differentiation is especially important. Treatments and dosages could not be assigned at random, manipulated, or controlled by the evaluators. USMES classes in the sample received diverse applications of the program. Some USMES classes experienced brief applications of the program throughout the entire school year, while others had their USMES time concentrated in intensive periods over a few weeks only. Many combinations of levels of intensity and duration of usage were reported by the sample USMES teachers, but, on the average, classes spent 1½ hours a day, for three days each week for 12 weeks on their USMES units.

For most USMES classes, the time for USMES came primarily from science time. The statistically "average USMES class" reportedly borrowed some additional time for USMES from other subjects, most notably from social studies and language arts. Precise data on how teachers fit USMES into their overall programs could not be obtained without continual monitoring of USMES classes during the USMES and non-USMES portions of their curricula, before or after and during periods of USMES use.

Design Lab facilities were present in all but 3 of the 19 sample USMES schools. Of the 16, 13 had recognized Design Lab managers, 6 of whom staffed their Labs on a regular basis. The managers were teachers or teacher aides. One of the teachers was assisted in the Lab by two mature high school students who enjoyed working with the younger children. Labs were staffed from ½ to 30 hours per week; the mean was 20 hours and the median was 12 hours.

The distribution of average times spent on non-USMES school subjects and activities for USMES classes when they did not pursue USMES was very similar to the distribution of average times for control classes. The measures of variability in these times for each subject or activity were very large in both the USMES group and the control group.

Similarly, both USMES and control groups represented tremendously variable patterns in the kinds of non-USMES curriculum materials and programs they used. There was as much variability within treatment groups as there was between treatment groups, except for the fact that no 1974-75 sample control classes were using or had used USMES.

Despite this variety in the nature and intensity of the treatments which the groups received, the results from the 1974-75 Class Activity Analysis indicated that there were clearly distinguishable differences observed in the kinds of activities pursued by USMES versus control students. Teachers continued to dominate class activity 16% to 20% of the time in both treatment groups. However, during the remaining observed class time, USMES students exhibited a wider repertoire of behaviors, and they spent larger amounts of time in more active, self-directed, and creative behavior than the control students. When the control students were not focusing on their teachers, they were spending much of the balance of the observed class time in very structured activities--prestructured reading, prestructured writing, and calculating--probably on worksheets or in workbooks for mathematics and/or science.

CHAPTER IV

INTERVIEWS WITH USMES TEACHERS AND THEIR STUDENTS

The interview was retained as a data collection strategy for the continued evaluation of USMES during 1974-75 because valuable insights about the program had been gained from the 1973-74 interview results.

The Interview Respondents

In 1973-74, we directed our interviews toward USMES teachers, their principals, and selected district level administrators who had authority over local USMES teacher training efforts. For the 1974-75 evaluation project, administrators were not interviewed with prepared questions since we felt that many administrators seen during 1973-74 were not sufficiently familiar with the program to have other than general comments. We hasten to add, however, that some administrators knew the program extremely well, and they offered many valuable insights about USMES training and USMES usage. Thus, for the 1974-75 site visitations, the chief administrator present in every sample school was seen by the Project Director and the Associate Director. Yet, in most instances, our visit was a courtesy call to thank the principal for his or her school's participation in the USMES evaluation program and to determine if we could correct any problem situations or misunderstandings which may have arisen about our testing program.

In those instances in which administrator wanted to discuss other issues, we did so. The administrators' comments about the USMES approach and its effects on students are contained in this chapter. Their insights about problems with program maintenance and dissemination, which are distinct from their views on the substance of the program, are presented in Chapter. VIII.

Another difference between the two years' evaluation programs was that for 1974-75, children were added to the interview schedule. This proved to be a valuable addition, for not only did it give the evaluators information about how children liked the program, and what they derived from USMES, but also it gave us a check on the teachers' perceptions of the effects of the program. Though by no means a new idea, this practice reinforced our notion that adults should consult with children to see how a program is affecting them.

USMES Teachers were retained as a major category of respondents for the 1974-75 interviews because, along with their students, they are closest to the program in practice. As a group, they are most knowledgeable about its use.

Arrangements for the Interviews

The two senior officers of the evaluation project staff conducted the interviews in a period from January through March, 1975. This time period intervened the periods of pre-testing and post-testing when the management of data collection activities consumed much attention from the evaluation project staff. Also, that time for interviewing was selected in the hope that all sample USMES teachers would have completed at least one unit with their present classes, and they may even have started a second unit. In fact, we found that approximately half of these teachers had just begun a first unit, and the other half had finished one unit. It was rare to meet a teacher who had begun a second unit by March.

Schedules for the interviews had been arranged by our field staff evaluators, so that the teachers and principals were prepared for our visits. Students were not advised about our visit, however, and they seldom knew in advance who we were or why we wanted to talk to them.

The co-operation from the schools to facilitate our conducting the interviews was outstanding. In no instance did a person with whom we requested an interview refuse to see us, or make a begrudging remark that we were imposing on his/her time. Our receptions at most schools were more than courteous; they were warm, friendly, interested, and very accomodating.

All 40 USMES teachers in the evaluation sample were interviewed. We also queried 120 students, three from each of these teachers' classes. No control teachers or control students were interviewed, but we did visit the principals of both USMES and control sample schools to thank them for co-operating in the USMES evaluation project. We extended this courtesy at the time of the site visits to any other administrators who had assisted us in securing permissions to test in the schools. The Project Director followed these visits with personal letters of thanks to these principals and other administrators. (A view held by several of our field staff evaluators was that these concerted and substantial public relations efforts helped to extend the schools' commitments to the evaluation program through the post-testing.)

Focus of the Interviews

Unlike the 1973-74 interviews, which encompassed discussion of the program's effects on students, the effectiveness of teacher training, the use of supportive materials, and other issues as well, the interviews for the

1974-75 evaluation project focused on proof of concept. Also during 1973-74, the first year of a comprehensive USMES evaluation, we felt that it would be inappropriate to use a structured interview technique based on our preconceptions of what was important to evaluate. Instead, the strategy of intensive interviewing with interview guides enabled us to ask questions about what interested us, but also to elicit from each respondent those concerns he or she considered to be of first importance in their use of the USMES program. But for the 1974-75 evaluation project, our increased level of information, our experience with the respondents, and our narrowed focus directed us to employ a more structured interview whose results were both pertinent to critical issues and easier to analyze.

Thus, the 1974-75 interviews focused on the USMES program as it effects students' cognitive and affective development and as it affects teachers' classroom behavior during USMES units and during non-USMES instruction. These issues provide the substance of this chapter.

Despite our focus on proof of concept, we concluded the interviews for teachers and principals with a general question as to whether they had any other comments about the program which our conversations hadn't touched upon. This question tended to elicit much more negative comments than our preceding, direct questions. In general, the interview respondents seemed to be expressing sincerely favorable comments about the merits of USMES approach and its effects on students--the substance of our interview questions designed to probe for proof of concept. However, most of the negative feedback promoted by our concluding question was directed toward other issues--problems with teacher training, logistical support for USMES teachers, and in turn, with program maintenance and widespread dissemination. Teachers' and administrators' comments on the latter group of issues are addressed in Chapter VIII.

The remainder of the present chapter will be divided into four sections: (a) interviews with USMES teachers; (b) interviews with USMES children; (c) conversations with principals; and (d) a synthesis of responses.

Interviews with USMES Teachers

All of the 40 USMES teachers whose classes were selected to participate in the USMES evaluation program were engaged in on-site interviews. The purpose was to evaluate the teachers' perceptions of the goals of USMES, and of the effectiveness of the USMES units in achieving these goals. The evaluators also sought to identify the teachers' perceptions of the effects USMES was having on their own teaching behavior.

The actual number of the sample of USMES teachers interviewed was 45. The extra interviews occurred when it was suggested at a school that we interview a teacher who had had extensive experience with the program, or when USMES teachers who were not selected for the sample overhead our introduction to a sample teacher for the interview and asked to be interviewed too. In the interest of good public relations, we did so.

A. The Strategy of Teacher Interviews

The interviews were conducted by a pair of evaluators; one actively queried each teacher while the other remained on the side and recorded the pertinent information. While the interviewer sought those specific perceptions indicated on the sample form (Cf. Appendix H), she did not follow the form with question-to-answer rigidity. The conversation was guided by the form, but was allowed to flow as the teacher freely related his/her own impressions of the USMES experience. Encouraging the teacher to speak openly and frankly, the interviewer fostered a non-judgmental climate. The "a" through "d" sets of responses found on the form provided support for the recorder but were not given as pre-defined alternatives to the teacher.

The observers provided an outside source for confirmation of the validity of the data acquired in these interviews. Having performed several previous classroom observational tasks both for the USMES staff and for the evaluation team, these observers were quite familiar with the classrooms in question.

B. Responses from USMES Teachers

Primary Goals-- Questions one and two inquired of the teachers' ability to clearly identify the primary purpose of the USMES program. The first question attempted to ask this without prejudicing the response:

> "What do you see as the primary goal of the USMES program?"

Most teachers readily identified the primary goal of USMES to be an increase in the student's ability to solve problems. Twenty-six teachers indicated this goal specifically and five others in equivalent terms. Thus, 69% affirmed this goal without prompting. Of these, 14 teachers also noted the secondary goal--to teach children to solve problems on their own, i.e. without the direct intervention and direction of the teacher. Only three teachers saw the USMES goal primarily as an attempt to integrate math and science in the curriculum.

The second question specifically mentioned "problem solving" and asked for affirmation of the teachers' unprompted response about the primary goal of USMES:

> "To what extend do you see real life problem solving as a goal?"

Thirty teachers saw it as "very important" and four more had already affirmed its primacy in their first response. Another four qualified their answers, but implicitly affirmed that problem solving was the intended major focus. With this prompting, then, 84% of the USMES teacher respondents expressed clear understanding of the basic goal of USMES.

Essential Characteristics-- The next set of questions, three through six, queried the teachers on their perceptions of some of the elements claimed to be essential to the USMES program. This third question asked:

> "Have the problems you've solved come about naturally, or were they contrived?"

One of the claims for the superiority of the USMES program is that their "challenges" are based on real, not artificial or "canned" problems. But a "real" problem arises naturally, and is not contrived by the instructor. In response to this question, 18 teachers identified their problems as contrived, and three more described a problem arising in a clearly contrived manner. Only 14, or 31%, based their units on real problems.

Those teachers who used contrived problems either defended their method by noting that regulated and planned teaching schedules do not ordinarily allow for spontaneous problems, or argued that even when a problem-producing event does occur, it is the teacher's introduction which turns it into a problematic frame of reference. Some teachers clearly indicated that they were specifically prepared to teach a single unit, such as "Weather," they found it interesting, and introduced it as such to their students. Yet another reason for contriving the problem was an agreement made between EDC and development teachers to try certain units. Several teachers were caught in the bind of not being able to follow "natural" problems, but being forced to continue the agreed-upon problem.

Another essential element in the USMES program is that the students be able to solve the problem they have engaged. Question four asks:

> "Was a solution to the problem found?"

Nineteen teachers responded in the affirmative, and 11 more reported that their units were still in progress. Of the eight who responded in the negative, some noted that their units--Zoo, Plants, Mass Media, Weather--were too broad in scope for the children to be able to envision them as a single problem. However, several small problems and solutions were found throughout the unit process.

Not only does the USMES philosophy require that a real problem be initiated and solved by the students themselves, but that they also experience their solution and feel that it has had some practical effect. Question five asked:

> "Were the children satisfied with their solution?"

While 22 responded in the affirmative and another 11 noted that their unit was still in process, the interviewers felt that these teachers had mentally transposed this question into "Were the children satisfied with their unit?" Also, the 25 who responded to question six:

> "Did the students feel the solution made a difference--was it implemented?"

appeared to be referring more to the various sub-problems and sub-solutions within the unit than to the unit as a whole.

Generalization of USMES Approach-- In this next set of questions, seven to nine, the evaluators wanted to know if the problem solving skills, learned through the USMES units, were learned well enough to be applied to non-USMES situations. Did these skills become part of the teachers' and students' general repertoire of teaching/learning behavior?

A strong majority (80%) of the teachers responded "yes" to the seventh question:

> "Have you used the USMES approach to solve problems for which USMES units don't exist?"

When asked for specific examples, the teachers typically referred to small, individual, often inter-personal problems, or to the indirect style of teaching fostered by USMES. Few had actually pursued the solution of problems with the potential magnitude and complexity of USMES challenges. To question eight:

> "Who raised the problem?"

the same 80% of the teachers again admitted that they had, rather than the students. Question nine:

> "Was the problem solved successfully?"

again brought an almost exclusive response of "yes."

The evaluators' intent in posing this series of questions was to probe whether USMES teachers were transferring the more indirect, USMES style of teaching to other areas of their teaching. After examining and reflecting upon their responses, we could see that the teachers were strongly affirming their support of the USMES philosophy and method because of their USMES training. Some insisted that they had always taught in this indirect approach; others felt that they had always wanted to teach in this manner, but that USMES had given them the structure and support they needed to accomplish it.

The question of transfer of the USMES method to other areas of teaching needs to be addressed in greater depth with a variety of indicators in subsequent USMES evaluations.

Manuals-- How useful are the manuals? Questions ten through thirteen were intended to investigate this area of conflict raised in the previous year's evaluation. Throughout the years of USMES implementation, teachers were not using the manuals. Among the majority who did not use this reference were some teachers who called for concise manuals with a "how-to-do-it" approach, particularly for the point of introducing a challenge.

However, another group of teachers noted that if the USMES philosophy requires student-discovered problems, and a process which supports their problem solving efforts, then of what support or value is a manual in which a specific problem is established and a structured procedure for solving that problem is set out in detail? Does the USMES manual conflict with the USMES philosophy in the eyes of most USMES teachers? And if manuals are developed which are more supportive than directive in format, will teachers use them?

After establishing the teachers' range of experience, the interviewers sought to focus on those units which the teachers had taught in their classrooms but had not been trained for in the USMES training sessions. Question eleven asked:

> "How many (units) had you seen presented before you used them?"

and twelve:

> "How did you learn to do the others?"

In response to these questions, only two teachers referred to the manuals as their resource. The others offered a great variety of alternatives, almost a different one for each person interviewed.

When they were prompted and asked specifically if the manuals had been helpful, their responses were equally divided into positive and negative comments. The following are a digest of these responses:

"flow charts are helpful"
"helpful for geometric designs"
"encouraged by other teachers failings"
"use it as a reference when I get into trouble"
"Read it before school started, but have not used it since"
"I look for ideas from the past"
"interesting to compare other situations"
"good for younger students, but not for 7th and 8th grades"
"suggests some activities"
"skill charts help anticipate skills"
"would be lost without it"
"it prejudiced my approach, so I stopped using it" (Five made similar comments.)
"I don't read them." (Five comments.)
"too technical...too deep...too wordy" (Six comments.)
"my workshop training was sufficient."

Further urged for suggestions on how the manuals might be improved, the teachers added these comments:

"they should be less directive"
"need a short summary of ideas"
"more diagrams are desired"
"make them shorter, more concise" (Several comments.)
"teacher needs only the philosophy and challenge"
"more activities at the 7th and 8th grade levels"
"more pictures"
"material on how to present a challenge" (four)
"more teacher comments on how skill sessions are used in class."

The evaluators concluded that as many as half the teachers would not use a manual of any sort for USMES; the balance wanted a concise, prescriptive reference despite its incompatibility with the USMES philosophy.

USMES in the Curriculum-- In questions fourteen through seventeen, the interviewers intended to reassess the place USMES was assuming as it found its way into the elementary school curriculum. Was it a new curriculum program, added to and supplementing the traditional subjects? Or, as a supposedly interdisciplinary curriculum, was it replacing one or more of these traditional areas? Question fourteen began:

> "Are you currently supplementing USMES with other math, science?"

Twenty-seven confirmed that they were supplementing this program with math/science, but all of these further specified that the addition had been made in the area of mathematics. Their explanation was that USMES was their science program, and therefore not supplemented by science.

But this affirmation raised an interesting paradox for the evaluators. When asked:

> "What math are they learning from the current unit?"

the teachers immediately responded with a lengthy list of mathematical application skills--addition, subtraction, multiplication, division, fractions, percent, measuring, graphing, drawing to scale, probability, the metric systen, and more--depending only on the particular USMES unit each teacher had used. When asked:

> "What science are they learning from the current unit?"

they responded hesitantly and in generalizations. USMES teaches the scientific method, but, for most of its units, not the content of science. The teachers identified its contents more readily as mathematics. Possiblv this reflects the general preference most elementary school teachers have for teaching mathematics over science. Or it possibly indicates that the set of skills defining "elementary mathematics" is better defined and agreed upon than any set of skills that might be called "elementary science." With USMES, they can appear to teach science while in reality they are teaching applied mathematics.

Question seventeen:

> "Where is the time coming from that you use for USMES?"

confirmed the response to question fourteen. Ten said "from science" and fourteen said "from science plus another area." This represents 53%. The remaining teachers offered a variety of non-patterned answers. Seven just insisted that the time needed was "simply taken;" no other subject suffered.

Changes in Behavior-- To question eighteen:

> "What is happening to children as a result of USMES?"

the teachers offered an assortment of answers, all positive. In essence, they said that their children are becoming more enthusiastic and capable students: "More inquisitive," "more verbal," "interact more," "look for facts," "better observation skills." They find that the students are learning to work together: "better class discussion," "form own groups," "respect for others," "better group inappropriate for some children. Those lacking the basic skills (reading, math) would be ill-suited for USMES. Those children who need strong structure or who are weakly motivated would also be "lost" in the program. Finally, there is the student who simply isn't peer-oriented and does not work well in groups. "No one method is appropriate for all kids," was the one comment which seemed to epitomize the others. USMES is most effective with that student who is relatively mature and more advanced in sociability and basic skills. It is least successful with those who are weak in basic skills, inner discipline, and social skills.

In question twenty, the teachers were asked:

> "Do you find that, as you use more units, you handle them differently?"

No radical changes were announced here. Each teacher who had moved into a second unit felt more comfortable with its process. They allowed the students to do more, were more confident that a solution would emerge, and responded to unplanned opportunities more readily. They became a bit more confident of their own response and teaching style, e.g., they became di-

> "Do you see a cumulative effect on children as they are exposed to more units?"

the sample was too small to offer any reliable insight.

This same kind of question was then directed to the USMES teachers themselves:

> "Has your perception of the program changed as you continue to use it?"

Their responses indicated that no radical changes had occurred. In general they were more enthusiastic. Some indicated that some direction on their part was appropriate at certain times in this basically non-directive program. Some had expected more science and math, but were more than compensated by the presence of the socialization and problem solving skills they had not expected in so high a degree.

The final question, twenty-three:

> "Do you see yourself as becoming more or less directive in the USMES units as you continue to use the program?"

C. Summary of Teacher Responses

These interviews with the teachers speak favorably for the USMES program. Occasionally their answers tended to become more general than specific; when this occurred, the highly positive attitude would have to be tempered in an objective evaluation as the evaluator recalled that these teachers had been selected and reselected by USMES, and that they themselves elected to remain and work with the program. Those with a lower, negative evaluation of their USMES experience were no longer with the program. Thus, the general evaluations are susceptible to bias. However, the specific comments, observations, and recommendations made within these interviews bear several marks of being fair and free of bias. Their accuracy was later confirmed by our observers as being for the large part valid.

USMES teachers perceive that they understand and practice the USMES teaching philosophy, with two exceptions: problems are more often contrived than discovered; and teachers are uncertain as to the amount of structure and direction they should supply. As a result of USMES use, their students seem to be developing some problem solving skills which they sometimes carry preparing them with nothing more than: "Ms. X would like to talk to you about school." The interviewers estimate that about 25% were selected deliberately on the basis of their high verbal ability, and the remaining 75% could be described as randomly selected. Clearly, the overall sample did represent students higher than average in verbal ability. Understandably, some teachers tried to prevail upon the observers to select such students for an interview situation. However, this bias toward students with higher verbal ability is also indicative of the population of largely suburban middle class students who receive the USMES treatment.

The intended numerical sample of 120 was increased slightly as four additional students were added to the group. Two were added to give information when two of the original 120 claimed not to have participated in an USMES unit; two more were mistakenly sent into the interview sessions and were allowed to remain.

A. The Strategy of Student Interviews

The student interview strategy differed from the process followed with the teachers. Interviewers related to the children one-to-one. We felt interviewer could serve both as correspondent and recorder. Fearing that tape recorders would intimidate the respondents, we avoided their use in both teacher and student interviews.

B. Responses from USMES Students

Orientation-- Items one and two were intended merely to introduce the topic and orient the student's thoughts. They also allowed the evaluators to determine if the student had been prompted for the interview.

> "Has your teacher told you why you've been asked to talk to me?"

If the student identified USMES in his answer, the interviewer broadened the focus to "other things you do in school" as well. If he did not know why he had been invited into the session, the interviewer said:

> "Well, I'd like to know what children like or don't like about school. (Indicated as Item 2.)

General Estimate of USMES-- Questions three through five were intended to explore the students' perceptions of USMES: Was he aware of it? Did he like it? Could he identify its basic directions?

Interestingly, another 50 responses indicated that mathematics was a favorite, while only 16 volunteered "science." This pattern of responses could be an indication that "science" is not identified as such when it is taught in the elementary grades. Or could it be that elementary school children simply prefer mathematics to science; mathematics is a relatively popular subject in those grades. Too, in the face of increased complexity and uncertainty, the students too may be finding greater appeal for some activities with immediate answers, simplicity, and closure.

Question four was limited to those students who did not mention USMES directly or indirectly in response to question three. The interviewers prompted them by calling to mind each of the specific USMES activities found in a unit and then asking for their response. Did they enjoy it, and if not, why not?

Here the interviewers became aware of the language difficulty between ordinary adult terminology about USMES on the one hand, and that of the child on the other. The interviewers had been briefed by the observers and were informed on the type of USMES activity experienced by each student.

as such, and most understood "problem solving" to mean solving mathematic problems or talking about personal problems. As noted above, "math" and each of its sub-categories are readily identified. The responses to question four were not always informative because the responses may indicate more the ability of the student to identify these items than they do a report on student preference toward USMES activities.

Question five initiated an open-ended description by the child of the USMES activity he/she experienced. Out of this discussion, the interviewer was to judge whether or not (a) the child saw that his individual work was a part of a group effort and (b) whether or not he saw the individual activities forming a series which moved toward the solution of an overall problem. While 84 affirmed the first topic and 73 affirmed the second, the interviewers again felt that these numbers did not truly reflect the students' actual perceptions. Their estimate is that 60% of the students saw themselves working in groups in a group effort, while 40% were conscious only of their individual activity. Further, of the total group of respondents, about 40% could not see that their individual activities were directed

Group Work-- Group work rather than individual work is the USMES method of learning to solve a problem. Do the students portray this aspect of the program? Question six asks:

> "When you work on USMES, do you work alone or with others?"

Twelve responded "alone" and 88 (72%) said "with others." The remaining 24 were mixed, sometimes one and sometimes the other--a response which could be compatible with the USMES format. An "others" answer, it should be noted, could indicate either the desired group work, or the traditional classroom format.

Question seven asked:

> "When you work with others, do you always work with the same people or with lots of different people?"

Fifty-two responded "with the same people," or "just with my friends," and 50 "with different people." To question eight:

> "Who chooses the people you work with?"

32 said the "teacher," 46 said "each child," and only six "a leader." Some of the remaining 38 combined their answers to include both teacher and

multiplication, fractions, and percentages; and about a dozen particular descriptive responses such as "how to find area," "how to take a poll," etc. Nine noted that they were learning the metric system.

Under "science," 60 could make no response at all. Those who could noted specific situations: "how to grow plants," "how to electrify something," "how to use a humidity chart," "how to use paint on various surfaces." Overall, there was a paucity of response to "science." Science does not have easily identified sub-categories like "subtraction" for these students, but a lower profile in the curriculum is also indicated.

To the question:

> "What kinds of things did you learn about solving problems?"

the students responded, for the most part, in terms of socialization skills they had acquired. "If you work together, its easier," "everyone has ideas--you choose the best one," "sit down and talk it over," "how to solve arguments," "you can't give up." Forty-four could not identify any problem-solving skills they had learned.

Question ten, also in three parts, asked the students to evaluate their

The students were evenly divided in their ability to see ways in which their USMES training could relate to other classroom uses--66 saw no further use; 66 infer its transferrability. Some of these areas of use were identified as subject areas--math, science, measurement--and others as socialization skills: "I feel more comfortable talking in front of groups;" "problem solving skills can be used in all areas;" perseverance, how to organize work, and how to collect information. Again, language was a problem. The younger children in particular could not grasp the intent of this question. And if a student saw USMES as a series of activities and not the solution of a problem, the question had no meaning.

Practical Problems-- "Real" problem solving, in the USMES philosophy, requires that the problem be practical and relate to the students' immediate situation. Questions eleven through thirteen explored this USMES quality through the child's perceptions. Question eleven asked:

> "Do you really think that (name problem) needed to be changed?"

Seventy-seven responded "yes" and only 16 "no." In general, the 16 simply couldn't see a problem which needed solving. For the balance of the respon-

Eighty-six of the children were of the conviction that children can truly solve problems, and only four totally deferred this power to adults. However, the remaining students added interesting qualifications: "both work together," "it takes the teacher to steer the students in the right direction," "if the teacher agrees," "its good to give kids a chance, but a grown-up should make certain the solution is o.k. before its used." Even in the indir approach of USMES, many of the students recognized a healthy interdependence between student and teacher. A different wording of the question may have produced a larger recognition of the interdependency.

<u>Design Lab</u>-- In the previous year, the Design Lab was a highly stressed element in the USMES program--so much so, according to the evaluators, that many hesitated to begin using the USMES program until they had acquired access to this Lab. Actually, the Design Lab was not intended to be as essential to USMES teaching as the teachers had perceived it to be. Questions fourteen, fifteen, and sixteen intended to follow up on this issue through the perceptions of the students: how important is the Design Lab in this year's program, and how is it being used?

"Do you use it as often as you like?"

Thirty-two responded "yes," 40 "no," and 15 made other comments somewhere between these two responses.

Clearly, the Labs, with further teacher experience in USMES, were not perceived to be as essential as in the previous year. They require continual supervision and some funding. Neither of these resources were as plentiful during this 1974-75 academic year.

How-to-Cards-- Finally, the evaluators wanted some student perceptions on the "how-to-cards" and their mode of use during this year. The type of card these opinions reflect were those used prior to the revisions most recently made by the USMES developers. Question seventeen asked:

"Have you ever seen a "How-to-Card?"

Thirty-five said "yes" and 85 "no." Fully 70% could not identify them. Those who could were further asked:

"Do you know how to use them?"

Twenty again affirmed "yes," and 11 could identify them but did not know how to use them. Finally:

"how-to-card" whereas individuals will not. Even when this group activity is pursued and students learn to use the cards, they still prefer to ask someone than go to the cards.

C. Summary of Student Responses

The interviews of students indicated that almost all of them enjoyed the USMES program, and that slightly over half were able to recognize their USMES units as problem-solving activity. The remainder perceived their USMES work as a single activity or group of activities, but with no problematic framework. Some of these students could not relate their activities to anything else because, in fact, neither students nor teacher had identified a problem. In other cases, the students were not able to perceive the problem because the unit did not lend itself to the identification of an overriding problem toward which the unit's activities could be directed.

The mathematics content of the various USMES units was readily identified by the students. Neither the science content nor scientific method was.

Those students who could sense the problem-solving focus of their units saw that their solutions could be employed and could actually effect change.

Conversations with Principals

There was no prescribed interview schedule for administrators. We met with every USMES principal and with every control principal to thank each one for his/her cooperation and to ask about difficulties encountered with the evaluation program.

The only persistent problem for both USMES and control principals involved the amount of time required for the testing and the resultant disruption which occurred, most notably for the SAT's. The time demands were resented by many of these persons. Disruptions to normal schedules were most severe in classes which were not self-contained, and especially in the seventh- and eighth-grade classes which faced rigid schedules for the change of classes. There simply was not enough time in their regular class periods to administer two SAT subtests; either four periods had to be used--one for each of four subtests--or elaborate changes had to be made which affected many more students.

Beyond this general complaint, the number and kinds of comments which administrators wanted to offer were related to their knowledge of the program. The largest group of administrators--half of the USMES principals and almost all of the control principals--had little to say. The USMES principals knew that the program was being used by some of their teachers; they were only generally aware of the intent of the program; they knew that our field staff evaluators were observing classes and administering tests at various times.

The control principals were also aware of the observers and the testing Some of the control principals expressed a desire to know more about the program, although all felt that their schools had a sufficient number of new programs and activities at the present time. There were no additional complaints from either these USMES and control principals. Everything seemed fine to them.

A smaller group of USMES principals and one control principal (who had previously been an USMES principal) were very knowledgeable about USMES. They had been to USMES workshops; a few had even taught workshops occasionally. They were interested in the program, and they kept themselves informed of the USMES activities and the units' progress in the building. They tended to be quite happy with the program.

A few of these principals mentioned that the lack of science content on the part of their teachers was a real hindrance in their dealing with USMES. (They attributed the same problem to SCIS usage.) Two principals arranged schedules so that science consultants could teach the science-oriented units, because they felt their teachers couldn't handle them. A few of these principals expressed dissatisfaction with the workships, both national and local, because they felt they were not preparing the teachers to use the program.

Some of these principals pointed to the problem of trying to keep teachers in the program, and of recruiting new teachers. They acknowledged the danger of the program's simply phasing itself out. But by-and-large, this group of principals were supportive of USMES. They felt the program was influencing student and teacher behavior, was accepted by parents, and enjoyed by students.

Another small but very persuasive group consisted of some of those principals who had been associated with the program for the longest period of time, and had been the most supportive of the program. They were very knowledgeable about the USMES philosophy and approach, and they tried to keep informed of the program's development. They were--to put it succinctly--

the kind of principal the program needs. From these principals emerged the greatest amount of dissatisfaction, but not with the conceptual framework of the program, because that's what attracted them to USMES in the first place. Their dissatisfaction stemmed from the way the program was being developed and implemented, the training models, and the selection of personnel for all these phases. Because of the importance of these issues and the fact that they came from both principals and teachers, we have devoted as separate chapter to the presentation and analysis of these comments.

A Synthesis of Responses

A sample of 40 USMES teachers was selected to be interviewed by the two senior members of the evaluation staff during the months of January, February, and March, 1975. For each USMES teacher interviewed three of his/her students were also interviewed. In addition, the evaluation staff members spoke informally with the administrators in the schools of both the USMES and control teachers.

The interviews focused on the effects of USMES or student and teacher behavior. Other issues arose in the course of the interviews with teachers and administrators, but these are discussed in Chapter VIII, since they do not relate directly to the questions in the interview schedules. (It is necessary to keep this in mind, as the material in the two chapters may appear contradictory. Although the data in both chapters were derived from the same interviews, the issues discussed in each chapter differ.)

Most of the points in the foregoing chapter came from pairs of sources: teachers and students; teachers and administrators. This built-in system of checks helped to establish the validity of the information. Another source of information, our discussions with the observer at each site, lent further credence. The issues are summarized with the objective of noting major points of agreement and disagreement from all sources.

There was no disagreement on the subject of children's enjoyment of USMES. Children did enjoy USMES and they looked forward to using it. All agreed that each child derived something from the program: increased knowledge in content areas, or ability to solve problems, or socialization skills, or increased feeling of self-worth (because every child could be successful in some aspect of USMES). What each child derived from USMES appeared to be a function of the teacher, the challenge, and the child.

The philosophy of USMES received complete support from teachers and administrators. There was not a single instance of anyone in either group questioning the value of a problem-solving approach in education. Since these largely self-selected USMES users favored a real problem-solving approach, it follows that they also favored an integrated approach to teaching the disciplines, in order to solve the problems. And in theory, they did. But in practice, there were difficulties. Departmentalized programs, rigid time schedules, and teachers with limited content background (especially science) made the integrated approach difficult.

The nature of the USMES challenge was another factor which made the problem-solving approach and the integration of the disciplines difficult to implement. Some challenges simply did not lend themselves to a problem solving approach. Very often, the challenge was not perceived as a problem by the children, who simply saw what they did as a series of unrelated activities. In some instances, even the teacher did not perceive the USMES unit as a problem.

Administrators and teachers supported this perception by asking whether USMES was teaching problem solving or was just a series of activities, often seen as "gimmicks." For those students and teachers who saw a challenge as a problem, there was some feeling that the method of solution was generalizing to other areas.

While USMES appeared to be teaching new skills, it was seen mainly as reinforcing old learning. Teachers and students had no difficulty identifying the specific aspects of mathematics being learned, but neither students nor teachers could identify very much science involved in the program. Other content areas, e.g., language arts, social science, were identified by some teachers as being heavily involved in the program.

Although the content emphasis was a function of the particular challenge, it was also a function of a specific teacher's likes and dislikes. Teachers still tended to stress those areas which interested them or which they felt most comfortable teaching. And so, they tended also to choose those challenges with which they felt most comfortable. As a result, those units which emphasized the social science contents of mathematical applications were most often used.

Teachers continued to learn to use the program through workshops or by word-of-mouth rather than by using the manual. Other materials developed specifically for USMES e.g., how-to-cards, technical papers were also getting minimal usage, both by students and teachers. Even the Design Lab usage declined noticeably over the last year.

All-in-all, however, the interviews revealed that USMES appeared to be fulfilling some of its promises. There were indications that children felt capable of dealing with their environment, and that teachers, through less directive teaching, were encouraging children to solve their own problems.

directive teaching, were encouraging children to solve their own problems. USMES seemed to be changing the behavior of both teachers and students in what the developers could view as a positive way.

CHAPTER V

THE EFFECTS OF USMES ON STUDENTS' BASIC SKILL DEVELOPMENT

Background for This Assessment

The paramount goal of the USMES program is the enhancement of students' abilities in real problem solving. However, two important criteria for the selection of an USMES unit challenge have been that the problem has the potential for a substantial acquisition of facts and scientific concepts and also for mathematical structuring appropriate to the age level (USMES, EDC; September, 1972, p. 5). Additionally, the USMES developers have suggested that the group communication required for the solution of USMES challenges would enhance language arts skills essential to improvement in other cognitive areas.

One aspect of the original conception of USMES was that of an integrated mathematics, science, and social science program. Inherent in this conception was the notion that these disciplines, particularly mathematics and science, could be taught not just as discrete skills but by an integrated problem solving approach. In their concern to remain successfully accountable for the communication of basic skills, USMES teachers and principals, as well as prospective USMES users, asked: "Will students' scores in the basic skills decrease if these subjects are no longer taught as separate content areas?"

Over the past few years, the empahsis of USMES has been modified gradually. USMES no longer claims to offer a replacement for the study of math, science, and social science as discrete disciplines in the curriculum. Indeed, the vast majority of USMES teachers interviewed during 1973-74 accepted

USMES as a supplement to regular class work, especially in math and social studies. (More frequently, USMES was the replacement for other science instruction.) Teachers felt that math skills in particular needed to be taught directly, and then they could be reinforced by the problem solving activities of USMES (Shann, August, 1975, p. 41).

With the shifting claims about the role of USMES in the school's program came a new version of the perennial question about accountability for basic skill development: "Since time spent on USMES may detract from time spent on instruction in basic skills, will students in USMES classes fall behind students in control classes in basic skill development?" It is this general question which the evaluators address in this chapter of the report

Use of the Stanford Achievement Tests

Previous evaluations of USMES have shown no difference in the rates of basic skill development for USMES and control groups, as measured by the Reading Comprehension and Mathematics Computation subtests of the Stanford Achievement Test Series (Shann, 1975; Shapiro, 1974). However, the usable data bases were small, and were not truly representative of the wider geographical and socio-economic distributions of USMES users. In the interest of continuity, the SAT was selected for the 1974-75 evaluation program, and the previously administered reading and math subtests were included in the group of subtests administered to sample classes in 1974-75.

One of the criteria for the original selection of the Stanford battery was that this series was used more widely than other test batteries under consideration. Its selection would maximize the probability that at least some sample schools would be using it, and the requests for additional testing time for the USMES evaluation could be minimized.

Very few schools in the 1974-75 sample used the SAT as part of their school testing program. Many principals and teachers, particularly those in the control sample, voiced objection to the time demands for the SAT administration. However, most of the sample schools did agree to cooperate in giving the SAT's for this USMES evaluation.

One notable exception was found in an urban school district where the earlier, 1964 edition of the SAT series was being used. Harcourt-Brace, the SAT publishers, had not formulated equations for transforming scores from the 1964 versions of the SAT to their equivalents from the 1973 edition, and it was understandable that the school district would not add to its already extensive testing schedule. The district's testing director has provided the evaluation staff with SAT data from the 1964 version of the SAT for the three sample classes and their controls in his district. However, the data have had to be analyzed separately.

A few teachers who were selected for the evaluation sample and who cooperated in other aspects of our testing program refused to allow administration of the SAT's. In two geographical areas, which encompassed a total of 10 sample classes (USMES and control), the limitations to testing were severe, and the procedures required for securing parental permission could not be fulfilled by the evaluators working at a distance. Pre-test data without any post-test results was the case for a few additional classes. Scores for two classes on two subtests had to be deleted from the analysis because the wrong subtests were administered at post-test time. The number of sample classes with both pre-test and post-test data on a given subtest is shown in Table 5.1.

TABLE 5.1

Number of Sample Classes with Pre-test and Post-test Scores Used in the Analysis of Stanford Achievement Test Data

SAT Subtest	Treatment	Grade Levels 2-4	5-6	7-8	Total
Reading Comprehension	USMES	11	13	7	31
	Control	11	12	6	29
Math Computation	USMES	10	13	7	30
	Control	10	12	6	28
Math Application	USMES	5	5	2	12
	Control	4	3	3	10
Math Concepts	USMES	5	7	3	15
	Control	6	7	3	16
Science	USMES	5	6	4	15
	Control	6	5	3	14
Social Science	USMES	5	6	2	13
	Control	4	5	3	12

Six subtests of the SAT's were given. As in prior years' evaluations, Reading Comprehension and Math Computation were administered to all sample classes in which SAT's could be given. Four other SAT subtests whose content related more directly to the focus of the USMES program were introduced in the 1974-75 USMES evaluation plan. These were the subtests of Math Application, Math Concepts, Science and Social Science. To reduce the demands for testing time, one-half of the USMES classes, and their controls received the Mathematics Concepts and Science subtests, while the other half of the sample received the Mathematics Application and Social Science subtests. The sample was divided to ensure that all grade levels would be represented by all distributions of subtest scores.

Procedures for Test Administration and Scoring

Pre-test and post-test administration of the SAT's was accomplished in the Fall (late September and early October) and in the Spring (during May) for both USMES and control classes. The tests were given by our trained field staff. Classroom teachers were encouraged to remain in their classrooms during the administration of these tests.

Table 5.2 lists the form of the test given to each grade, and to those sample classes encompassing combinations of grades. These designations were based on information contained in the publisher's test catalogue (Test Department, Harcourt, Brace, Jovanovich, Inc., 1973, p. 10).

The sample classes with more than one grade represented were found in "non-graded" schools. From consultation with the teachers for those sample classes we determined that most of the combinations were based on ability groupings; still the teachers and students could identify each child's placement by using the traditional grade level designation. The level

TABLE 5.2

Form of SAT for Pre-test and Post-test by Single Grade and Combinations of Grades

Grade	Pre-test	Post-test
2	Primary I	Primary II
3	Primary II	Primary III
4	Primary III	Intermediate I
5	Intermediate I	Intermediate II
6	Intermediate II	Intermediate II
7	Advanced	Advanced
8	Advanced	Advanced
Combined Grade	**Pre-test**	**Post-test**
2,3	Primary II	Primary II
3,4	Primary III	Primary III
4,5,6	Intermediate I	Intermediate II
6,7	Intermediate I & II	Advanced
7,8	Advanced	Advanced

of the SAT administered in such classes was the level which was appropriate for the average of the given combination of grades. It was not feasible to administer more than one level of single classroom because testing times varied across level directions differed as well.

Students who received Primary I, II, or III level SAT booklets responded in the booklets, and their answers were transferred to Digitek answer sheets for optical scanning. The older student got Intermediate I, Intermediate II, or Advanced level tests, responded directly on the machine scorable answer sheets.

Test scoring was accomplished at the Boston University Computing Center. The scoring programs yielded, for each student, a raw score and its corresponding scaled score on each of the subtests he took. The calculation of the scaled scores is described in the following paragraph taken from the Norms Booklet, Form A, of the Stanford Achievement Test (1973, p. 13):

> Scaled scores on the Stanford Achievement Test were obtained through a computerized application of the Thurstone absolute scaling procedure. This resulted in the development of a system of inter-battery standard scores which permitted the translation of raw scores at each level to standard scores with comparability across levels for a test area. The scale values were derived by setting the median raw scores of grade 3 and grade 8 in the Fall standardization equal to 132 and 182 respectively.

Data Analysis

The scaled scores, rather than the raw scores, were used in all analysis. The analyses were computed with packaged statistical programs at the Boston University Computing Center.

As expected, there were in most classrooms some instances of loss of data from pre-test to post-test due to both student absenteeism and to the mobility of students. However, unlike the 1973-74 evaluation when the usable results were so limited that studen ather than classroom had to be used as the unit of analysis, the 1974-, evaluation could employ the classroom, which had been the sampling unit, as the more appropriate unit of analysis.

Of course, treatment group was one factor which was of interest as an independent variable in the analysis of SAT scores. Its levels included "USMES" classes, which received some form of the USMES treatment as described in Chapter III, and "control" classes, i.e., those non-USMES sample classes which received their regular school program. Grade was also used as an independent variable in the analysis of SAT scores, because performance in the basic skills should be expected to increase over grade levels. Blocking on grades was achieved with the following groupings: grades 2-4; grades 5-6; and grades 7-8. These particular groupings were selected because they yielded reasonably large cell n's for an approximately proportional design. None of the combination grade sample classes in the "non-graded" schools for which we could obtain SAT data included students in a class which overlapped our grade blocks, i.e., we had no classes of both fourth- and fifth-graders, nor of both sixth- and seventh-graders.

The data for each subtest were submitted to a repeated measures analysis of variance to investigate whether classes from either treatment group realized statistically significant gains in any of the six content areas from Fall to Spring. This method of analysis also enabled us to examine, for each test, whether USMES classes differed from control classes, and whether there were significant differences among grade levels, as grouped for this analysis. Several interaction effects could also be investigated.

The data were also submitted to analysis of covariance to test whether post-test scores between USMES and control groups differ significantly once adequate statistical [illegible]ances were made for pre-test differences between the treatment gr[illegible] [illegible]wever, these ANCOVA results are not presented here because the validity of those results are suspect.

The covariance procedure assumes that the correct form of regression equation has been fitted. We had anticipated that linear regressions would be appropriate however, our continued investigations of the SAT data suggest that curvilinear equations may give a better fit for some SAT subtest data, more so for the control group than the USMES group.

Another assumption required for the valid use of covariance is that error effects have a common variance. The SAT data do exhibit departures from homogeneity of variance. Since this assumption is the natural extension of one required for an analysis of variance, the problem of heterogeneity affects the inferences one can draw even from the repeated measures analysis.

Questionable scaling procedures also affect both sets of analyses. We are concerned that the formulae given by the test developers really have not achieved comparability of translated, scaled scores across test levels for a given test area.

The more serious problems with the analyses of covariance disuaded us from presenting those results here, but indeed the problems render suspect even the results of the repeated measures analyses, which we do present here.

Therefore, we urge the reader to examine too our presentations of the SAT data analyses in their simplest forms: the tables of means and standard deviations by treatment groups and grade levels, and the graphs of the means for each subtest.

Before observing the actual results of the analysis of SAT data, one should consider what results to expect. Significant increases in basis skills scores from the lower to higher elementary grade levels for both treatment groups would be consistent with expected growth patterns for these areas o. [illegible] Further one would hope to find significant growth from pre-test to post-test administration for both treatment groups within each grade level. Those concerned that the USMES program does not attempt to increase problem solving abilities at the expense of basic skill development would hope to find no significant differences between treatment groups.

The actual results of the analysis of SAT data are presented below by subtest:

A. <u>Reading Comprehension</u>

The Reading Comprehension subtest measures reading comprehensional tasks varying from simple recognition to making inferences from several related sentences in varying content areas.

The test questions sample the following skills:

- Comprehension of global meaning.
- Comprehension of the meaning of detailed information.
- Comprehension of implied meaning.
- Use of context for word and paragraph meanings.
- Drawing inferences from what has been said.

The difficulty of the items and the length of the selected paragraphs increases from the Primary I through the Advanced Forms.

Results of the two-factor repeated measures analysis of variance for Reading Comprehension subtest scores are shown in Table 5.3. The means and standard deviations for treatment groups at each grade level are given in Table 5.4

There were no significant differences in Reading Comprehension scores between treatment groups. As expected, there was a significant overall difference ($p < .01$) among grade levels, with the higher grades scoring higher on this subtest. Further, there were significant differences ($p < .01$) between pre-test and post-test administration. However, the size of these differences was dependent on grade level, as indicated by the significant interaction effect ($p < .05$) for grade by test administration.

Examination of the means in Table 5.4 shows that the gains made by USMES and control groups were very similar. The means table also reveals the nature of the significant interaction between test administration and grade level. The classes at grades 2-4 gained approximately 11 points; the classes in grades 5 and 6 averaged about a 7-point gain; those in grades 7 and 8 gained about 1 point from pre-test to post-test administration. This result may be the function of a ceiling effect of the test. More likely, the interaction effect simply mirrors the growth curve for gains in reading comprehension ability.

B. Mathematics Computation

All forms of the Mathematics Computation subtest of the SAT measure skills in the four basic operations - addition, subtraction, multiplication and division. The numbers used in the problems become larger as one progresses from the Primary I test through the Advanced test. In addition to basic computation, there are questions requiring knowledge of

TABLE 5.3

Repeated Measures Analysis of Variance
for SAT Reading Comprehension

Source	df	Sum of Squares	Mean Square	F-Ratio
Between Ss	59	37276.00	631.80	
Grade (G)	2	25406.00	12703.00	63.269**
Treatment (T)	1	163.00	163.00	.812
T x G	2	865.00	432.00	2.154
Error	54	10842.00	200.78	
Within Ss	60	4709.00	78.48	
Tests (A)	1	1684.00	1684.00	35.633**
G x A	2	471.00	235.00	4.983*
T x A	1	1.00	1.00	0.021
T x G x A	2	1.00	0.50	0.011
Error	54	2552.00	47.26	
Total	119	41985.00	352.82	

*$p < .05$

**$p < .01$

TABLE 5.4

Means[a] and Standard Deviations for
SAT Reading Comprehension

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	11	13	7	31
	Pre	137.25 (12.06)	165.50 (14.44)	183.50 ((6.93)	159.54
	Post	148.85 (5.56)	172.79 (14.39)	184.89 (5.96)	167.03
Control	N	11	12	6	29
	Pre	138.08 (11.30)	167.46 (12.27)	171.69 (9.90)	157.19
	Post	149.80 (8.47)	174.62 (11.73)	172.13 (10.10)	164.69
Total for Grades	N	22	25	13	60
	Pre	137.67	166.44	178.05	
	Post	149.33	173.67	179.00	

[a]Class means were used as the unit of analysis.

greater-than and less-than relationships, common and decimal fractions, percent, average, exponents, simplification of expressions, and graphing.

The results of the repeated measures analysis of variance for this Mathematics Computation subtest scores are reported in Table 5.5. Table 5.6 contains the means and standard deviations for treatment groups at each grade level on this subtest.

Again, grade differences were statistically significant grades achieving higher scores on math computation, as expected. Overall pre-to-post differences were also significant, but these had to be qualified by grade level. Inspection of the means table reveals that both treatment groups at all grade levels realized gains in math computation subtest scores, except the control group at the seventh- and eighth-grade level, which witnessed a slight decline in performance. However, both treatment groups achieved lesser increases from pre-test to post-test as the grade level increased. This result is consistent with what one should expect, given the growth curve for many areas of basic skill development.

C. Mathematics Application

This test is designed to assess the student's ability to compute in order to solve problems which occur in life situations. The pupil is required to analyze the problem and utilize his knowledge of mathematics properly. At the primary level, the problems are read to the pupil in an attempt to separate his reading ability from his mathematical competence.

In Table 5.7, the results of the repeated measures analysis of Mathematics Application scores are given, while Table 5.8 contains the means and standard deviations for this subtest.

TABLE 5.5

Repeated Measures Analysis of Variance for
SAT Mathematics Computation

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	57	31917.24	559.95	
Grade(G)	2	24567.63	12283.81	94.406**
Treatment(T)	1	70.60	70.60	0.543
T x G	2	512.93	256.47	1.971
Error	52	6766.08	130.12	
Within Ss	58	2363.07	40.74	
Tests(A)	1	980.91	980.91	72.115**
G x A	2	640.83	320.42	23.556**
T x A	1	3.51	3.51	0.258
T x G x A	2	30.51	15.25	1.121
Error	52	707.31	13.60	
Total	115	34280.32		

*$p < .05$

**$p < .01$

TABLE 5.6

Means[a] and Standard Deviations for
SAT Mathematics Computation

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	10	13	7	30
	Pre	141.53 (7.78)	166.04 (10.34)	186.53 (4.50)	162.65
	Post	152.63 (8.88)	172.39 (8.64)	188.28 (5.78)	169.51
Control	N	10	12	6	28
	Pre	140.58 (5.25)	169.14 (8.6)	180.60 (7.31)	161.40
	Post	153.57 (8.22)	174.76 (10.8)	179.02 (9.98)	168.84
Total for Grades	N	20	25	13	58
	Pre	141.1	167.5	183.8	
	Post	153.1	173.5	184.0	

[a]Class means were used as the unit of analysis.

TABLE 5.7

Repeated Measures Analysis of Variance for
SAT Mathematics Application

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	21	17473.00	832.05	
Grade(G)	2	12847.00	6423.50	23.935**
Treatment(T)	1	4.00	4.00	0.015
T x G	2	328.00	164.00	0.611
Error	16	4294.00	268.38	
Within Ss	22	3038.00	138.09	
Tests(A)	1	803.00	803.00	8.978**
G x A	2	489.00	244.50	2.734
T x A	1	215.00	215.00	2.401
T x G x A	2	100.00	50.00	0.559
Error	16	1431.00	89.44	
Total	43	20511.00	477.00	

*$p < .05$

**$p < .01$

TABLE 5.8

Means[a] and Standard Deviations for
SAT Mathematics Application

Treatment Group		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	5	5	2	12
	Pre	130.89 (2.84)	163.87 (18.6)	183.0 (1.41)	153.32
	Post	153.91 (24.0)	167.79 (16.07)	191.14 (8.10)	165.9
Control	N	4	3	3	10
	Pre	133.74 (3.08)	168.66 (8.05)	180.67 (5.43)	158.29
	Post	142.04 (11.43)	172.89 (12.14)	177.66 (0.85)	163.4
Total for Grades	N	9	8	5	22
	Pre	132.2	165.7	181.6	
	Post	148.6	169.7	183.1	

[a]Class means were used as the unit of analysis.

No treatment differences were statistically significant. However, the grade level differences were highly significant at $p < .0001$; not surprisingly, classes at the higher grade levels scored higher in Mathematics Application.

There were also significant pre-to-post differences. Examination of the means in Table 5.8 reveals a similar pattern to the one found for Mathematics Computation data, with a decline in score increases at higher grade levels and with a slightly lower post-test than post-test mean score only for the control group at the seventh- and eighth-grade level. However, the interaction effect one would expect to be associated with this distribution of mean scores was not significant, at least in part because the number of classes which received the Mathematics Application subtest was less than half the number which received the Mathematics Computation subtest. Correspondingly, the number of degrees of freedom for the analysis of Mathematic Application scores was much reduced, and so, the test of significance was less sensitive.

D. <u>Mathematical Concepts</u>

The items in this subtest are concerned principally with instructional objectives related to number, notation, operations, geometry, and measurement. The items are dictated to those students who take the Primary levels of this subtest, so that the effect of reading ability on their scores is minimized.

The results of the repeated measures analysis of Mathematical Concepts scores are shown in Table 5.9. Means and standard deviations for this subtest are presented in Table 5.10

TABLE 5.9

Repeated Measures Analysis of Variance
for SAT Mathematics Concepts

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	30	24268.00	808.93	
Grade(G)	2	15002.00	7501.00	20.955**
Treatment(T)	1	31.00	31.00	0.087
T x G	2	286.00	143.00	0.400
Error	25	8949.00	357.96	
Within Ss	31	2637.00	85.06	
Tests(A)	1	119.00	119.00	1.377
G x A	2	333.00	166.50	1.929
T x A	1	6.00	6.00	0.069
T x G x A	2	18.00	9.00	0.104
Error	25	2161.00	86.44	
Total	61	26905.00	441.07	

*$p < .05$

**$p < .01$

TABLE 5.10

Means[a] and Standard Deviations for
SAT Mathematics Concepts

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	5	7	3	15
	Pre	141.93 (7.32)	166.83 (8.87)	182.63 (9.93)	161.69
	Post	137.28 (27.67)	175.42 (12.70)	186.31 (11.07)	164.88
Control	N	6	7	3	16
	Pre	144.99 (7.47)	167.71 (9.57)	175.44 (14.46)	160.64
	Post	143.30 (26.20)	173.85 (12.15)	177.08 (14.00)	163.00
Total for Grades	N	11	14	6	31
	Pre	143.60	167.30	179.00	
	Post	140.60	174.60	181.70	

[a]Class means were used as the unit of analysis.

There is evidence of growth in math concepts scores over grade levels, as indicated by the significant F-ratio ($p < .01$) in Table 5.9 for grade differences. However, within each block of grades, there was no significant increase in these scores from pre-test to post-test administration. Of all of the six SAT subtests which were administered, Mathematical Concepts is the only one for which there is no statistically significant difference on the factor of the repeated measure. In fact, an examination of the means in Table 5.10 reveals a slight loss from pre-to-post scores for both the USMES and control groups in grades 2-4, with the USMES group losing more than the control group. Slight gains occurred at the higher grade levels.

There was no significant difference in Mathematical Concepts scores between treatment groups. The means in Table 5.10 show that the pattern of insignificant gains or losses within a grade block, and the pattern of significant differences among grade levels, were similar for both treatment groups.

E. Science

The SAT Science subtest purports to measure the child's ability to understand basic concepts in the natural and physical sciences. The concepts included are: matter; energy; change in the physical universe; the environmental interaction of living objects; the effect of heredity and environment upon living objects; the interdependence of living objects; the basic processes of science; the basic measurement skills of science, and the ability to test hypotheses.

Results of the two-factor repeated measures analysis of variance for Science subtest scores are shown in Table 5.11. The means and standard deviations for treatment group at each grade level are given in Table 5.12.

TABLE 5.11

Repeated Measures Analysis of Variance for SAT Science

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	28	23670.00	845.36	
Grade(G)	2	18150.00	9075.00	43.648**
Treatment(T)	1	287.00	287.00	1.38
T x G	2	451.00	225.50	1.085
Error	23	4782.00	207.91	
Within Ss	29	1345.00	46.38	
Tests(A)	1	402.00	402.00	12.913**
G x A	2	197.00	98.50	3.164
T x A	1	20.00	20.00	0.643
T x G x A	2	10.00	5.00	0.161
Error	23	716.00	31.13	
Total	57	25015.00	438.86	

$^{*}p < .05$

$^{**}p < .01$

TABLE 5.12

Means[a] and Standard Deviations for SAT Science

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	5	6		15
	Pre	136.89 (9.13)	165.69 (14.04)	187.72 (9.44)	162.00
	Post	144.31 (5.41)	174.74 (17.05)	188.63 (11.75)	168.30
Control	N	6	5	3	14
	Pre	138.03 (8.93)	172.21 (8.67)	177.10 (8.22)	158.61
	Post	144.22 (11.39)	178.73 (5.45)	173.10 (12.88)	162.73
Total for Grades	N	11	11	7	29
	Pre	137.50	168.70	183.20	
	Post	144.30	176.60	182.00	

[a]Class means were used as the unit of analysis.

Grade level differences were statistically significant at $p < .0001$, and overall difference scores between pre-test and post-test administration were significant at $p < .01$. As expected, classes at the higher grade levels tended to score higher on this subtest, and in general, post-test performance was significantly better than pre-test performance.

Inspection of the means in Table 5.12 reveals, however, that the rates of change from pre-test to post-test administration varied according to grade level. There was steady growth from pre- to post-test scores observed for both USMES and control groups in grades 2-4 and in grades 5-6. In contrast, there was virtually no growth from pre-test to post-test means for the USMES group in grades 7 and 8, and there was a decline in the averages for the control group in this grade block. In fact, the control classes at the fifth- and sixth-grade level out performed the control classes at the seventh- and eighth-grade level.

Referring to the results of the analysis of variance for Science scores in Table 5.11, one notes that this interaction effect for grade by test administration was not significant at the .05 level. However, the F-ratio of 3.16 approached closely the value of 3.42 which is required for statistical significance at $p=.05$ and the nature of this interaction is a noteworthy aspect of the Science subtest data.

F. Social Science

The Social Science Test is designed to measure the pupil's ability to understand concepts in six social science disciplines: Geography, History, Economics, Political Science, Anthropology, and Sociology. The abilities to infer, to reason, to predict, and to conclude are measured in different ways through the reading of globes, interrelated maps, demographic data, political posters, as well as through questions which call for the display of these inquiry skills directly.

The results of the repeated measures analysis of variance for the Social Science subtest scores are reported in Table 5.13, while Table 5.14 contains the means and standard deviations for treatment groups at each grade level on this subtest.

As shown in Table 5.13, the test score differences attributable to the main effect of grade were highly significant at $p < .001$. Furthermore, pre-to-post administration differences were significant at $p < .01$. However, two first order interaction effects were also significant. Interpretation of pre-test to post-test differences required qualification by grade level and by treatment group.

The USMES groups at all three grade level blocks showed gains from pre-test to post-test, but their rates of growth declined at the higher grade levels. The patterns of change for the control classes were less consistent. The control group at the lowest grade level, grades 2-4, showed virtually no change in average performance from pre-to-post test time. The control classes in grades 5 and 6 averaged six points higher on the post-test than on the pre-test, and the control group at the seventh- and eighth-grade level witnessed a 7-point decline in average performance.

Results of the analyses of all SAT subtest data are summarized and discussed later in a concluding section to this chapter.

Additional Problems for the Analysis and Interpretation of SAT Data

As indicated earlier in this chapter the analysis of covariance was rejected as an appropriate method for analyzing the differences between the treatment groups. A covariance adjustment for pre-test differences between

TABLE 5.13

Repeated Measures Analysis of Variance
for SAT Social Science

Source	df	Sum of Squares	Mean Square	F-Ratio
Between Ss	24	16547.00	689.46	
Grade (G)	2	11967.00	5983.50	28.071**
Treatment (T)	1	90.00	90.00	0.422
T x G	2	440.00	220.00	1.032
Error	19	4050.00	213.16	
Within Ss	25	953.00	38.12	
Tests (A)	1	192.00	192.00	8.387**
G x A	2	159.00	79.50	3.472*
T x A	1	138.00	138.00	6.028*
T x G x A	2	29.00	14.50	0.633
Error	19	435.00	22.89	
Total	49	17500.00	357.14	

*p < .05

**p < .01

TABLE 5.14

Means[a] and Standard Deviations for
SAT Social Science

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	5	6	2	13
	Pre	139.10 (6.89)	166.50 (15.83)	184.60 (3.42)	158.80
	Post	147.80 (6.54)	173.50 (13.89)	188.00 (0.00)	165.80
Control	N	4	5	3	12
	Pre	143.20 (4.27)	173.50 (12.36)	178.80 (4.64)	164.70
	Post	142.60 (12.08)	179.50 (12.70)	171.50 (7.20)	165.20
Total for Grades	N	9	11	5	25
	Pre	[illegible].90	169.70	181.10	
	Post	145.50	176.20	178.10	

[a] Class means were used as the unit of analysis.

USMES and control groups would have been desirable, but a critical assumption, of homogeneity of variance, could not be justified.

A careful review of the standard deviations, which were presented along with the means and cell n's in the preceding section, alerted us to the presence of additional problems for analyzing and interpreting the data. Table 5.15 contains only the standard deviations on pre-test and post-test for both treatment groups, at each grade level. In studying these data we were struck with what seemed to be a very large range among the standard deviations of the cells. When we compared these standard deviations with those in the SAT norms booklet, our apprehension was reinforced. In Table 5.15 we have included the standard deviations from the SAT Norms Booklets, (1973) for comparison. (The standard deviations for the treatment groups should be consistently smaller than those presented in the norms booklet, because in this evaluation class means were used as the unit of analysis.)

Trying to reach some generalization about the reasons for this large spread in the standard deviations proved futile: not all of them were larger than expected (we even observed an S of 0.00 for a cell containing two classes which achieved identical means); they were not concentrated at one grade level; the curious result was not limited to a single test area; the size of the standard deviations was not consistent even from pre-test to post-test within a specific test area and level. We considered several explanations to account for the extremely varied standard deviations within the cells.

The very small standard deviations occurred only at the seventh- and eighth-grade level. It seemed that these extremely small s's could be attributed, in part, to the small number of classes within the cells, since the standard deviations in Table 5.15 are measures of the deviation of class means, the unit of analysis, from the cell means. The extremely large

TABLE 5.15

Standard Deviations of Scores
on 6 SAT Subtests

SAT Subtest	Treatment Group	Test	Grade Levels 2-4	5-6	7-8
Reading Comprehension	USMES	Pre	12.06	14.44	6.93
		Post	5.56	14.39	5.96
	Control	Pre	11.30	13.37	9.90
		Post	8.47	11.73	10.10
	Norms*		~(11-12)	~(14-15)	~(14-15)
Mathematics Computation	USMES	Pre	7.78	10.34	4.50
		Post	8.88	8.64	5.78
	Control	Pre	5.25	8.6	7.31
		Post	8.22	10.9	9.98
	Norms		~(6-[illegible])	~(8-10)	~(8-10)
Mathematics Application	USMES	Pre	2.84	18.50	1.41
		Post	24.0	16.57	8.10
	Control	Pre	3.08	8.05	5.43
		Post	11.43	12.14	0.85
	Norms		~(5-6)	~(9-10)	~(9-10)
Mathematics Concepts	USMES	Pre	7.32	8.87	9.93
		Post	27.67	12.[illegible]	11.07
	Control	Pre	7.47	9.57	14.46
		Post	26.2	12.15	14.0
	Norms		~(6-7)	~(6-7)	~(6-7)
Science	USMES	Pre	9.13	14.04	9.44
		Post	5.41	17.05	11.75
	Control	Pre	8.93	9.67	8.22
		Post	11.39	5.45	12.88
	Norms		~(4-5)	~(11-12)	~(11-12)
Social Science	USMES	Pre	6.89	15.83	3.42
		Post	6.54	13.89	0.00
	Control	Pre	4.27	12.36	4.64
		Post	12.08	12.70	7.2
	Norms		~(3-4)	~(9-10)*	~(12-13)*

*The approximate standard deviations from the SAT Norms Booklet.

standard deviations could not be explained so readily.

The observation that the very large standard deviations were found principally in grades 2-4 drew our attention to the fact that the data for this block of grades did differ in one important respect from the data for grades 5-6 and for grades 7-8. This difference is illustrated in Table 5.16. The levels of the SAT which were required for both pre-test and post-test administration to classes in grades 2-4 encompassed five levels of the SAT. For grades 5 and 6, only two levels were needed for both pre-test and post-test administration. In the block for grades 7 and 8, only one level was required.

Scaled socres for the SAT were developed in such a way that this use of various levels of the sequential battery should not have made a difference. The test developers claim that their scaling procedure resulted in a system of inter-battery standard scores which should permit the translation of raw scores at each level to standard scores with comparability across levels for a test area. Thus, if the appropriate conversion tables were used, the choice of test level, within the narrow limits offered for a grade level, should not have made a difference in the scaled score which an individual obtains. Our observation of extremely varied standard deviations among the cells of Table 5.15 lead us to question the validity of this claim.

The seriousness of this problem of heterogeneity of variances for the USMES evaluation is diminished greatly by the consistency of the results of the analysis of basic skill development, over previous evaluations and within this evaluation. Sophisticated analyses are not needed to ascertain that the national sample of USMES classes performed at least as well as their control matches on the six selected subtests of the Stanford Achievement Test battery.

TABLE 5.16

Forms of the SAT Administered Within
Groups of Grades

Test	Grades 2-4	Grades 5-6	Grades 7-8
Pre-test	Primary I	Intermediate I	Advanced
	Primary II	Intermediate II	
	Primary III		
	Intermediate I		
Post-test	Primary II	Intermediate II	Advanced
	Primary III		
	Intermediate I		
	Intermediate II		

Discussion and Summary

A pre-test, post-test control group design was used to investigate whether USMES students maintain the same level of basic skill development as control students, even though USMES usage may detract from the amount of basic skills instruction which USMES students can receive. Basic skill development was measured with Fall and Spring administrations of selected subtests of the Stanford Achievement Test battery: (a) Reading Comprehension, (b) Mathematics Computation, (c) Mathematics Application, (d) Mathematics Concepts, (e) Science, and (f) Social Science.

Several problems were encountered in the collection, analysis and interpretation of the SAT data. However, none of these problems prevented an unequivocal response to the expressions of concern about accountability for the communication of basic skills. Clearly, USMES students do not fall behind their control counterparts in their performance on tests of basic skills. On all six post-test measures, the overall USMES mean was higher than the overall control mean.

The odd-numbered figures, from Figure 5.1 through Figure 5.11, are presented to aid one's review of the pre-to-post changes in mean scores on each subtest for USMES and control groups at each grade level. The F-ratio for overall pre-post differences from each of the repeated measures analysis was statistically significant, except when scores from the Mathematical Concepts subtest were used as the dependent variable. This general pattern of increases in score means from pre-test to post-test administration for both treatment groups can be observed in the graphs.

The USMES program purports to enhance the problem solving ability of elementary school students without imparing their basic skill development.

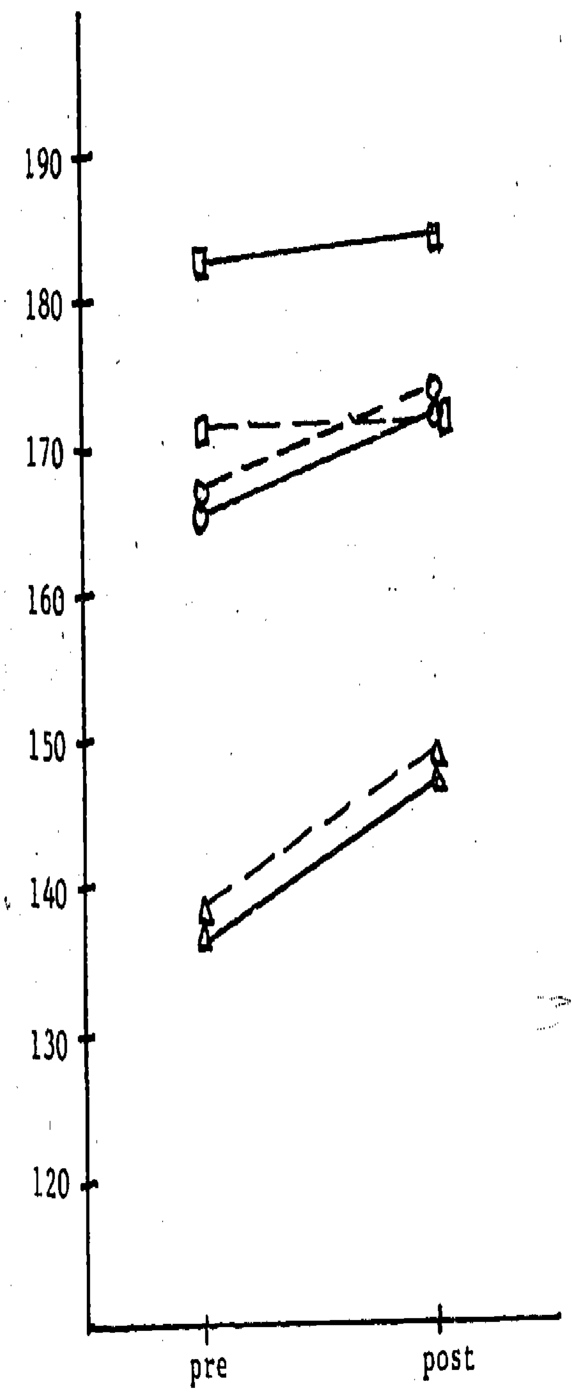

Figure 5.1-- Graph of means on SAT Reading Comprehension Subtest showing pre-to-post differences.

Key: Grades 2-4: Δ USMES: ———
Grades 5-6: O Control: -------
Grades 7-8: □

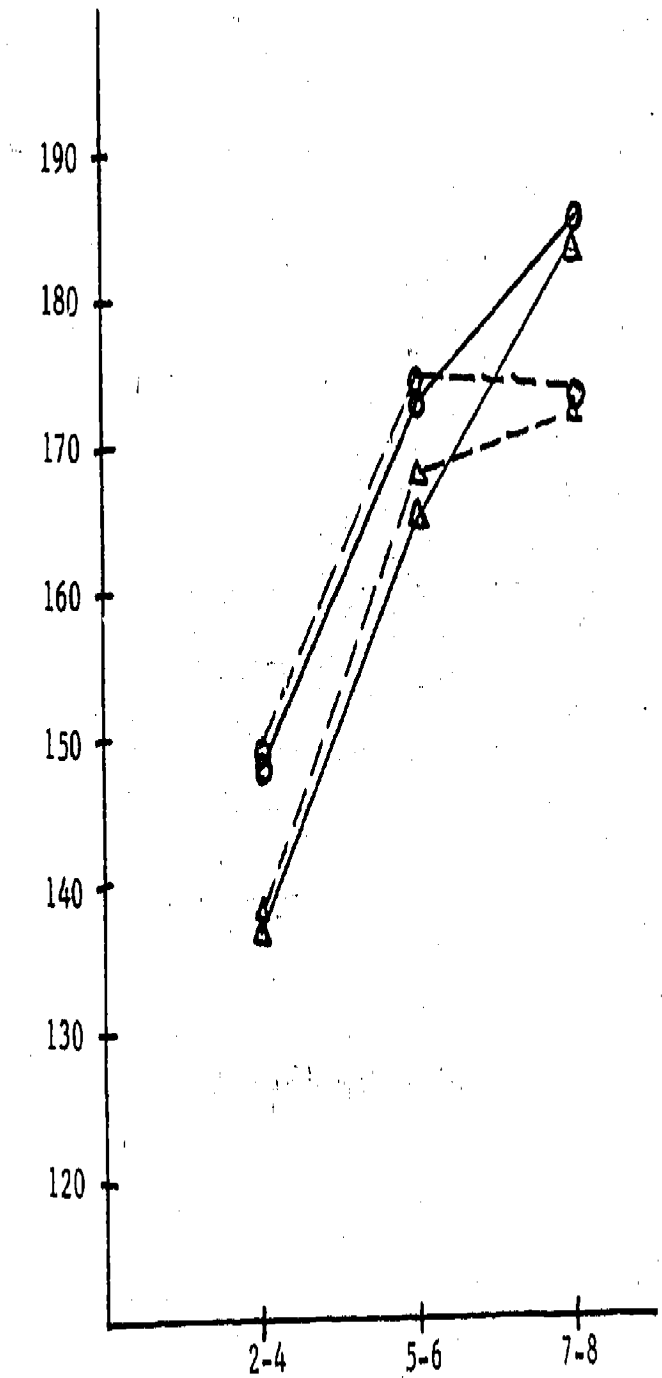

Figure 5.2-- Graph of means on SAT Reading Comprehension Subtest showing grade level differences.

Key: Pre: Δ USMES: ———
Post: O Control: -----------

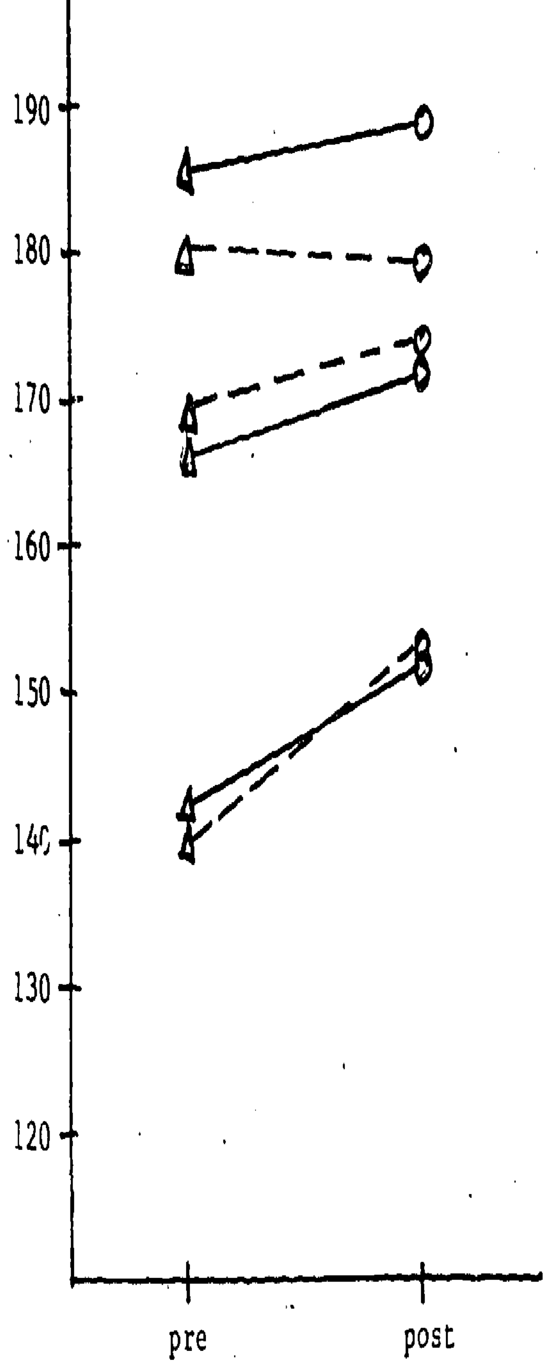

Figure 5.3-- Graph of means on SAT Mathematics Computation Subtest showing pre-to-post differences.

Key: Grades 2-4: Δ
Grades 5-6: O
Grades 7-8: □

USMES: ———
Control: ------

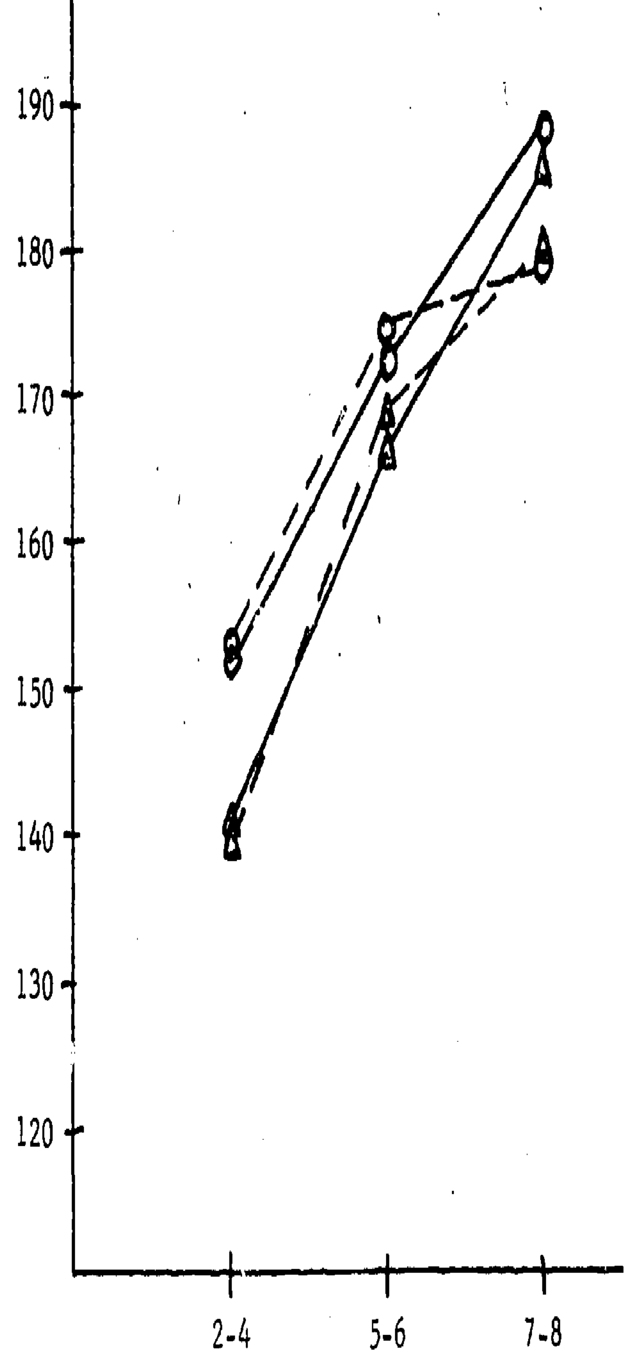

Figure 5.4-- Graph of means on SAT Mathematics Computation Subtest showing grade level differences.

Key: Pre: Δ
Post: O

USMES: ———
Control: ---------

-119-

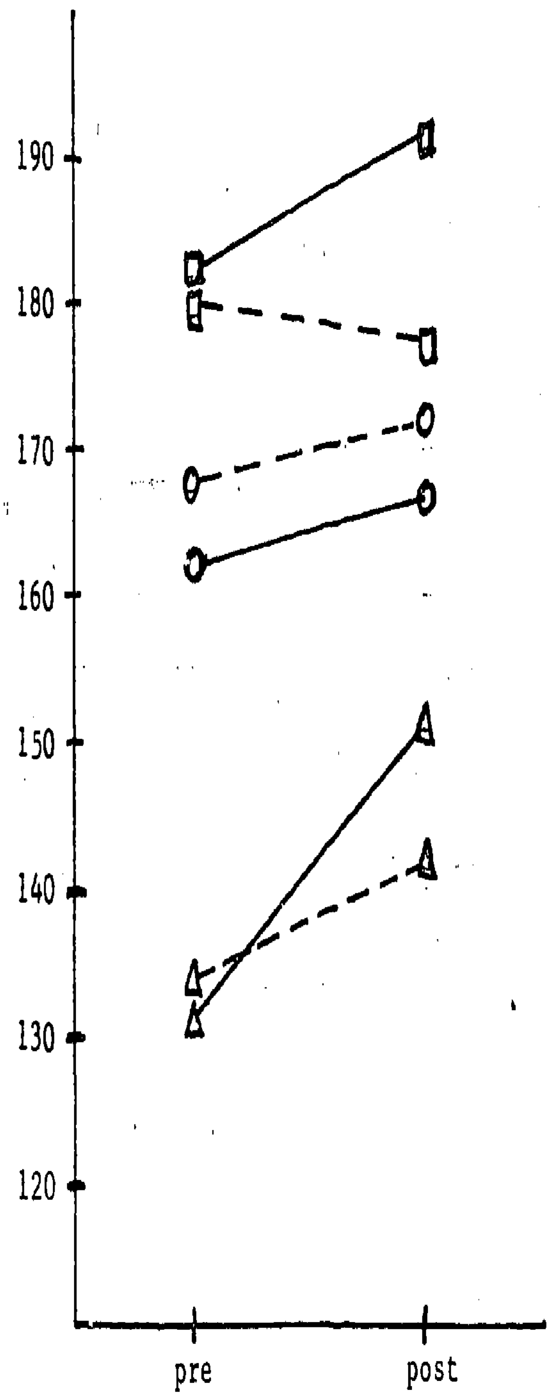

Figure 5.5-- Graph of means on SAT Mathematics Application Subtest showing pre-to-post differences.

Key: Grades 2-4: △
Grades 5-6: ○
Grades 7-8: □
USMES: ————
Control: --------

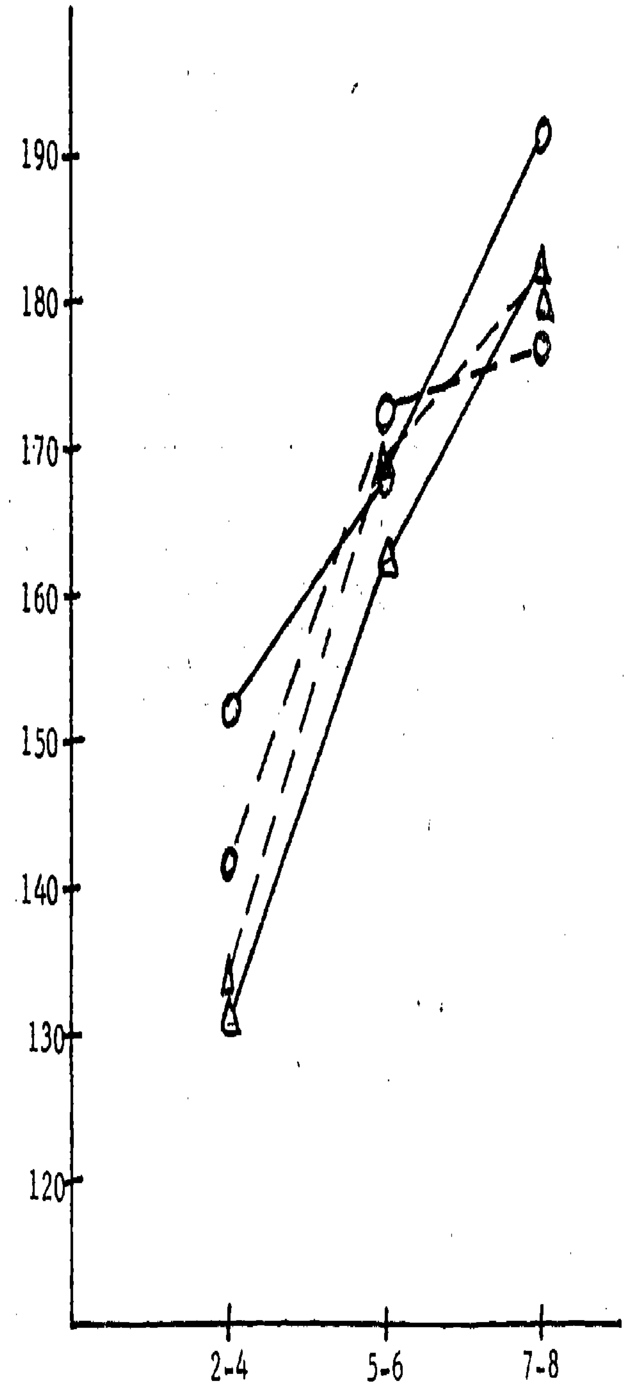

Figure 5.6-- Graph of means on SAT Mathematics Application Subtest showing grade level differences.

Key: Pre: △
Post: ○
USMES: ————
Control: -----------

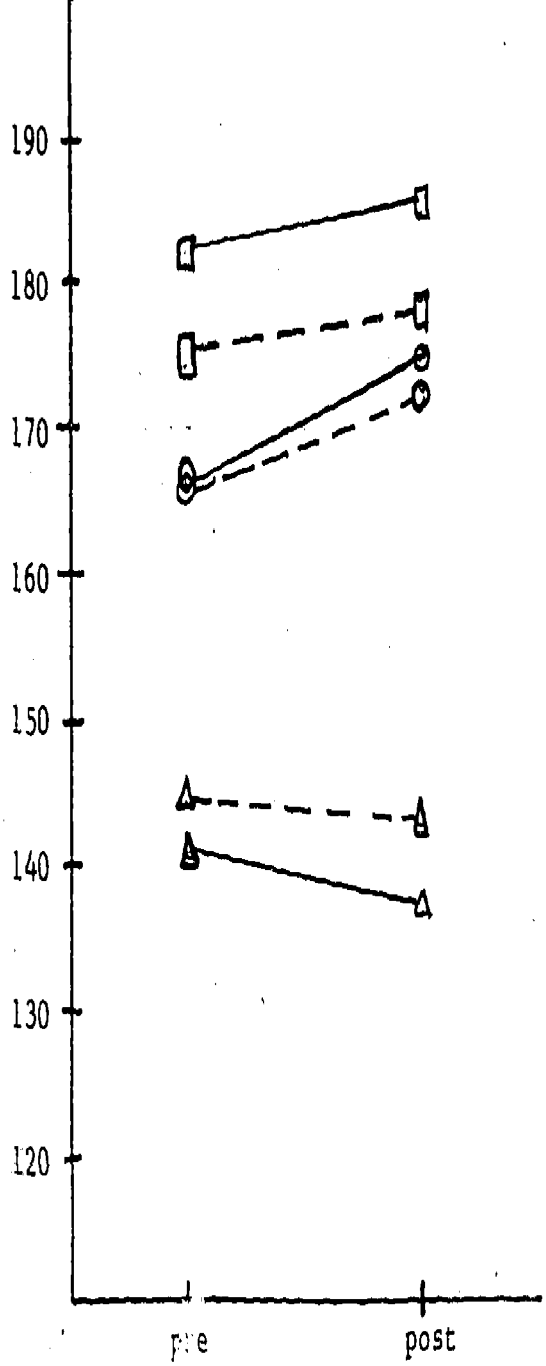

Figure 5.7-- Graph of means on SAT Mathematics Concepts Subtest showing pre-to-post differences.

Key: Grades 2-4: Δ
Grades 5-6: O
Grades 7-8: □
USMES: ————
Control: --------

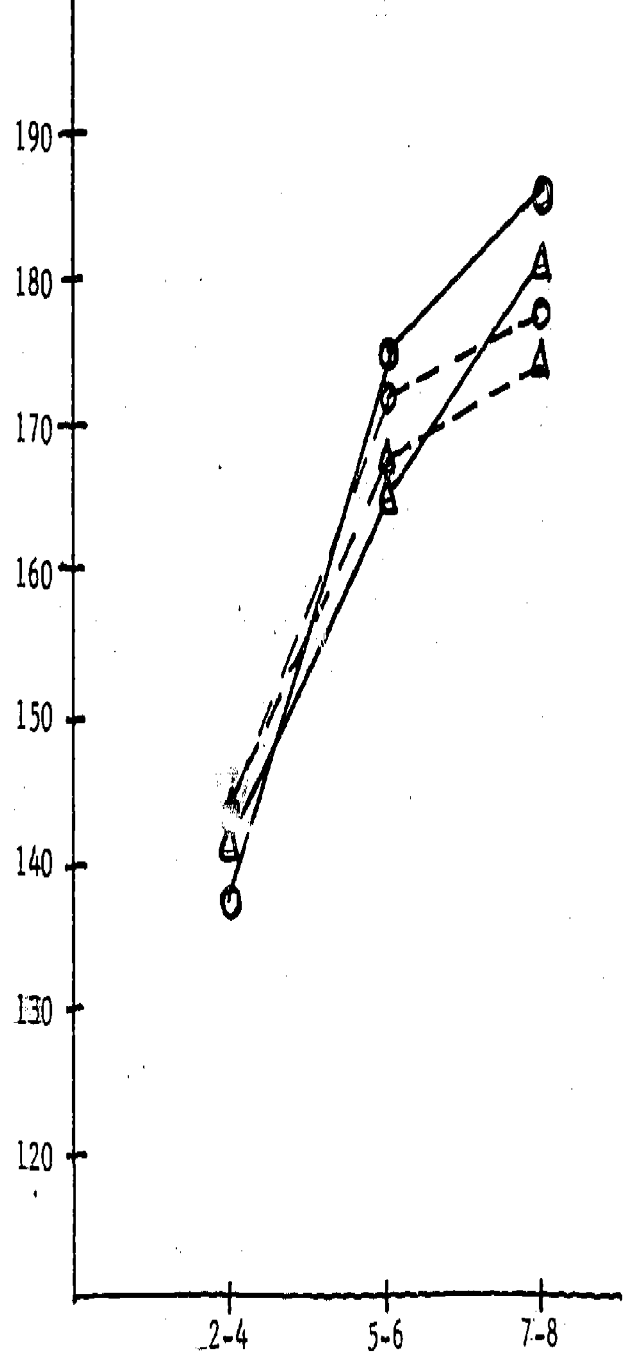

Figure 5.8-- Graph of Means on SAT Mathematics Concepts Subtest showing grade level differences.

Key: Pre: Δ
Post: O
USMES: ————
Control: -----------

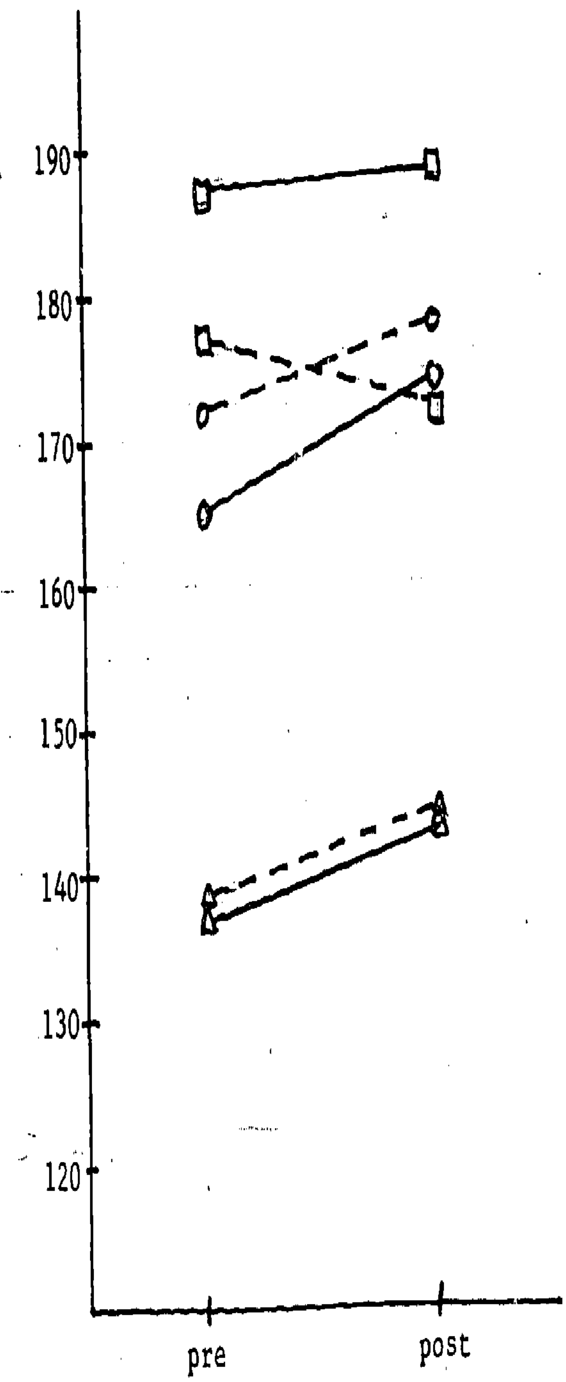

Figure 5.9-- Graph of means on SAT Science Subtest showing pre-to-post differences.

Key: Grades 2-4: Δ
Grades 5-6: O
Grades 7-8: □
USMES: ————
Control: ---------

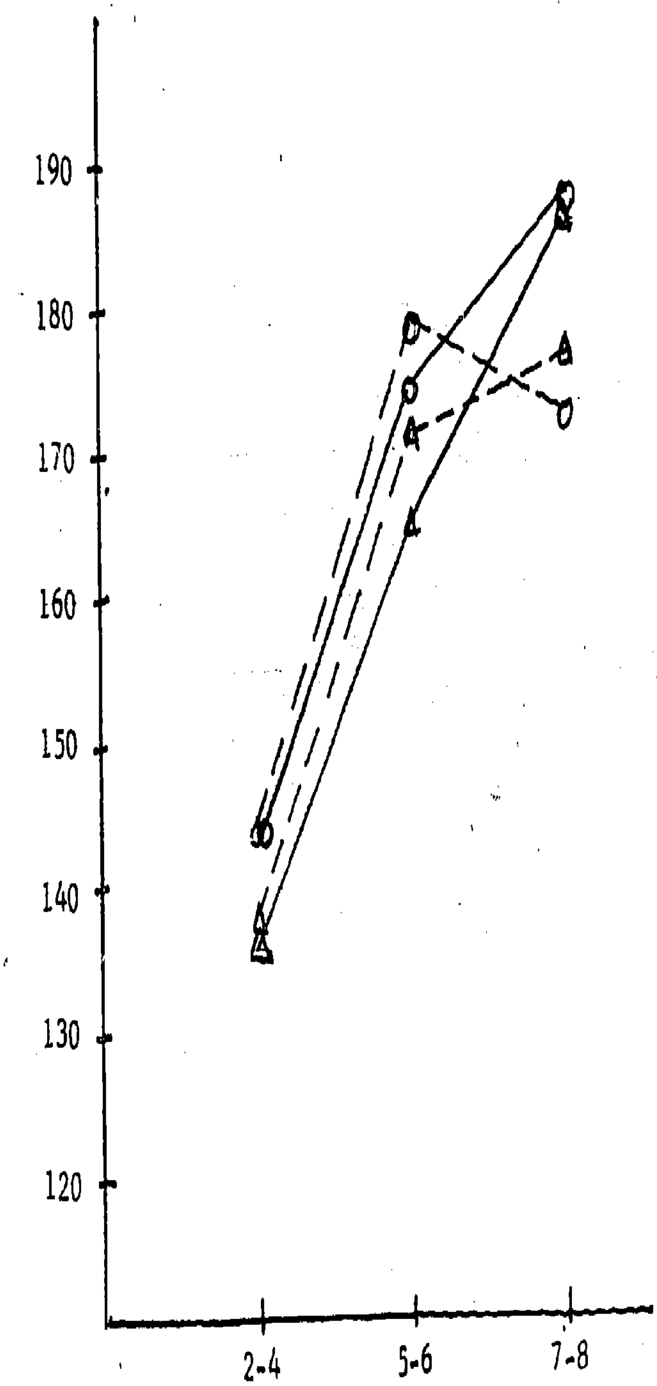

Figure 5.10-- Graph of means on SAT Science subtest showing grade level differences.

Key: Pre: Δ
Post: O
USMES: ————
Control: -----------

144

145

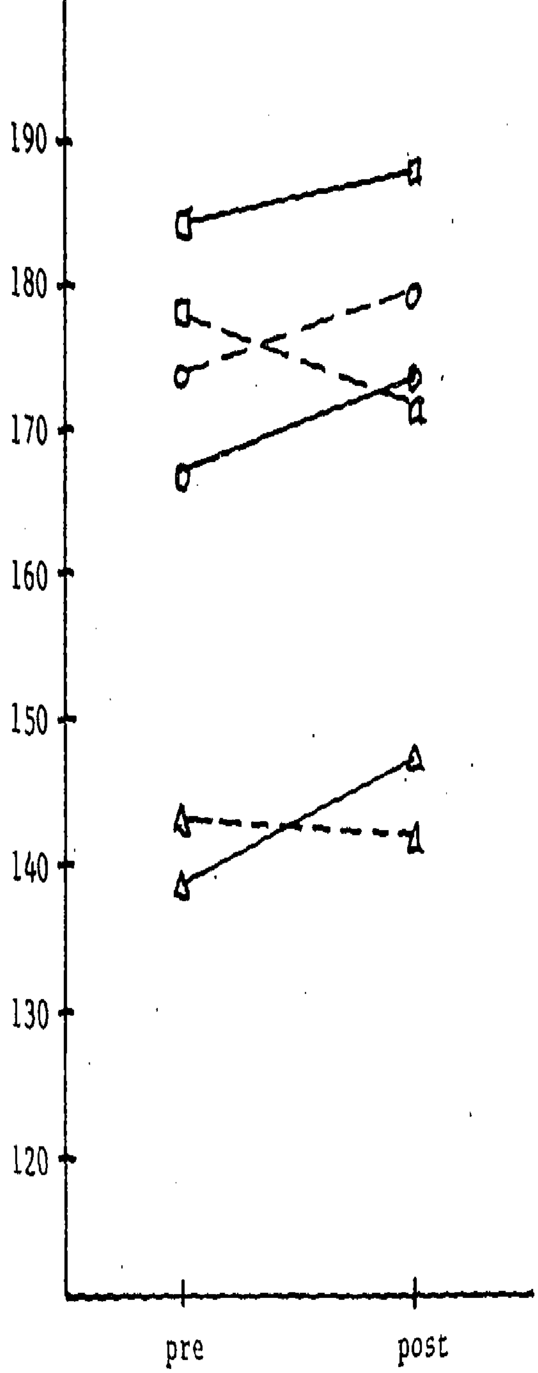

Figure 5.11-- Graph of means on SAT Social Science Subtest showing pre-to-post differences.

Key: Grades 2-4: △ Grades 5-6: ○ Grades 7-8: □

USMES: ————— Control: - - - - - - -

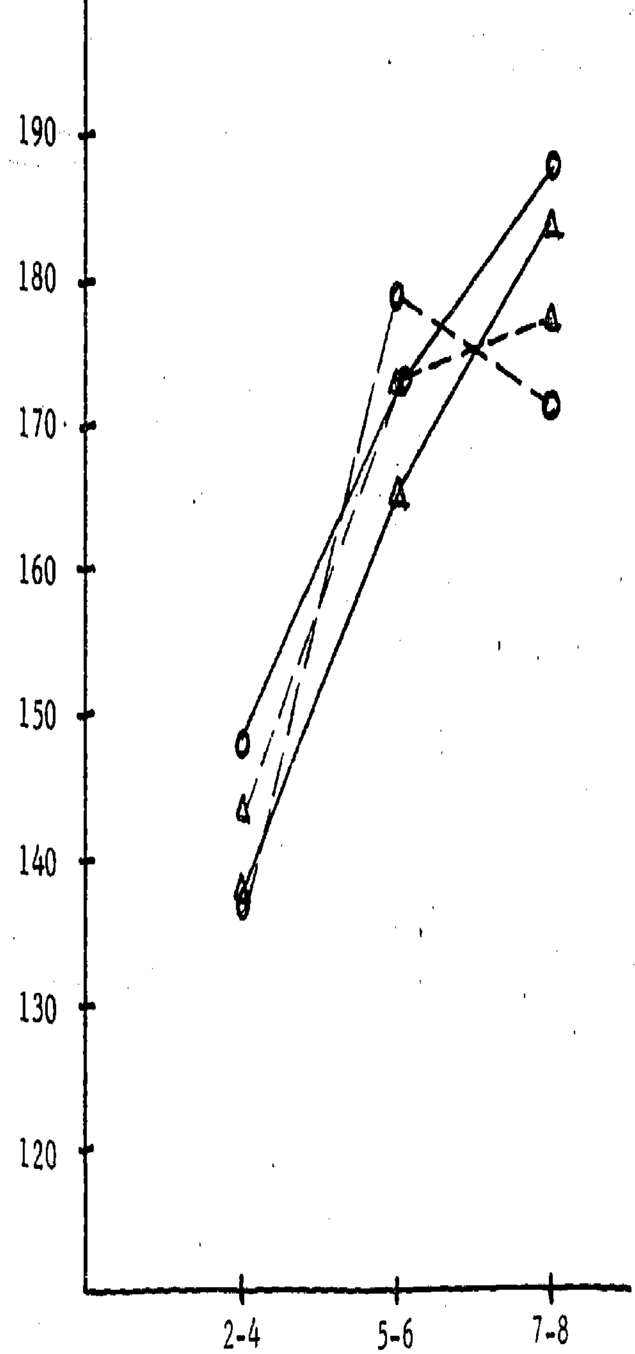

Figure 5.12-- Graph of means on SAT Social Science Subtest showing grade level differences.

Key: Pre: △ Post: ○

USMES: ————— Control: - - - - - - -

time from these areas. Some of the classes are not self-contained, and for these classes, schedules and amounts of time are mandated for basic skills instruction.

Clearly, USMES usage, as practiced by sample classes representing a wide distribution of geographic areas and socioeconomic levels, did not affect basic skill development adversely. Previous investigations on this issue yielded similar results. The measurement of basic skills has been a costly and time consuming activity and sample teachers, principals, and their students have become increasingly resentful that this kind of test administration is disruptive of the school day, and sometimes is threatening to students.

We recommend that the resources devoted to comparing the basic skill development of USMES and control students should not be expended in the future. Moreover, the issue of basic skill development should be of diminished importance in light of the patterns of USMES usage in most schools. Most frequently, the time for USMES comes from regularly scheduled science time, and to a lesser extend from project time. Hence, one should not expect USMES children to fall behind in the basic skills areas of reading, language arts, and mathematics which are of primary concern to most elementary school personnel. Stated simply, the issue is not an issue.

CHAPTER VI

STUDENT PERFORMANCE ON SIMULATED REAL-LIFE PROBLEM TASKS: RESULTS OF THE PLAYGROUND AND PICNIC PROBLEMS

The primary objective of the USMES program is the enhancement of elementary school students' abilities in real, complex problem solving. Accordingly, the primary responsibility of the USMES evaluation project staff has been the investigation of whether USMES is achieving this goal.

This determination of proof of concept has been difficult and challenging for the evaluators, because the "state of the art" of measuring the problem solving abilities of children is itself so limited. After an exhaustive review of the problem solving measures reported in the research literature in psychology, general education, science education, and mathematics education, we determined that no available measure was appropriate for evaluating USMES-styled problem solving.

Existing tests were faulted as measures of "real" complex, USMES-like problems for a variety of reasons: they measured skill on arithmetic word problems; or they required only quick insight into artificial puzzle situations; or their items relied heavily on specific content which many USMES and control students could not be expected to have acquired; or the tests purported to measure problem solving processes, but in our judgement and that of our consultants, they measured factual recall. Other tests were dismissed because their application was limited to high-school-age students and adults. Still other tests of problem solving designed for younger students were rejected because they were judged technically inadequate, with little or no reliability evidence, poor item statistics, ambiguous items, or low quality testing materials.

This assessment of the state of the art concerning the measurement of problem solving necessitated that the evaluation project assume a two-fold thrust: (1) program evaluation, and (2) new instrument development for measuring children's abilities in complex problem solving. Yet, the immediate need shared by the funding agency and the program developers for proof of concept forced us to apply the best available techniques for the 1974-75 evaluation, as we concurrently pursued new instrument development.

Our research on problem solving and the results of our new instrument development work are reported in a separate document whose publication and distribution will follow closely the availability of this evaluation report. The 1974-75 evaluation of USMES students' participation and growth in problem solving skills offered in the present document is based on the results of several indicators: classroom acitivity analyses, teacher interviews, student interviews, and the Playground and Picnic Problems. It is the results of the latter two indices which are reported in this chapter.

Conceptual Basis for the Playground and Picnic Problems

The USMES philosophy is an eclectic one; it encompasses features of the theoretical positions expressed by Dewey, Bruner, Gagne, and others. Most consistently evident in the USMES developers' written statements about the USMES approach, however, are references which call to mind John Dewey's "five logically distinct steps" of the problem solving process:

1. Recognizing that a problem exists
2. Identifying the nature of the problem
3. Searching for possible solutions
4. Analyzing the adequacy of the tentative solutions
5. Testing the most promising of the tentative solutions (Dewey, 1910)

Guilford (1965, p. 8) compared many recent theories about creative production, and he concluded that the most remarkable thing about them was their similarity to those of Dewey. The USMES developers' parallel to Dewey's conceptualization is illustrated in the following statement:

> "To learn the process of real problem solving, the students themselves, not the teacher, determine the route they will take. The children analyze the problem, choose the variables that should be investigated, search out the facts, and judge the correctness of the hypotheses and conclusions (USMES Guide, 1975, p.

The Playground Problem was conceptualized according to this description of the problem solving process. This test required that students develop a plan for a playground which would serve children in their school and/or neighborhood. A catalog of equipment, cost data, and measuring instruments were given to the students along with the information that they could spend up to $2000.

The pre-test, post-test control group design used in the evaluation necessitated that a parallel form for the Playground Problem be developed, since retest results from such a unique test would be affected by memory factors. To answer this need, the Picnic Problem was developed. This test challenged students to develop plans for a class picnic. The students were provided with a photograph of various foods available to them and a map drawn to scale which included the locations of their school and three park areas as possible sites for the picnic. Along with measuring instruments, the students were given cost data and the information that they could spend up to $25. They were to assume that 25 students would be going on the picnic, and that a school bus would be provided for their transportation, free of charge.

Neither the playground problem nor the picnic problem satisfied the developers' concern that these tests meet all the criteria for "realness." The tests were simulated problems whose solutions would not have immediate, practical effects on students' lives. Nevertheless, data shown later in this chapter indicate that the vast majority of students tested with the Playground and Picnic tasks were motivated to accept the problems. In that sense. we can say the tasks were meaningful to the students.

Another of the developers' criteria for "realness" is that real challenges are "big" enough to require many phases of class activity for any effective solution. The Playground and Picnic Problems did not meet this criterion. In the interest of observing reasonably larger samples of children we had to abbreviate test times to approximately one hour.

Depsite these limitations, the Playground and Picnic Problems have other important features in common with USMES-styled, real problems: they have no "right" solutions; they have no clear boundaries; they require students to use their own ideas for solving the problems; and they elicit group efforts toward the solutions to the problems.

Test Administration

The field staffers of the evaluation team were specially trained to administer the Playground and Picnic Problems. To help standardize administration procedures, the Project Director developed Administrator's Manuals for both the Playground and Picnic Problems. These are shown in Appendices D and E.

Both tests were designed for administration to small groups of children. The pre-test, post-test control group design called for Fall and Spring testing in both USMES and control sample classes. Two groups of five children were randomly selected from each sample class. In the Fall, one group received the Playground Problem, the other, the Picnic Problem. In the Spring, the same two groups of children from each class were to receive the alternate form of the test they had worked on in the Fall. If any student was absent for the post-testing, another student was to be selected randomly to take the place of the absentee.

The distributions of sample USMES and control classes with pre-test and post-test scores used in the analysis of the Playground and Picnic Problems are shown by grade level in Table 6.1

Scoring Procedures and Scorer Reliability

The USMES developers have emphasized repeatedly the importance of student involvement in all aspects of the real problem solving process. Some of these aspects are: determination of the important factors in a problem; determination of how to measure these factors; data collection, recording, and analysis; discussion and group work. A scoring protocol for the Playground Problem was developed which would enhance reliable assessments of students' performance on several of these aspects.

A behavioral assessment included rating scales on four aspects: motivation to accept the problem; commitment to task; allocation of responsibilities for efficiency of manpower; and the nature of group leadership. The cognitive assessment included four summary rating scores on variable identification, measurement, calculation, and recording. Observers' notes and students' audio tapes were the bases for these behavioral and cognitive assessments. The students' drawings of their proposed playgrounds were analyzed to yield four product scores: scale, labels, landmarks, and area designation. In summary, then, for the Playground Problem, 12 scores were

TABLE 6.1

Number of Sample Classes with Pre-test and Post-test Scores Used in the Analysis of the Playground Problem and the Picnic Problem Grade Level by Treatment

Treatment Group	Grade Levels 2-4	5-6	7-8	Total
Playground Problem				
USMES	14	13	6	33
Control	12	13	8	33
Picnic Problem				
USMES	14	13	6	33
Control	12	14	8	34

derived from the scoring protocol: four behavioral, four cognitive, and four product scores. The actual derivation of scores is described in the Scoring Manual shown in Appendix F.

A parallel form of the scoring protocol was developed for the Picnic Problem. The specific variables likely to be identified and the descriptions provided in the rating categories had to be changed to fit the context of the Picnic Problem. In addition, no product akin to a drawing of their playground design resulted from the Picnic Problem, so product assessments were not available for the Picnic test. The Picnic Problem Scoring Protocol, shown in Appendix G, yielded assessments on four behavioral aspects and four cognitive aspects of the students' work to the solution of the problem.

Four persons, college graduates with backgrounds in education and psychometrics, were involved in scoring the Playground and Picnic Problems. The original training only involved two scorers. During the first training session, a staff member explained the conceptual bases of the Playground Problem and its scoring protocol. Then, individually, at home, each scorer analyzed the same 10 tests. Only Playground Problems were used in this original training session.

During the second training session, a scorer interreliability check was done. The two scorers compared their score sheets for any discrepancies. There were several areas of disagreement. The taped interviews with the children were replayed, and the observer sheets were reexamined. The problems were resolved after considerable discussion. By the end of this session, there was general agreement on the procedures for scoring and on how to rate various kinds of responses.

The third training session was held several weeks later, when the scorers were very familiar with the Playground Problem. They were given 10 Picnic Problems to score in the same manner that the Playground Problems had been socred, after a general explanation of the Picnic Problem. At the following session, the score sheets were compared. This time, careful attention was paid to discrepancies on scoring of the behavioral aspects, since the scoring schemes for these aspects were identical for both the Playground and Picnic Problems. General agreement was found for the most part, and discrepancies were resolved in the same manner as for the Playground Problem.

Three months later, two additional scorers had to be trained to complete the scoring of the tests. Their training was similar to that received by the original scorers. The new scorers continued to score the 20 Playground and Picnic Problems used in the original training until close agreement occurred. This was completed in 3 sessions. Towards the end of the scoring process, all coders reconvened and compared their scores on the same three tests. Agreement among raters remained high.

Other Reliability and Validity Evidence

The Playground Problem and the Picnic Problem as tests are unusual, even unique. Appropriate use of these tests requires skilled administration by a trained examiner to a group of five children who should work as a team on the creation of a solution to a complex problem. There is no one, right, readily derived solution for either problem. These features make an appropriate reliability assessment somewhat difficult and exceedingly rigorous.

Test-retest reliability is not an appropriate method because the tests are so unique that, according to our observers, students can remember their work on the Playground and Picnic Problems eight months after their administration.

Internal consistency reliability estimates were inappropriate, given the variety of distinct behavioral and cognitive assessments which the scoring protocols were designed to yield. Low interscale and scale and scale-total correlations may indicate only that the scoring protocols were assessing various distinct aspects of problem solving.

Equivalent forms reliability checks would have been most appropriate, and the most rigorous assessment of the problem solving tests' reliability. Indeed, it had been our intention to conduct studies with children who were not part of the evaluation sample to determine the equivalence and stability of the Playground and Picnic Problem as assessments of problem solving and to investigate the influence of time limits, degree of structure in directions, and group vs. individual performance on students' performance with these tests. However, the USMES developers had serious objections to the use of simulated problems in these tests, so that costs and efforts for additional study of these tests by evaluation team did not seem to be warranted.

We did compute correlations between the scores on the Playground Problem and the Picnic Problem, at pre-test time and at post-test time, on the four cognitive scales which both tests offer: identifies, measures, calculates, and records. These coefficients are shown in Table 6.2 for USMES classes, for control classes, and for the total sample. In one sense, these correlation coefficients are indices of the equivalence within classes of the two forms of the problem solving test. However, the tests were administered to assess pre- and post-test performance of two randomly selected groups within each class, not to study equivalence reliability in the evaluation sample. Therefore, each correlation coefficient in Table 6.2 reflects the correspondence in performance of two different groups of children within the same class. Not only

TABLE 6.2

Correlations Between Picnic and Playground Problem Scores on Pre-test and Post-test for USMES and Control Classes, and for Total Sample

Variable	Treatment Group	Pre	Post
Identification	USMES	0.356*	0.172
	Control	0.351*	0.305*
	Total	0.354**	0.248*
Measures	USMES	0.450**	0.348*
	Control	0.511**	0.277
	Total	0.479**	0.299**
Calculates	USMES	0.561**	0.351*
	Control	0.570**	0.402**
	Total	0.558**	0.370**
Records	USMES	0.429**	0.335*
	Control	0.526**	0.231
	Total	0.476**	0.271*

*$p < .05$

**$p < .01$

would the size of the correlation coefficients be diminished by non-equivalence of the test forms, but also, the values would be suppressed by non-equivalence of the two groups randomly selected from each class, even though each group of five was sought to be representative of all the children in a class.

The correlation coefficients between pre-test and post-test performances on the Playground Problem and on the Picnic Problem are shown in Table 6.3. Again, the coefficients are based on scores for two different groups of five children in each class. This, one group took the Playground Problem in the Fall, another group from the same class worked on the Playground Problem in the Spring. Both groups' scores on a variable constituted a pair, and the pairs for all classes were the data on which the correlation coefficient for a given variable was computed.

The pre-post correlations for the Picnic Problem socres were somewhat higher than the pre-post correlations for the Playground scores on the same variables, as the coefficients in Table 6.3 indicate. This result can be explained by the difference in the number of factors which the children could investigate readily on the two tests. There are several variables in the Picnic Problem which, for the older children, were very obvious factors to take into consideration: cost of admission to the parks, cost of food, time, distance, and size of parks. For the Playground Problem, only the cost, size, and placement of equipment were the more obvious factors to study, and the latter two variables did not lend themselves to measurement and calculation as readily as the variables from the Picnic Problem. (The required covariation in the pre-post scores on the Picnic Problem which must be present for correlation can be attributed largely to grade differences, as

TABLE 6.3

Correlations Between Pre-test and Post-test on Playground and Picnic Problems for USMES and Control Classes and Total Sample

Variable	Treatment Group	Playground	Picnic
Identification	USMES	0.334*	0.308*
	Control	0.253	0.525**
	Total	0.303**	0.414**
Measures	USMES	0.328*	0.503**
	Control	0.223	0.542**
	Total	0.282*	0.516**
Calculates	USMES	0.323*	0.565**
	Control	0.571**	0.617**
	Total	0.439**	0.578**
Records	USMES	0.227	0.294
	Control	0.317*	0.497**
	Total	0.275*	0.395
Behavior	USMES	-0.020	0.341*
	Control	0.492	0.177
Product	USMES	0.334*	
	Control	0.219	
	Total	0.251*	

*p < .05

**p < .01

[a]The scoring protocol for the Picnic Problem does not yeild a product score.

indicated by the analysis of variance results presented later in this chapter.)

The correlation coefficients in Tables 6.2 and 6.3 are not satisfactory as reliability evidence for the problem solving tests because two different groups of children from a class were involved in a pair. The problem of experimental mortality suppressed the correlations even further. In some classes one or more of the children randomly selected to take part in the pre-test seven or eight months earlier had moved or was absent at post-test time.

The validity of the Playground and Picnic Problems was established only through content validation. (If a criterion was readily available, the tests need not have been developed.) The conceptual bases for both the tests and for their scoring protocols were designed to match as closely as possible the developers' viewpoints on what constitutes problem solving, as they expressed them in written materials about USMES and in conferences with the evaluation staff.

Some additional validity evidence is available in the test results themselves. From the tapes and the test administrators' notations, it can be observed that groups of children who offered solutions to the Playground and Picnic Problems did perform activities which the developers regard as elements of the problem solving process. The children identified variables or or factors to investigate; they collected data and performed calculations; they thought of various solutions and picked the one they deemed best.

Factors likely to detract from the tests' reliability and validity were anticipated in the scoring protocols for the tests. The numbers of sets of test results which contained any evidence of specific reliability or validity problems are listed in Table 6.4 for the Playground Problem and in Table 6.5 for the Picnic Problem. Although inclement weather affected 25 administrations of the Playground Problem, part of which had to be done outdoors, the scorers felt that the problems were not serious enough to warrant invalidating those tests. Other instances of problems affecting reliability and validity tended to be correlated. Thus, "outside interference" was almost always accompanied by the problem of a "noisy testing environment." And, "prompting by an observer" was certainly regarded too as a "deviation from correct procedure."

Negative evidence for the reliability and validity of a set of Playground or Picnic test data had to be substantial on one of the factors in Table 6.4 or 6.5 before the set was invalidated. More rigorous standards would have resulted in too small a data base for the analysis. Three or four sets of both Playground and Picnic test results both for USMES and for control classes had to be discounted principally because in those instances the observers deviated seriously from the scripts provided in the manuals.

Results

The scoring procedures for the Playground and Picnic Problems incorporated both behavioral and cognitive assessments. Additionally the Playground Problem yielded a product, the students' drawing of their playground design, which was rated for various features too. The analysis and results from each kind of assessment are given below in separate sections.

A. Behavioral Aspects

Rating scales were developed to assess four behavioral aspects of the students' work to the solution to a problem: (1) motivation to accept the problem; (2) commitment to task; (3) efficient allocation of responsibilities;

TABLE 6.4

Distribution of Pre-test and Post-test Ratings of Reliability/ Validity Problems for USMES and Control Classes on the Playground Problem

Nature of Problem	USMES		Control	
	Pre	Post	Pre	Post
A. Biased selection of students.	1 (3.0)[1]	0 (0.0)	0 (0.0)	0 (0.0)
B. Prompting by observer.	4 (12.1)	1 (3.0)	6 (18.2)	2 (6.1)
C. Prior student experience with test.	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
D. Inclement weather.	3 (9.1)	8 (24.2)	4 (12.1)	9 (27.3)
E. Noisy testing environment.	2 (6.1)	1 (3.0)	2 (6.1)	0 (0.0)
F. Outside interference, interruptions.	2 (6.1)	1 (3.0)	3 (9.1)	0 (0.0)
G. Observer deviated from procedure.	5 (15.2)	4 (12.1)	7 (21.2)	8 (24.2)

[1]Figures in parentheses are percentages.

TABLE 6.5

Distribution of Pre-test and Post-test Ratings of Reliability/ Validity Problems for USMES and Control Classes on the Picnic Problem

Nature of Problem	USMES		Control	
	Pre	Post	Pre	Post
A. Biased selection of students.	0 (0.0)[1]	0 (0.0)	0 (0.0)	0 (0.0)
B. Prompting by observer.	0 (0.0)	4 (12.1)	0 (0.0)	3 (9.1)
C. Prior student experience with test.	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
D. Inclement weather.	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
E. Noisy testing environment.	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
F. Outside interference, interruptions.	1 (3.0)	2 (6.1)	0 (0.0)	0 (0.0)
G. Observer deviated from procedure.	0 (0.0)	3 (9.1)	4 (11.8)	2 (5.9)

[1] Figures in parentheses are percentages of the number of groups which took the test.

and (4) style of group leadership. The same scales could be applied to both the Playground Problem and the Picnic Problem. (The actual rating stimuli and categories are described in the scoring protocols shown in Appendices F and G.)

Distributions of these ratings of the four behavioral aspects are shown in Table 6.6 for the Playground Problem and in Table 6.7 for the Picnic Problem. These tables show that almost all of the children in both USMES and control groups who received the Playground Problem were motivated at least initially to attempt to solve it, and they were only slightly less accepting of the Picnic Problem.

Since the rating on motivation to accept the problem is the number of children in a group of five who show any interest in solving the problem, the raw data on this scale can be treated meaningfully with parametric statistics. The means on this scale for the Playground Problem were, for the USMES group, pre=4.64 and post=4.82. For control classes, the pre-test mean was 4.61 and the post-test mean was 4.85. The average numbers of children who attempted the Picnic Problem were similarly high in USMES and control groups. The pre-test average for USMES classes was 4.52; the post-test average 4.58. For the control group, the pre-test mean was 4.61, the post-test mean was 4.85.

There are no noteworthy differences between USMES and control students which appeared on any of the four behavioral rating scales for either test of problem solving. Overall, both treatment groups obtained slightly higher ratings on the Playground Problem than the Picnic Problem.

B. Cognitive Aspects

The scoring protocol for cognitive aspects of the students' problem

TABLE 6.6

Distributions of Pre-test and Post-test Ratings on Behavioral Aspects of Students' Performance on the Playground Problem

Treatment Group	Test	Ratings 0	1	2	3	4	5
Motivation to Accept the Problem							
USMES (N=33)	Pre	6.1[a]				6.1	87.9
	Post	3.0				12.1	84.8
Control (N=33)	Pre		3.0		3.0	21.2	72.7
	Post					15.2	84.8
Commitment to Task							
USMES	Pre	6.1	9.1	9.1	36.4	39.4	*
	Post			9.1	30.3	60.6	*
Control	Pre			3.0	27.3	69.7	*
	Post		6.1	6.1	21.2	66.7	*
Efficient Allocation of Responsibilities							
USMES	Pre	6.1	21.2	9.1	36.4	15.2	12.1
	Post		33.3	6.1	30.3	6.1	24.2
Control	Pre		21.2	3.0	63.6	6.1	6.1
	Post		36.4	3.0	36.4		24.2
Style of Group Leadership							
USMES	Pre	30.3	12.1	12.1		45.5	*
	Post	21.2	12.1	6.1	3.0	57.6	*
Control	Pre	12.1	9.1	9.1	6.1	63.6	*
	Post	15.2	12.1	15.2	6.1	51.5	*

[a]The data are expressed as percentages of the N for that treatment group.

*The rating options for these categories were 0 through 4 only.

TABLE 6.7

Distributions of Pre-test and Post-test Ratings on Behavioral Aspects of Students' Performance on the Picnic Problem

Treatment Group	Test	Ratings					
		0	1	2	3	4	5
		Motivation to Accept the Problem					
USMES (N=33)	Pre	3.0[a]	3.0		3.0	15.2	75.8
	Post			3.0	9.1	15.2	72.7
Control (N=34)	Pre	2.9	2.9	2.9	5.9	17.6	67.6
	Post			2.9	11.8	23.5	61.8
		Commitment to Task					
USMES	Pre	3.0	6.1	3.0	21.2	66.7	*
	Post		15.2	12.1	30.3	42.4	*
Control	Pre	2.9	5.9	11.8	26.5	52.9	*
	Post		5.9	5.9	23.5	64.7	*
		Efficient Allocation of Responsibilities					
USMES	Pre	6.1	39.4	9.1	30.3	3.0	12.1
	Post		51.5	9.1	18.2	3.0	18.2
Control	Pre	2.9	50.0	11.8	35.3		
	Post		44.1		35.3	2.9	17.6
		Style of Group Leadership					
USMES	Pre	27.3	6.1	15.2	3.0	48.5	*
	Post	30.3	3.0	15.2		51.8	*
Control	Pre	26.5	14.7	17.6	35.3		*
	Post	14.7	11.8	17.6	14.7	41.2	*

[a]The data are expressed as percentages of the N for that treatment group.

*The rating options for these categories were 0 through 4 only.

solving behaviors involved coding the variables or factors which each group identified as salient to the solution of the Playground and Picnic Problems. Seven possible variables were anticipated for the Playground Problem; ten variables were anticipated for the Picnic Problem. For each test, one additional variable, "other considerations," could be accommodated. The number of factors which each group identified for consideration were summed. This sum is termed the "identification" score.

Summations for each group were made for the levels of measurement the groups achieved for each variable they identified. Similarly, summations were obtained across calculation ratings for each variable and across ratings on the adequacy of data recordings. Data analyses were based on these four summary measures: identification, measurement, calculation and recording.

Two types of analyses were conducted. First, repeat measures analyses of variance were conducted to determine if USMES and/or control classes realized statistically significant gains in any of the four cognitive summary measures. Second, analyses of covariance were used to test the hypothesis that there were no statistically significant differences in post-test differences among the groups or across grades.

1. Repeated Measures Analyses of Cognitive Scores. Source tables for the two-factor repeated measures analyses of variance using each of the four cognitive summary measures as dependent variables are shown in Tables 6.8 through 6.11 for the Playground Problem data. Corresponding results for the Picnic Problem are shown in Tables 6.12 through 6.15. Without exception, grade level differences were found to be significant at $p < .01$ on all four cognitive measures for both the Playground Problem and the Picnic Problem.

TABLE 6.8

Repeated Measures Analysis of Variance for the Playground Problem Variable, Identification

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	65	213.98	3.29	
Grade(G)	2	64.73	32.36	13.746**
Treatment(T)	1	2.73	2.73	1.162
G x T	2	5.25	2.62	1.114
Error	60	141.27	2.35	
Within Ss	66	131.50	1.99	
Tests(A)	1	18.19	18.19	10.717**
G x A	2	1.36	0.68	0.400
T x A	1	0.01	0.01	0.005
T x G x A	2	10.11	5.05	2.978
Error	60	101.84	1.70	
Total	131	345.48	2.64	

*$p < .05$

**$p < .01$

TABLE 6.9

Repeated Measures Analysis of Variance
for the Playground Problem Variable,
Measuring

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	65	745.00	11.46	
Grade (G)	2	271.98	135.99	17.813**
Treatment (T)	1	5.12	5.12	0.671
G x T	2	9.84	4.92	0.644
Error	60	458.06	7.63	
Within Ss	66	442.00	6.70	
Tests (A)	1	27.27	27.27	4.504**
G x A	2	22.18	11.09	1.832
T x A	1	0.00	0.00	0.000
T x G x A	2	29.25	14.63	2.415
Error	60	363.30	6.06	
Total	131	1187.00	9.06	

*$p < .05$

**$p < .01$

TABLE 6.10

Repeated Measures Analysis of Variance
for the Playground Problem Variable,
Calculates

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	65	549.88	8.46	
Grade(G)	2	274.29	137.15	29.918**
Treatment(T)	1	0.27	0.27	0.060
G x T	2	0.27	0.14	0.030
Error	60	275.04	4.58	
Within Ss	66	210.00	3.18	
Tests(A)	1	3.03	3.03	0.983
G x A	2	9.47	4.73	1.536
T x A	1	2.45	2.45	0.796
T x G x A	2	10.03	5.01	1.626
Error	60	185.02	3.08	
Total	131	759.88	5.80	

*p $< .05$

**p $< .01$

TABLE 6.11

Repeated Measures Analysis of Variance for the Playground Problem Variable, Records

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	65	75.22	1.16	
Grade(G)	2	23.69	11.85	14.432**
Treatment(T)	1	0.92	0.92	1.117
G x T	2	1.35	0.67	0.824
Error	60	49.26	0.82	
Within Ss	66	42.50	0.64	
Tests(A)	1	0.92	0.92	1.404
G x A	2	1.09	0.55	0.836
T x A	1	0.07	0.07	0.105
T x G x A	2	1.25	0.63	0.961
Error	60	39.17	0.65	
Total	131	117.72	0.90	

*$p < .05$
**$p < .01$

TABLE 6.12

Repeated Measures Analysis of Variance
for the Picnic Problem Variable,
Identification

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	66	402.15	6.09	
Grade (G)	2	112.36	56.18	12.315**
Treatment (T)	1	1.38	1.38	0.302
G x T	2	10.11	5.06	1.108
Error	61	278.30	4.56	
Within Ss	67	180.00	2.69	
Tests (A)	1	1.46	1.46	0.538
G x A	1	0.11	0.11	0.039
T x A	2	1.60	0.80	0.294
T x G x A	2	10.68	5.34	1.961
Error	61	166.14	2.72	
Total	133	582.15	4.38	

*$p < .05$

**$p < .01$

TABLE 6.13

Repeated Measures Analysis of Variance
for the Picnic Problem Variable,
Measuring

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	66	3263.93	49.45	
Grade (G)	2	1305.70	652.85	20.955**
Treatment (T)	1	18.79	18.79	0.603
G x T	2	38.99	19.49	0.626
Error	61	1900.44	31.15	
Within Ss	67	1099.00	16.40	
Tests (A)	1	55.19	55.19	3.436
G x A	2	44.54	22.27	1.387
T x A	1	11.19	11.19	0.697
T x G x A	2	8.32	4.16	0.259
Error	61	979.76	16.06	
Total	133	4362.93	32.80	

*$p < .05$

**$p < .01$

TABLE 6.14

Repeated Measures Analysis of Variance
for the Picnic Problem Variable,
Calculates

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	66	3546.90	53.74	
Grade (G)	2	1260.54	630.27	17.148**
Treatment (T)	1	2.65	2.65	0.072
G x T	2	41.62	20.81	0.[illegible]66
Error	61	2242.09	36.76	
Within Ss	67	991.50	14.80	
Tests (A)	1	76.13	76.13	5.389*
G x A	2	23.20	11.60	0.821
T x A	1	24.67	24.67	1.746
T x G x A	2	5.72	2.86	0.203
Error	61	861.77	14.13	
Total	133	4538.40	34.12	

*p < .05

**p < .01

TABLE 6.15

Repeated Measures Analysis of Variance
for the Picnic Problem Variable,
Records

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	66	795.43	12.05	
Grade(G)	2	349.80	174.90	24.047**
Treatment(T)	1	0.18	0.18	0.025
G x T	2	1.78	0.89	0.123
Error	61	443.67	7.27	
Within Ss	67	360.00	5.37	
Tests(A)	1	13.16	13.16	2.635
G x A	2	39.38	19.69	3.942*
T x A	1	0.33	0.33	0.066
T x G x A	2	2.41	1.21	0.241
Error	61	304.72	4.99	
Total	133	1155.43	8.69	

*$p < .05$

**$p < .01$

The means in Tables 6.16 through 6.23 reveal the nature of these grade differences. As one would expect, the higher grades achieved higher scores on these cognitive measures. The growth is very consistent on three of the four summary measures: identification, measuring, and calculating. The trend of development over grade levels is not as clear on the variable, "records," though overall grade level differences were significant for this variable too. Perhaps the results for "records" were less clear because the variable itself could not be rated as precisely as the other three cognitive measures.

To summarize other results from the repeated measures analyses of variance, significant growth from pre-test to post-test administrators was found only for the following scores: identification and measuring on the Playground Problem, and calculating on the Picnic Problem. No other F-ratios for pre-test to post-test administration differences was significant, as shown in Tables 6.8 through 6.15.

The nature of the significant differences between pre- and post-test administration in Playground factor identification scores can be observed in Table 6.16. The means reveal that both treatment groups at all grade levels (except the 7th and 8th grade control group) identified on the average, one more variable on the post-test than on the pre-test for consideration in their solution to the Playground Problem.

There were also significant differences in measurement scores from pre- to post-test administration ($p < .05$) for the Playground Problem, as Table 6.9 indicates. The means in Table 6.17 indicate that both treatment groups (except 5-6th grade USMES and 7-8th grade control) showed an increase in the average amoung of measuring done by their groups. Therefore, on the Playground

TABLE 6.16

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Identification for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	2.43 (1.50)	3.54 (2.11)	3.50 (1.22)	3.06
	Post (1.58)	3.12 (1.58)	3.92 (1.26)	5.00 (1.41)	3.82
	Adjusted	3.30	3.86	4.94	4.03
Control	N	12	13	8	33
	Pre	1.67 (1.15)	2.77 (1.09)	4.50 (1.77)	2.79
	Post	2.50 (1.00)	4.15 (1.21)	4.00 (1.31)	3.52
	Adjusted	2.70	4.20	3.80	3.56
Total for Grades	N	26	26	14	66
	Pre	2.08	3.15	4.07	
	Post	2.89	4.04	4.43	
	Adjusted	3.00	4.03	4.37	

TABLE 6.17

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Measuring for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	4.29 (2.43)	7.77 (4.36)	7.50 (1.52)	6.24
	Post	6.14 (2.57)	7.15 (2.08)	9.50 (2.07)	7.15
	Adjusted	6.33	7.02	9.39	7.58
Control	N	12	13	8	33
	Pre	3.25 (2.70)	6.54 (1.90)	8.63 (3.16)	5.85
	Post	5.25 (1.82)	7.77 (2.52)	7.38 (2.39)	6.76
	Adjusted	5.54	7.75	7.16	6.82
Total for Grades	N	26	26	14	66
	Pre	3.81	7.15	8.14	
	Post	5.73	7.46	8.29	
	Adjusted	5.93	7.39	8.28	

TABLE 6.18

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Calculations for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	1.64 (1.74)	5.15 (2.88)	5.00 (2.00)	3.64
	Post	2.93 (1.59)	4.62 (1.71)	6.33 (2.73)	4.21
	Adjusted	3.33	4.44	6.18	4.65
Control	N	12	13	8	33
	Pre	1.83 (1.99)	4.77 (1.30)	6.00 (2.62)	4.00
	Post	2.42 (1.73)	4.92 (1.75)	5.00 (1.41)	4.03
	Adjusted	2.79	4.81	4.68	4.09
Total for Grades	N	26	26	14	66
	Pre	1.73	4.96	5.57	
	Post	2.69	2.77	5.57	
	Adjusted	3.06	4.62	5.43	

TABLE 6.19

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Records for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	0.64 (0.93)	1.69 (1.18)	1.50 (0.84)	1.21
	Post	0.93 (0.62)	1.77 (0.83)	1.83 (1.17)	1.42
	Adjusted	0.98	1.73	1.81	1.51
Control	N	12	13	8	33
	Pre	0.42 (0.67)	1.23 (0.93)	1.88 (1.13)	1.09
	Post	0.83 (0.58)	1.46 (0.78)	1.38 (0.52)	1.21
	Adjusted	0.91	1.46	1.32	1.23
Total for Grades	N	24	26	14	66
	Pre	0.54	1.46	1.71	
	Post	0.88	1.62	1.57	
	Adjusted	0.94	1.59	1.56	

TABLE 6.20

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Identification for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	4.46 (1.85)	6.21 (2.46)	7.83 (1.94)	5.82
	Post	5.38 (1.71)	6.00 (1.92)	7.17 (1.47)	5.97
	Adjusted	5.69	5.94	6.77	6.13
Control	N	12	14	8	34
	Pre	4.50 (2.39)	5.86 (1.83)	6.63 (1.30)	5.56
	Post	4.17 (1.64)	6.71 (1.77)	6.75 (1.67)	5.82
	Adjusted	4.46	6.73	6.60	5.93
Total for Grades	N	25	28	14	67
	Pre	4.48	6.04	7.14	
	Post	4.80	6.36	6.93	
	Adjusted	5.07	6.33	6.69	

TABLE 6.21

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Measuring for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	9.00 (5.21)	14.50 (6.86)	19.17 (3.06)	13.18
	Post	11.69 (5.56)	14.14 (4.67)	18.00 (4.10)	13.88
	Adjusted	13.06	13.99	16.55	14.53
Control	N	12	14	8	34
	Pre	8.08 (5.11)	15.71 (4.65)	17.13 (3.04)	13.35
	Post	11.42 (4.81)	16.43 (3.03)	18.75 (3.73)	15.21
	Adjusted	13.04	15.94	17.87	15.61
Total for Grades	N	25	28	14	67
	Pre	8.56	15.11	18.00	
	Post	11.56	15.29	18.43	
	Adjusted	13.05	14.96	17.21	

TABLE 6.22

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Calculations for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	6.46 (6.10)	12.36 (6.63)	15.50 (2.88)	10.61
	Post	8.38 (5.99)	11.93 (4.36)	15.83 (4.02)	11.24
	Adjusted	10.04	11.35	14.07	11.82
Control	N	12	14	8	34
	Pre	5.00 (5.61)	12.57 (4.62)	13.13 (2.42)	10.03
	Post	8.50 (4.38)	13.57 (3.74)	16.13 (4.26)	12.38
	Adjusted	10.71	12.92	15.26	12.96
Total for Grades	N	25	28	14	67
	Pre	5.76	12.46	14.14	
	Post	8.44	12.75	16.00	
	Adjusted	10.37	12.13	14.66	

TABLE 6.23

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Records for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	2.46 (1.85)	5.86 (4.00)	7.67 (2.34)	4.85
	Post	3.15 (2.41)	4.14 (1.92)	6.17 (2.56)	4.12
	Adjusted	3.56	4.04	5.80	4.47
Control	N	12	14	8	34
	Pre	2.25 (2.83)	5.71 (2.43)	7.13 (2.64)	4.82
	Post	2.92 (2.07)	4.00 (1.84)	6.88 (1.96)	4.29
	Adjusted	3.35	3.92	6.59	4.62
Total for Grades	N	25	28	14	67
	Pre	2.36	5.79	7.36	
	Post	3.04	4.07	6.57	
	Adjusted	3.46	3.98	6.19	

Problem post-test, more variables were identified, and groups spent more time measuring, in order to help find a solution to the problem. There were no significant changes from pre- to post-test for calculation or recording scores. This result applied to both USMES and control classes.

For the Picnic Problem, no significant changes from pre-test to post-test occurred on identification or measurement summary scores. For calculation, there was a significant difference in these summary scores from pre-test to post-test administration ($p < .05$). While increases for the USMES groups at the upper grade levels were negligible, the means in Table 6.22 show that USMES classes at the lower grades (2-4) and the control groups at all grade levels averaged more pronounced pre-to-post increases.

Also from the repeated measures analysis of the Picnic Problem data, one interaction effect involving pre-post differences and grades was significant. This result is indicated by the F-ratio of 3.94 ($p < .05$) in Table 6.15 for the interaction of test administration with grade level on the variable "records." The means table for this variable, Table 6.23, shows that the lowest grades (2-4) in both treatment groups achieved slight increases in their "records" scores, but the middle and upper grades' averages on "records" declined from pre-to-post administration, perhaps because the older students were able to retain more information in memory.

2. <u>Covariance Analyses of Cognitive Scores</u>. Results of the two-way analyses of covariance used to test the significance of adjusted post-test differences between the two treatment groups are shown in Table 6.24, for Playground Problem scores and in Table 6.25 for Picnic Problem scores. Grade level was used as a factor in these two-way analyses, and it accounted for a significant portion of the total variance in each analysis. Grade

TABLE 6.24

Two Factor Analyses of Covariance for Playground Problem Scores

Source	df	Sum of Square	Mean Square	F-Ratio	P
		Identification Scores			
Treatment(T)	1	0.15	.15	0.683	.412
Grade(G)	2	16.16	8.08	4.808	.012
G x T	2	5.48	2.74	1.632	.204
Error	59	91.12	1.68		
Total	64	113.91	1.78		
		Measuring Scores			
Treatment(T)	1	2.10	2.10	0.403	.528
Grade(G)	2	35.46	17.73	3.408	.040
G x T	2	20.26	10.13	1.947	.152
Error	59	306.80	5.20		
Total	64	364.62	5.70		
		Calculate Scores			
Treatment(T)	1	0.96	0.96	0.312	.579
Grade(G)	2	33.70	16.85	5.473	.008
G x T	2	8.02	4.01	1.301	.280
Error	59	181.72	3.08		
Total	64	224.40	3.51		
		Records Scores			
Treatment(T)	1	0.66	0.66	1.211	.276
Grade(G)	2	5.02	2.51	4.567	.014
G x T	2	0.40	0.20	0.360	
Error	59	32.45	0.55		
Total	64	38.53	0.60		
		Products Scores			
Treatment(T)	1	1.83	1.83	0.186	.668
Grade(G)	2	101.38	50.69	5.136	.009
G x T	2	11.84	5.92	0.600	.552
Error	59	582.33	9.87		
Total	64	697.38	10.89		

TABLE 6.25

Two Factor Analyses of Covariance for
Picnic Problem Scores

Source	df	Sum of Square	Mean Square	F-Ratio	P
		Identification Scores			
Treatment(T)	1	0.14	0.14	0.049	.825
Grade(G)	2	23.68	11.84	4.112	.021
G x T	2	13.36	6.68	2.320	.107
Error	60	172.80	2.88		
Total	65	209.98	3.23		
		Measuring Scores			
Treatment(T)	1	27.42	27.42	1.514	.228
Grade(G)	2	103.32	51.66	2.853	.066
G x T	2	12.98	6.49	0.358	.701
Error	60	1086.60	18.11		
Total	64	1230.32	19.22		
		Calculate Scores			
Treatment(T)	1	30.78	30.78	1.790	.186
Grade(G)	2	120.88	60.44	3.515	.036
G x T	2	2.64	1.32	0.077	.927
Error	60	1032.00	17.20		
Total	65	1186.30	18.25		
		Records Scores			
Treatment(T)	1	0.52	0.52	0.122	.728
Grade(G)	2	56.38	28.19	6.575	.003
G x T	2	2.48	1.24	0.289	.750
Error	60	257.40	4.29		
Total	65	316.78	4.87		

differences were consistently significant for all cognitive summary scores on both the Playground and Picnic Problems. (The grade effect was also significant in the analysis of covariance of product scores from the Playground Problem.)

Despite statistical adjustments for pre-test differences in the cognitive summary scores derived from the Playground and Picnic Problems, Tables 6.24 and 6.25 show that no significant differences were found between USMES and control treatment groups for any of the dependent cognitive variables. Nor were there any significant interaction effects between treatment and grade level on any of these measures.

3. <u>Specific Factors Which Students Considered in Their Solutions to the Playground and Picnic Problems.</u> The repeated measures and covariance analyses of cognitive scores from the Playground and Picnic Problems reported in the preceeding sections were based on summary scores rather than on the ratings for specific factors derived from the scoring protocol. Thus, composite scores for identification, measuring, calculating, and recording were obtained for each group by summing the ratings they received on these aspects for each factor. However, the specific factors which the groups took into consideration were obscured by the summations, and the identity of these factors may be of interest.

The percentages of USMES and control classes which considered each of seven specific factors in their solution to the Playground Problem are shown in Table 6.26. Cost of equipment was an obvious consideration for most classes working on the problem, and a majority of the classes also considered placement of equipment. Tables 6.27 and 6.28 contain the percentages of classes which proceeded with measuring and calculating on the basis of any of the seven factors they took into consideration for their solutions to the Playground

TABLE 6.15

Repeated Measures Analysis of Variance for the Picnic Problem Variable, Records

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	66	795.43	12.05	
Grade (G)	2	349.80	174.90	24.047**
Treatment (T)	1	0.18	0.18	0.025
G x T	2	1.78	0.89	0.123
Error	61	443.67	7.27	
Within Ss	67	360.00	5.37	
Tests (A)	1	13.16	13.16	2.635
G x A	2	39.38	19.69	3.942*
T x A	1	0.33	0.33	0.066
T x G x A	2	2.41	1.21	0.241
Error	61	304.72	4.99	
Total	133	1155.43	8.69	

*$p < .05$

**$p < .01$

The means in Tables 6.16 through 6.23 reveal the nature of these grade differences. As one would expect, the higher grades achieved higher scores on these cognitive measures. The growth is very consistent on three of the four summary measures: identification, measuring, and calculating. The trend of development over grade levels is not as clear on the variable, "records," though overall grade level differences were significant for this variable too. Perhaps the results for "records" were less clear because the variable itself could not be rated as precisely as the other three cognitive measures.

To summarize other results from the repeated measures analyses of variance, significant growth from pre-test to post-test administrators was found only for the following scores: identification and measuring on the Playground Problem, and calculating on the Picnic Problem. No other F-ratios for pre-test to post-test administration differences was significant, as shown in Tables 6.8 through 6.15.

The nature of the significant differences between pre- and post-test administration in Playground factor identification scores can be observed in Table 6.16. The means reveal that both treatment groups at all grade levels (except the 7th and 8th grade control group) identified on the average, one more variable on the post-test than on the pre-test for consideration in their solution to the Playground Problem.

TABLE 6.16

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Identification for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	2.43 (1.50)	3.54 (2.11)	3.50 (1.22)	3.06
	Post (1.58)	3.12 (1.58)	3.92 (1.26)	5.00 (1.41)	3.82
	Adjusted	3.30	3.86	4.94	4.03
Control	N	12	13	8	33
	Pre	1.67 (1.15)	2.77 (1.09)	4.50 (1.77)	2.79
	Post	2.50 (1.00)	4.15 (1.21)	4.00 (1.31)	3.52
	Adjusted	2.70	4.20	3.80	3.56
Total for Grades	N	26	26	14	66
	Pre	2.08	3.15	4.07	
	Post	2.89	4.04	4.43	
	Adjusted	3.00	4.03	4.37	

TABLE 6.17

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Measuring for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	4.29 (2.43)	7.77 (4.36)	7.50 (1.52)	6.24
	Post	6.14 (2.57)	7.15 (2.08)	9.50 (2.07)	7.15
	Adjusted	6.33	7.02	9.39	7.58
Control	N	12	13	8	33
	Pre	3.25 (2.70)	6.54 (1.90)	8.63 (3.16)	5.85
	Post	5.25 (1.82)	7.77 (2.52)	7.38 (2.39)	6.76
	Adjusted	5.54	7.75	7.16	6.82
Total for Grades	N	26	26	14	66
	Pre	3.81	7.15	8.14	
	Post	5.73	7.46	8.29	
	Adjusted	5.93	7.39	8.28	

TABLE 6.18

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Calculations for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	1.64 (1.74)	5.15 (2.88)	5.00 (2.00)	3.64
	Post	2.93 (1.59)	4.62 (1.71)	6.33 (2.73)	4.21
	Adjusted	3.33	4.44	6.18	4.65
Control	N	12	13	8	33
	Pre	1.83 (1.99)	4.77 (1.30)	6.00 (2.62)	4.00
	Post	2.42 (1.73)	4.92 (1.75)	5.00 (1.41)	4.03
	Adjusted	2.79	4.81	4.68	4.09
Total for Grades	N	26	26	14	66
	Pre	1.73	4.96	5.57	
	Post	2.69	2.77	5.57	
	Adjusted	3.06	4.62	5.43	

TABLE 6.19

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Records for the Playground Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	14	13	6	33
	Pre	0.64 (0.93)	1.69 (1.18)	1.50 (0.84)	1.21
	Post	0.93 (0.62)	1.77 (0.83)	1.83 (1.17)	1.42
	Adjusted	0.98	1.73	1.81	1.51
Control	N	12	13	8	33
	Pre	0.42 (0.67)	1.23 (0.93)	1.88 (1.13)	1.09
	Post	0.83 (0.58)	1.46 (0.78)	1.38 (0.52)	1.21
	Adjusted	0.91	1.46	1.32	1.23
Total for Grades	N	24	26	14	66
	Pre	0.54	1.46	1.71	
	Post	0.88	1.62	1.57	
	Adjusted	0.94	1.59	1.56	

TABLE 6.20

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Identification for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	4.46 (1.85)	6.21 (2.46)	7.83 (1.94)	5.82
	Post	5.38 (1.71)	6.00 (1.92)	7.17 (1.47)	5.97
	Adjusted	5.69	5.94	6.77	6.13
Control	N	12	14	8	34
	Pre	4.50 (2.39)	5.86 (1.83)	6.63 (1.30)	5.56
	Post	4.17 (1.64)	6.71 (1.77)	6.75 (1.67)	5.82
	Adjusted	4.46	6.73	6.60	5.93
Total for Grades	N	25	28	14	67
	Pre	4.48	6.04	7.14	
	Post	4.80	6.36	6.93	
	Adjusted	5.07	6.33	6.69	

TABLE 6.21

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Measuring for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	9.00 (5.21)	14.50 (6.86)	19.17 (3.06)	13.18
	Post	11.69 (5.56)	14.14 (4.67)	18.00 (4.10)	13.88
	Adjusted	13.06	13.99	16.55	14.53
Control	N	12	14	8	34
	Pre	8.08 (5.11)	15.71 (4.65)	17.13 (3.04)	13.35
	Post	11.42 (4.81)	16.43 (3.03)	18.75 (3.73)	15.21
	Adjusted	13.04	15.94	17.87	15.61
Total for Grades	N	25	28	14	67
	Pre	8.56	15.11	18.00	
	Post	11.56	15.29	18.43	
	Adjusted	13.05	14.96	17.21	

TABLE 6.22

Treatment Group Means and Standard Deviations by Grade Level for Pre-test and Post-test Ratings on Calculations for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	6.46 (6.10)	12.36 (6.63)	15.50 (2.88)	10.61
	Post	8.38 (5.99)	11.93 (4.36)	15.83 (4.02)	11.24
	Adjusted	10.04	11.35	14.07	11.82
Control	N	12	14	8	34
	Pre	5.00 (5.61)	12.57 (4.62)	13.13 (2.42)	10.03
	Post	8.50 (4.38)	13.57 (3.74)	16.13 (4.26)	12.38
	Adjusted	10.71	12.92	15.26	12.96
Total for Grades	N	25	28	14	67
	Pre	5.76	12.46	14.14	
	Post	8.44	12.75	16.00	
	Adjusted	10.37	12.13	14.66	

TABLE 6.23

Treatment Group Means and Standard Deviations by Grade Level
for Pre-test and Post-test Ratings on Records
for the Picnic Problem

Treatment Groups		Grade Levels 2-4	5-6	7-8	Total for Treatments
USMES	N	13	14	6	33
	Pre	2.46 (1.85)	5.86 (4.00)	7.67 (2.34)	4.85
	Post	3.15 (2.41)	4.14 (1.92)	6.17 (2.56)	4.12
	Adjusted	3.56	4.04	5.80	4.47
Control	N	12	14	8	34
	Pre	2.25 (2.83)	5.71 (2.43)	7.13 (2.64)	4.82
	Post	2.92 (2.07)	4.00 (1.84)	6.88 (1.96)	4.29
	Adjusted	3.35	3.92	6.59	4.62
Total for Grades	N	25	28	14	67
	Pre	2.36	5.79	7.36	
	Post	3.04	4.07	6.57	
	Adjusted	3.46	3.98	6.19	

Problem post-test, more variables were identified, and groups spent more time measuring, in order to help find a solution to the problem. There were no significant changes from pre- to post-test for calculation or recording scores. This result applied to both USMES and control classes.

For the Picnic Problem, no significant changes from pre-test to post-test occurred on identification or measurement summary scores. For calculation, there was a significant difference in these summary scores from pre-test to post-test administration ($p < .05$). While increases for the USMES groups at the upper grade levels were negligible, the means in Table 6.22 show that USMES classes at the lower grades (2-4) and the control groups at all grade levels averaged more pronounced pre-to-post increases.

Also from the repeated measures analysis of the Picnic Problem data, one interaction effect involving pre-post differences and grades was significant. This result is indicated by the F-ratio of 3.94 ($p < .05$) in Table 6.15 for the interaction of test administration with grade level on the variable "records." The means table for this variable, Table 6.23, shows that the lowest grades (2-4) i[illegible]h treatment groups achieved slight increases in their "records" scores, but the middle and upper grades' averages on "records" declined from pre-to-post administration, perhaps because the older students were able to retain more information in memory.

2. Covariance Analyses of Cognitive Scores. Results of the two-way

TABLE 6.24

Two Factor Analyses of Covariance for
Playground Problem Scores

Source	df	Sum of Square	Mean Square	F-Ratio	P
		Identification Scores			
Treatment(T)	1	0.15	.15	0.683	.412
Grade(G)	2	16.16	8.08	4.808	.012
G x T	2	5.48	2.74	1.632	.204
Error	59	91.12	1.68		
Total	64	113.91	1.78		
		Measuring Scores			
Treatment(T)	1	2.10	2.10	0.403	.528
Grade(G)	2	35.46	17.73	3.408	.040
G x T	2	20.26	10.13	1.947	.152
Error	59	306.80	5.20		
Total	64	364.62	5.70		
		Calculate Scores			
Treatment(T)	1	0.96	0.96	0.312	.579
Grade(G)	2	33.70	16.85	5.473	.008
G x T	2	8.02	4.01	1.301	.280
Error	59	181.72	3.08		
Total	64	224.40	3.51		
		Records Scores			
Treatment(T)	1	0.66	0.66	1.211	.276
Grade(G)	2	5.02	2.51	4.567	.014
G x T	2	0.40	0.20	0.360	
Error	59	32.45	0.55		
Total	64	38.53	0.60		
		Products Scores			
Treatment(T)	1	[illegible]	[illegible]	[illegible]	[illegible]

TABLE 6.25

Two Factor Analyses of Covariance for
Picnic Problem Scores

Source	df	Sum of Square	Mean Square	F-Ratio	P
		Identification Scores			
Treatment(T)	1	0.14	0.14	0.049	.825
Grade(G)	2	23.68	11.84	4.112	.021
G x T	2	13.36	6.68	2.320	.107
Error	60	172.80	2.88		
Total	65	209.98	3.23		
		Measuring Scores			
Treatment(T)	1	27.42	27.42	1.514	.228
Grade(G)	2	103.32	51.66	2.853	.066
G x T	2	12.98	6.49	0.358	.701
Error	60	1086.60	18.11		
Total	64	1230.32	19.22		
		Calculate Scores			
Treatment(T)	1	30.78	30.78	1.790	.186
Grade(G)	2	120.88	60.44	3.515	.036
G x T	2	2.64	1.32	0.077	.927
Error	60	1032.00	17.20		
Total	65	1186.30	18.25		
		Records Scores			
Treatment(T)	1	0.52	0.52	0.122	.728
Grade(G)	2	56.38	28.19	6.575	.003
G x T	2	2.48	1.24	0.289	.750
Error	60	257.40	4.29		
Total	65	316.78	4.87		

differences were consistently significant for all cognitive summary scores on both the Playground and Picnic Problems. (The grade effect was also significant in the analysis of covariance of product scores from the Playground Problem.)

Despite statistical adjustments for pre-test differences in the cognitive summary scores derived from the Playground and Picnic Problems, Tables 6.24 and 6.25 show that no significant differences were found between USMES and control treatment groups for any of the dependent cognitive variables. Nor were there any significant interaction effects between treatment and grade level on any of these measures.

3. Specific Factors Which Students Considered in Their Solutions to the Playground and Picnic Problems. The repeated measures and covariance analyses of cognitive scores from the Playground and Picnic Problems reported in the preceeding sections were based on summary scores rather than on the ratings for specific factors derived from the scoring protocol. Thus, composite scores for identification, measuring, calculating, and recording were obtained for each group by summing the ratings they received on these aspects for each factor. However, the specific factors which the groups took into consideration were obscured by the summations, and the identity of these factors may be of interest.

The percentages of USMES and control classes which considered each of

TABLE 6.26.

Distribution of Pre-test and Post-test Ratings of Identification of Selected Variables for USMES and Control Classes on the Playground Problem

Variable	USMES		Control	
	Pre	Post	Pre	Post
A. Cost of equipment.	29 (87.9)[1]	33 (100.0)	28 (84.8)	32 (100.0)
B. Size of equipment vs. size of children.	10 (30.3)	14 (42.4)	9 (27.3)	18 (54.5)
C. Size of equipment vs. area available.	20 (60.6)	16 (48.5)	15 (45.5)	12 (36.4)
D. Capacity of equipment.	6 (18.2)	4 (12.1)	3 (9.1)	8 (24.2)
E. Durability of equipment.	1 (3.0)	7 (21.2)	3 (9.1)	3 (9.1)
F. Placement of equipment. for safety considerations.	16 (48.5)	18 (54.5)	13 (39.4)	21 (63.6)
G. Placement of equipment for efficient utilization of area.	19 (57.6)	24 72.7)	18 (54.5)	15 (45.5)
H. Other considerations.	0 (0.0)	10 (30.3)	3 (9.1)	7 (21.2)

[1]Figures in parentheses are percentages of the number of classes tested which identified the variable.

TABLE 6.27

Distribution of Pre-test and Post-test Ratings of Measuring[1] of Selected Variables for USMES and Control Classes on the Playground Problem

Variable	USMES Pre	USMES Post	Control Pre	Control Post
A. Cost of Equipment.	25 (78.8)[2]	31 (93.9)	25 (78.8)	30 (90.9)
B. Size of equipment vs. size of children.	0 (0.0)	0 (0.0)	0 (0.0)	1 (3.0)
C. Size of equipment vs. area available.	17 (51.5)	10 (30.4)	11 (33.3)	4 (12.1)
D. Capacity of equipment.	1 (3.0)	0 (0.0)	1 (3.0)	1 (3.0)
E. Durability of equipment.	1 (3.0)	5 (15.2)	1 (3.0)	3 (9.1)
F. Placement of equipment for safety considerations.	7 (21.2)	7 (21.2)	4 (12.1)	6 (18.2)
G. Placement of equipment for efficient utilization of area.	4 (12.1)	5 (15.2)	4 (12.1)	3 (9.1)
H. Other considerations.	0 (0.0)	10 (30.3)	9 (27.3)	14 (42.4)

[1] Measuring means were rated 2 or above with the scoring protocol. (See Appendix F.)

[2] Figures in parentheses are percentages.

TABLE 6.28

Distribution of Pre-test and Post-test Ratings of Calculations[1] on Selected Variables for USMES and Control Classes on the Playground Problem

Variable	USMES		Control	
	Pre	Post	Pre	Post
A. Cost of equipment.	23 (69.9)[2]	30 (90.9)	26 (78.8)	28 (84.8)
B. Size of equipment vs. size size of children.	0 (0.0)	0 (0.0)	0 (0.0)	1 (3.0)
C. Size of equipment vs. area area available.	4 (12.1)	4 (12.1)	5 (15.1)	1 (3.0)
D. Capacity of equipment.	1 (3.0)	0 (0.0)	1 (3.0)	1 (3.0)
E. Durability of equipment.	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
F. Placement of equipment for safety considerations.	0 (0.0)	1 (3.0)	0 (0.0)	0 (0.0)
G. Placement of equipment for efficient utilization of area.	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
H. Other considerations.	0 (0.0)	0 (0.0)	1 (3.0)	0 (0.0)

[1]Calculates means was rated 2 or above with the scoring protocol. (See Appendix F.)

[2]Figures in parentheses are percentages.

Problem. Only ratings of two or higher from the scoring protocol were considered as evidence of measuring and calculating for the specific factors. The large majority of classes tested became involved in measuring and calculating for data on the cost of playground equipment. Measuring the size of equipment for the area available, and measuring for placement of equipment occurred less frequently. Few additional calculations were performed, as Table 6.28 indicates.

Comparable distributions of raw scores from the Pcinic Problem are shown in Tables 6.29, 6.30, and 6.31. The students identified several specific factors as relevant to their solutions of the Picnic Problem, and virtually all of the classes considered the cost of food! The cost of admission to parks and weighing this cost against the cost of food were identified with only slightly lesser frequency by the classes. Many of the classes also considered the time available for the picnic, the distance to parks, and the size of park facilities in making their selections.

Measuring and calculating for specific factors in the solution to the Picnic Problem occurred somewhat less frequently than the mere identification of any relevant factors. However, the variables identified most frequently were the variables on which measurements and calculations were performed more frequently.

No tables are included which show ratings for recording on specific factors in the Playground and Picnic Problems. The "records" feature of students' consideration and study of a factor in the solution to a problem could not be rated carefully; for some variables, recording was not appropriate; and older classes recorded less frequently, perhaps because they did not need to rely on written references as much as younger students.

TABLE 6.29

Distribution of Pre-test and Post-test Ratings of Identification of Selected Variables for USMES and Control Classes on the Picnic Problem

Variable	USMES		Control	
	Pre	Post	Pre	Post
A. Cost of Admission.	31 (93.9)[1]	32 (97.0)	31 (91.2)	32 (94.1)
B. Cost of food.	33 (100.0)	30 (90.0)	33 (97.1)	33 (97.1)
C. Cost of food vs. cost of admission.	27 (81.8)	30 (90.9)	30 (88.2)	33 (97.1)
D. Time available for picnic.	19 (57.6)	20 (60.6)	16 (47.1)	18 (52.9)
E. Travel time vs. play-time.	12 (36.4)	12 (36.4)	12 (35.3)	12 (35.3)
F. Relation of food and admission costs and time.	8 (24.2)	8 (24.2)	6 (17.6)	9 (26.5)
G. Distance to parks.	21 (63.6)	27 (81.8)	21 (61.8)	22 (64.7)
H. Size of facilities.	19 (57.6)	23 (69.7)	19 (55.9)	25 (73.5)
I. Play equipment.	7 (21.2)	3 (9.1)	7 20.6)	1 (2.9)
J. Safety considerations for trip.	4 (21.1)	5 (15.2)	5 (14.7)	7 (20.6)
K. Other.	10 (30.3)	7 (21.2)	8 (23.5)	6 (17.6)

[1]Figures in parentheses are percentages.

TABLE 6.30

Distribution of Pre-test and Post-test Ratings of Measuring[1] of Selected Variables for USMES and Control Classes on the Picnic Problem

Variable	USMES		Control	
	Pre	Post	Pre	Post
A. Cost of admission.	23 (69.7)[2]	28 (84.9)	27 (82.3)	31 (94.1)
B. Cost of food.	27 (81.9)	28 (84.9)	28 (82.3)	32 (97.1)
C. Cost of food vs. cost of admission.	20 (60.6)	25 (75.7)	23 (67.6)	30 (88.2)
D. Time available for picnic.	13 (39.4)	8 (24.2)	11 (32.4)	14 (41.2)
E. Travel time vs. playtime.	7 (21.2)	3 (9.1)	10 (29.4)	4 (11.8)
F. Relation of food and admission cost and time.	3 (9.1)	3 (9.1)	2 (5.9)	2 (5.9)
G. Distance to parks.	18 (54.6)	15 (45.4)	14 (41.2)	13 (38.2)
H. Size of facilities.	1 (3.0)	3 (9.1)	0 (0.0)	2 (5.9)
I. Play equipment.	1 (3.0)	1 (3.0)	1 (2.9)	1 (2.9)
J. Safety considerations for trip.	1 (3.0)	5 (15.2)	2 (5.9)	3 (8.8)
K. Other.	0 (0.0)	3 (9.1)	5 (14.7)	0 (0.0)

[1]Measures here means rated 2 or above with the scoring protocol. (See Appendix G.)

[2]Figures in parentheses are percentages.

TABLE 6.31

Distribution of Pre-test and Post-test Ratings of Calculations[1] of Selected Variables for USMES and Control Classes on the Picnic Problem

Variable	USMES		Control	
	Pre	Post	Pre	Post
A. Cost of admission.	22 (66.7)[2]	27 (81.8)	24 (70.6)	31 (91.2)
B. Cost of food.	21 (63.6)	28 (84.8)	27 (79.5)	30 (88.2)
C. Cost of food vs. cost of admission.	21 (63.6)	25 (75.7)	24 (70.6)	30 (88.2)
D. Time available for picnic.	7 (21.2)	3 (9.1)	7 (20.6)	6 (17.6)
E. Travel time vs. play-time.	5 (15.2)	2 (6.1)	3 (8.8)	4 (11.8)
F. Relation of food and admission costs and time.	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
G. Distance to parks.	15 (45.4)	14 (42.5)	11 (32.3)	13 (38.3)
H. Size of facilities.	1 (3.0)	5 (15.2)	0 (0.0)	1 (2.9)
I. Play equipment.	1 (3.0)	0 (0.0)	1 (2.9)	0 (0.0)
J. Safety considerations for trip.	1 (3.0)	0 (0.0)	2 (5.9)	3 (8.8)
K. Other.	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

[1]Calculation here means rated 2 or above with the scoring protocol. (See Appendix G.)

[2]Figures in parentheses are percentages.

C. Product Aspects

The Playground Problem yielded a product which could be analyzed--the students' drawing of their playground design. (No comparable product resulted from the Picnic Problem.) The playground drawings were rated on four features. Higher scores were awarded (1) if a plan was drawn to scale; (2) if labels were present; (3) if landmarks were included; and (4) if the play area was delimited. Percentage distributions of pre-test and post-test ratings for students' drawings of the playground designs are shown in Table 6.32.

Chi squares were computed to determine if there were significant differences among treatment groups on any of the pre-test or post-test product ratings. None of the chi square results was significant. Summary product scores were formed by adding together the four product ratings for each group of five students working together on the Playground Problem. Analysis of covariance of the summary product scores reported in Table 6.24 did reveal significant grade differences but no significant treatment differences. The significant grade effect occurred because older students more frequently than younger students incorporated the four product features into their drawings of the plans for their playground design.

Chapter Summary

An objective assessment of proof of concept of the USMES curriculum was limited by the limited state of the art of measuring the problem solving abilities of elementary school children. As the evaluation team pursued a two-fold thrust of (1) program evaluation and (2) new instrument development, we applied the most satisfactory existing measures of problem solving to answer immediate needs shared by the developers and the funding agency about the progress of USMES students in real, complex problem solving.

TABLE 6.32

Distributions of Pre-test and Post-test Ratings for Students' Drawings of the Playground Designs

Treatment Group	Test	Ratings 0	1	2
		Drawn to Scale		
USMES (N=33)	Pre	81.8	12.1	6.1
	Post	84.8	12.1	3.0
Control (N=33)	Pre	81.8	15.2	3.0
	Post	90.9	9.1	0.0
		Labels Present		
USMES	Pre	48.5	51.5	*
	Post	42.4	57.6	*
Control	Pre	57.6	42.4	*
	Post	42.4	57.6	*
		Landmarks Included		
USMES	Pre	72.7	15.2	12.1
	Post	78.8	18.2	3.0
Control	Pre	75.8	18.2	6.1
	Post	75.8	15.2	9.1
		Area Delimited		
USMES	Pre	57.6	42.4	*
	Post	66.7	33.3	*
Control	Pre	72.7	27.3	*
	Post	75.8	24.2	*

*This rating position was not appropriate for this feature of the students' drawings.

These measures were the Playground Problem and the Picnic Problem. The conceptual bases for these simulated, real-life-relevant problem tasks reflected John Dewey's conceptualization of the problem solving process, whose "five logically distinct steps" permeate much of the literature about USMES prepared by the USMES Central Staff.

Designed as parallel forms of one another, both problem tests are accompanied by Manual's for trained administrators' presentation of the tests to groups of five children.

The scoring protocols developed for the tests offer both cognitive and affective assessments. The cognitive scores provide indices of the students' abilities to identify, measure, calculate, and record data on factors which they think are salient to the solution of the problems. The behavioral assessments include ratings on motivation to accept the problem, commitment to task, efficiency of manpower and the nature of group leadership. For the Playground Problem only, the protocol, afforded an assessment of the students' product--their drawing of the play area design.

Neither the Playground Problem nor the Picnic Problem satisfied the program developers' concerns that these tests meet all of their criteria for "realness." Therefore, rigorous investigation of these tests' reliability and statistical validity did not seem to be warranted. Content validation of the tests as simulated measures of life-like, complex problem solving was established.

No differences between USMES and control students were noted in the behavioral aspects of their work on the problems. The four cognitive scores were subjected to repeated measures analyses of variance and to analyses of covariance. Consistently, significant differences among grade levels were observed for all four cognitive aspects of the students performance. As one might expect, the older students in both treatment groups outperformed the younger students. They identified more factors and progressed to more frequent, higher level measuring, calculating, and recording on these factors. However, no significant differences between treatment groups were found on any of the ratings derived from the scoring protocol. More than this, no consistent patterns could be seen.

CHAPTER VII

SURVEYING ATTITUDE CHANGE IN USMES AND CONTROL STUDENTS

This chapter included a discussion of the rationale for assessing the effects of USMES on students' attitudes, a description of the instrument development procedures for this assessment, and an examination of the pre- and post-test attitude scale scores of USMES and control groups.

The Importance of Attitude Assessment

The realization of cognitive objectives continues to be the primary focus of most classroom practice, while the development of positive attitudes--toward subject matter, school, teacher, education, and other referents--is hoped to be, or assumed to be, a by-product of cognitive learning. Several popular works exposed this belief, that desirable attitudes follow cognitive mastery, as an untenable assumption (Friedenburg, 1965; Holt, 1964; Silberman, 1970).

With renewed emphasis on attitudes, values, and interests as important outcomes of education, has come increased concern for evaluating affective goals and responding to legitimate pressures for social accountability (Messick, 1970; Scriven, 1966; Stake, 1970). The need to teach and aprasie affective responses was summarized by Kahn & Weiss:

> What happens in teaching situations is highly related to the affective responses acquired related to school, teachers and the subject-matter area. If desirable affective goals are to be realized as a result of the educational process, relevant fromal learning situations have to be developed and the effects of such learning experiences will have to be systematically appraised. (Kahn & Weiss, 1973, p. 760.)

Attitudes and USMES

Claims that positive affective response w... ...ny USMES students' growth in real problem solving abilities ... the written material produced by the program developers. For example, in the context of correlating the content of USMES materials with SCIS (Science Curriculum Improvement Study) materials, the USMES developers noted:

> ....That while not all of the USMES challenges are based on a science topic, the investigation of each challenge does teach the problem-solving process. The students learn skills and concepts within the context of a problem that is real to them. Thus, they can see the need for acquiring a particular skill and will learn it more willingly and quickly. When the students can immediately see the practical application of some particular skill, they will react more positively with a desire to learn the skill in question. USMES provides the bridge between the abstractions of the school curriculum and the world of the student (USMES Curriculum Correlation Guide, April, 1975; p. 23.)

Indeed, the USMES philosophy itself emphasizes the importance of task relevancy, co-operative enterprise, and intrinsic rewards in the students' learning activities. USMES "challenges" undertaken by the students must embody some valid aspect of school or community life rather than being an invented problem imposed prepackaged by the curriculum.

The USMES developers have ruled out trivial problems, puzzles, contrived situations as USMES problems. They contend that like serious problems of the adult world, USMES problems typically require combined efforts of groups of students, not just an individual student working alone. While some work may be done individually, the USMES approach requires a division of labor and an exchange of ideas--a total group effort.

Finally, an USMES problem must be practical, so that students may hold the expectation of useful accomplishment. The success and correctness of analyses are determined by the students' tests of utility, not by teacher judgements (The USMES Guide; May, 1975).

Many USMES users and others familiar with USMES have cited the students' enthusiasm for the program as one of its most important assets. This perception was corroborated by the results of our 1973-74 teacher interviews and the observations by the evaluation team on site visitations for the 1973-74 evaluation project. USMES teachers and principals noted that having learned about USMES from their peers, non-USMES students were urging their teachers to adopt an USMES unit. Our conversation with a limited number of USMES students working in Design Labs during Winter and Spring, 1974 also supported the contention that students were very enthusiastic about USMES.

However, we felt that these reports may have been prejudiced, parochial or otherwise nonrepresentative. Therefore, as part of the 1974-75 evaluation efforts, we sought more objective evidence from a wider data base, directly from students, to investigate the impact of USMES on attitudes.

What Attitudes to Measure and How

It was implicit in their oral and written discussions of USMES that the developers considered the students' attitude development in setting general goals for the program and in devising the curriculum itself. However, no explicit, direct statements about affective objectives for USMES were available prior to Fall, 1974, when the evaluation team sought to pre-test student attitudes. Therefore, we asked the developers to commit themselves to a specific list of affective goals, which they sought to achieve through USMES, which might not be complete or final, but which we could use as one basis for assessing the affective behaviors of USMES students.

In response to our request, the developers shared with us a working draft of their goals for USMES students in the areas of attitudes and values. That list follows:

- Being open to new ideas and info
- Appreciating the importance of the many facets of problem solving
- Developing self-reliance, curiosity and initiative
- Making value judgements
  - Recognizing differences in values according to age, experience, occupation, income and interests (culture, race, religion, ethnic background)
  - Recognizing that facts alone do not determine decisions, that problematic situations have no set answers
  - Recognizing core values of daily living: fair play and justice, free speech, opportunity for decision making, opportunity for self-respect, choice, right to privacy, acceptance of the life styles of the community, group identity.
- Accepting responsibility for work being done
- Participating in decision making relevant to their lives
- Learning to work cooperatively in large and small groups; recognizing the values of cooperation among individuals, group work and division of labor
- Respecting the views, thoughts, and feelings of others.

The generality of these affective goals precluded our using them to determine what specific attitudes we should measure and how. Nevertheless, the list was useful as a screening device for existing items.

The research literature and our experience with attitude assessment directed us to consider several techniques to measure attitudes: (1) observational methods, [illegible]; and (3) self-report methods like questionnaires and attitude scales. Already, we had collected general statements from teachers about their views on the impact of USMES on students' attitudes, and more precise, structured observations of students' affective behaviors were not feasible. USMES student interviews which were to be conducted later in the 1974-75 academic year included questions about their affective response to the program, but the data base was relatively small (120 students), the interview time was limited, and control comparisons could not be made.

A self-report technique was deemed most satisfactory. Of the self-report methods, Likert scales have been used most frequently by researchers because these scales are usually easier to construct than other types (Aiken, 1970, p. 554; Kahn & Weiss, 1973, p. 768-769; Shaw & Wright, 1967). The greater availability of Likert-type scales was a favorable consideration for choosing this technique. However, ease of reading, ease of interpretation, and ease of response were more impelling reasons for our choice of this method. As Aiken cautions:

> A difficulty with self-report inventories at the elementary-school level is the readability and interpretability of the attitude instrument; another problem concerns the degree of self-insight and conscientiousness with which the pupils fill out the inventory (Aiken, 1970, p. 559-560).

Not surprisingly, a review of existing attitude scales revealed that no intact scale would be appropriate for the USMES evaluation. However, the review did serve to suggest ideas on what affective areas, relevant to USMES, could be measured. The evaluation team assembled for review a large pool of potentially usable items. Concurrently, we developed a list of the salient affective issues and referents to measure. Our initial decisions about retention of items in the pool were based on two criteria: (1) whether the items were related to the USMES philosophy and approach; and (2) whether the items were statements about the "content" of USMES--problem solving and the disciplines of math, science, or social science.

The sources of the original items which we selected for pilot-testing are identified below by topic:

a. Math items were taken from scales by Aiken (1963), Dutton & Blum (1968), and Fedon (1958).

b. Science items were culled from instruments developed by Allison (1967) and by Dutton & Stephens (1963).

c. Suggestions for social studies items were found in the works of Easton & Dennis (1966) and Kehoe (1970).

d. Items tapping attitudes toward the USMES style of learning were adapted from selected scales of the Minnesota School Affect Assessment. Those scales were entitled "general school interest," "autonomy," "self-expression," "failure anxiety," "academic fellowship," "co-operation," "non-mastery," "need for direction," and "non-communication." Only one to three items from each of these MSAA scales could be considered for inclusion in the pilot-testing for the new instrument.

e. Several items from Covington's (1967) scale were thought to be useable and appropriate for measuring students' attitudes toward USMES-styled, real problem solving.

Readability was a factor which eliminated several items from consideration, especially the longer items from the Covington scale. We decided to limit the attitude testing to students in grades four through eight, and to strive for a scale which could be readily understood by most beginning fourth graders.

From the item pool, two forms were assembled which were judged to be roughly parrallel by content analysis. A reading specialist critiqued the forms for level of difficulty on vocabulary, ideational depth, and logical construction. The final versions of the pre-test forms had been revised in accordance with her suggestions. (See Appendices K and L for Forms 1 and 2 of the pre-test attitude scale.)

<u>Pre-testing</u>

For the Fall pre-testing, item sampling was accomplished by giving alternate forms of the attitude scale to every other child in each sample class. This was done to reduce the demands for testing time but retain the opportunity to collect data on a larger number of items. Thus, instead of obtaining scores for every student in the sample on every item across both forms, we obtained estimates of class means on each of the 51 items from randomly selected halves of the groups in each class.

Our trained field staff administered these scales, and they were directed to note which words, ideas, items, etc. proved to be difficult for their sample classes.

Pilot Testing

The item sampling procedure used in the pre-testing would have enabled factor analysis of the total pool of 51 attitude items only if class means were used as the unit of analysis. However, this practice would have yielded too small an N.

To study the factor structure of the scale and to examine the internal consistency reliability estimates for the factors, we decided to administer the combined forms 1 and 2 as a single instrument. The resulting instrument, "Form 0," is shown in Appendix J. It was administered to approximately 180 students in grades 4 through 6 and to another 180 students in grades 7 and 8 in an urban school system in the metropolitan Boston area. Again test administrators noted words which students found difficult. The socioeconomic levels of these students could be characterized as lower-middle to middle class. The USMES program was not used in that school system, nor were any students in the pilot test sample serving as control students.

For our exploratory study of the attitude instrument, a variety of factor analyses were run on the data. Several computing methods for factor solutions were used; both orthogonal and oblique rotations were tried. The methods were applied to the data for grades 4 through 6, to the data for grades 7 and 8, and to the total data base. The SPSS programs (Nie, Bent, and Hull; 1970) for these factor analyses were run at the Boston University Computer Center. The results of these several analysis--lengthy tables of factor loadings--have been omitted from this report. Only a summary of the most important observations from these pilot study analyses follows.

Post-testing

The post-test version, or "Form 3," of the attitude scale is shown in Appendix M. The test was administered in May, 1975 to 31 USMES and 26 control classes. Usable returns were obtained for a total of 1491 students.

Determining the Factor Structure of the Instrument

A large number of factor analysis programs and options were available to us at the Boston University Computing Center. The procedure we deemed more appropriate for our purpose was the SPSS program for the alpha factor solution (Nie, Bent, and Hull, 1970, p. 220).

In alpha factoring, variables included in the factor analysis are considered to be a sample from the universe of variables, an assumption we held in selecting items for the attitude scale. The alpha method follows the classical factor model with the basic factor postualte that variables are assumed to consist of two parts: one that is determined by common factors and one that is unique to each variable. Using alpha factoring, we sought to define factors that have maximum generalizability, the measure of which is known as Kuder-Richardson's reliability coefficient or Cronbach's α. In other words, we wanted to achieve factors with as high as possible internal consistency reliability.

Only those factors whose eigenvalues were greater than unity were retained for subsequent interpretation. Orthogonal rotations were applied.

Separate analyses were conducted on Part I and on Part II of the post-test attitude data for the entire sample of 1491 students. Part I items included a variety of statements on classroom climate, instructional strategies, and specific content areas. The items on Part II were designed to measure attitudes toward complex problem solving.

Regardless of the factor analytic procedure or the data base which was used, some of the emerging factors were very similar across analyses. This was so, most notably for the "math" and the "science" factors. For other relatively stable factors, a few items loaded on one factor in one analysis but on another factor when another method or data base was used. In no case did the items intended to measure selected attitudes toward social studies load on factors in a logically satisfying way or with any sizable loadings.

The Covington attitude items based on the engineering problem in Part II of the scale did not factor clearly, probably because these items were quite homogeneous and the sample size might have been larger for the number of items used. We noted too that the references to the engineers problem in subsequent items after the story may have been too remote for the children to associate "problem" with that particular problem.

The results of the pilot study suggested a few changes in the attitude scale. We rewrote certain items which were difficult to read. "Social studies" items were deleted because of the poor item statistics and factor structure effects with these items. In our judgement, the item referents were unclear to an elementary school child. What constitutes social studies and by what name--social science, civics, geography, etc.--would the child refer to this expansive area? More specific issues or values in this area, e.g., prejudice, democracy, politics, would not be any more meaningful to young students and could not be sampled adequately with a small number of items. A final change based on the results of the pilot study was the combination of all revised, usable items into one form which would take approximately 30 minutes for post-test administration.

Also, the alpha method of factoring was applied separately to the USMES group data and to the control group data. Essentially the same factor structure was obtained for each data base, so we decided to use the total group results to determine which items would be included in the subtest scales.

Table 7.1 shows the factors, the item loadings, and the communality estimates which emerged from the analysis of Part I data. Table 7.2 gives the results of the factor analysis of Part II of the attitude scale. Also presented in these tables are item analysis data which show Hoyt's estimate of reliability (internal consistency) and the standard error of measurement for each factor. The last column in each table contains the item-subtest total correlation for each item, i.e., the correlation between the item score and the scale score. In general, the scale scores are simply additive raw score weights for the items loading $\geq .3$ on that factor. (In a few cases, items with factor loadings slightly less than .3 were included on a scale because there was a logical fit between the item and the factor, and because including those items improved the reliability of their respective scales.) As shown in Table 7.1, factor analysis of Part I of the attitude scale yielded seven significant factors: (1) science appreciation; (2) academic insecurity; (3) non-mastery; (4) preference for group learning activities; (5) arithmetic enjoyment; (6) self-directed learning; and (7) arithmetic value. It is interesting to observe that all of the "science" items in the instrument loaded on one general factor of science appreciation, however, the "arithmetic" items split to yield two factors—enjoyment versus value. Another note is that the third factor which we

TABLE 7.1

Results of Factor Analysis and Item Analysis of Part I of the Attitude Scale

Factor Number and Name	Percent of Explained Variance	Hoyt[a] & S_e	Item	Loading	Comm.	I.T.C.
1. Science Appreciation	36.9%	H= .77 S_e=2.29	7. I wish I had more science in school.	.77396	.5601	.648
			8. I am interested in learning more about science.	.77021	.5919	.678
			9. Science is not useful for children.	-.30030	.23211	.375
			10. Studying science bores me.	-.70206	.51271	.607
			11. Science helps to improve the	.29336	.2021	.317
			12. What you learn in science is often the basis of a good hobby.	.51902	.3193	.479
2. Academic Insecurity	19.5%	H= .35 S_e=2.20	14. I don't like to talk to the whole class about my ideas.	.34041	.2731	.157
			22. I have alot of questions I n never get a chance to ask.	.29158	.1550	.209
			29. I get confused when I don't know why I'm studying some things.	.30583	.1969	.219
			31. I hate to make a mistake in class class.	.28691	.1044	.151
3. Non-mastery	13.1%	H= .38 S_e=1.15	19. I like to study lots of things, even if I don't learn them well.	.51477	.4488	.232
			28. I like to go on to new topics, even if I haven't learned much about the topics I studied before.	.36910	.1564	.232

TABLE 7.1 (Cont.)

Factor Number and Name	Percent of Explained Variance	Hoyt[a] & S_e	Item	Loading	Comm.	I.T.C.
4. Preference for Group Learning Activities	10.6%	H= .50 S_e=2.57	14. I don't like to talk to the whole class about my ideas.	.34012	.2731	.196
			15. Talking with other students in small groups is fun.	-.54105	.3147	.247
			16. I like to help other students learn.	-.32114	.2202	.236
			24. I like to talk in a small group about my ideas.	-.58695	.3560	.296
			26. I enjoy talking with other people in large groups.	-.53546	.3440	.223
			27. I like to learn by working with other students.	-.46644	.2929	.343
5. Arithmetic Enjoyment	9.8%	H= .69 S_e=1.40	1. I try not to do much arithmetic because I am not good with numbers.	-.47003	.2449	.451
			2. I enjoy arithmetic.	.62005	.3414	.530
			4. When I hear the word arithmetic, I have an unpleasant feeling.	-.58491	.3292	.540
6. Self-Directed Learning	6.1%	H= .65 S_e=2.30	13. I like to choose what I want to learn.	.64606	.3399	.494
			18. I like my teachers to tell me what I'm supposed to learn.	-.37705	.1970	.320
			21. I like to decide for myself what I study in school.	.72877	.3481	.518
			23. I prefer to choose the people I want to work with in class.	.39567	.1849	.329

TABLE 7.1 (Cont.)

Factor Number and Name	Percent of Explained Variance	Hoyt[a] & S_e	Item	Loading	Comm.	I.T.C.
6. (cont.)			25. I like to listen to the teacher talk to the whole class.	-.36708	.2651	.340
7. Arithmetic Value	4.0%	H= .50 S_e=1.33	3. Arithmetic is not useful to children.	-.48793	.1892	.307
			5. Arithmetic is as important as any other subject.	.34346	.1708	.297
			6. I won't need arithmetic when I grow up.	-.47011	.1600	.362

[a]This column contains Hoyt's estimate of reliability and the standard error of measurement for each factor.

TABLE 7.2

Results of Factor Analysis and Item Analysis of Part II of the Attitude Scale

Factor Number and Name	Percent of Explained Variance	Hoyt[a] & S_e		Item	Loading	Comm.	I.T.C.
1. Need to Please--Acquiesce	54.1%	H=.60 S_e=2.69	3.	In a problem like this one, the best answer will be the one that most of the class decides is right.	.41160	.1825	.28[illegible]
			5.	The best answer is the one that the teacher thinks is right.	.52256	.4310	.442
			6.	It is best to make sure that an idea is a good one before sharing it with the class.	.33935	.1659	.268
			10.	I don't think I should ask too many questions about problems in class.	.47772	.3018	.388
			11.	Other students know more about problems like this than I do.	.38633	.1674	.273
			12.	If I already have one good idea, I would rather stick with it than look for more ideas.	.42613	.2824	.368

TABLE 7.2 (Cont.)

Factor Number and Name	Percent of Explained Variance	Hoyt[a]		Item	Loading	Comm.[9]	I.T.C.
2. Divergent Thinking	27.0%	H = .42 S_e = 2.21	4.	An idea for solving a problem could lead to a wrong answer but still be a good idea.	.21791	.0777	.109
			8.	I would like to work on a problem like this one.	.57430	.4528	.323
			9.	If I worked on this problem, I would get alot of good ideas.	.57263	.3822	.316
			13.	I would like to work on a problem like this, even though I might not be able to solve it.	.64051	.4062	.461
			15.	I am very good at thinking and solving problems.	.35550	.2142	.261
3. Academic Competition	11.5%	H = .44 S_e = 1.00	2.	If someone gets an idea that no one else has thought of, he should keep it to himself.	.42360	.3293	.283
			14.	I think that my ideas for solving this problem would be better than ideas given by other students.	.57987	.4308	.283
4. Convergent Thinking	7.5%	H = .69 S_e = .87	1.	There is probably only one answer to a problem like this one.	.61304	.4676	.532
			7.	There is probably only one best way to solve a problem like this one.	.76670	.5890	.532

[a]This column contains Hoyt's estimate of reliability and the standard error of measurement for each factor.

titled "non-mastery" could also be an indicator of "need for change," or a "short attention span."

The factor analysis of Part II of the attitude scale yielded four factors with eigenvalues greater than one. These factors were titled as follows: (1) need to please--acquiesce; (2) divergent thinking; (3) academic competition; and (4) convergent thinking. At first glance the items on factor 1 and factor 4 of Part II all seem to deal with closure and/or conformity. However, it is the nature of the statements--personality statements on factor 1, and cognitive statements on factor 4--which is the basis for a logically satisfying separation of these two factors.

Scoring and Analysis

The class mean for each attitude factor was the unit of analysis used in the investigation of treatment group differences. To obtain these means, several procedures were followed, First, those items with negative item--subtest-total correlations were reverse scored. Second, for the post-test data, subtest totals were found for each student by adding the raw scores for items which loaded .3 or greater on the factors. Then these subtest totals were average across students in a class to get class means for each factor.

The pre-testing of the attitude scale involved item sampling through the use of two forms. Thus, to obtain the pre-test estimates of class means for attitude factor, partial subtest scores were found for S's who took Form 1, then for those who received Form 2. Class averages for the partial subtest scores were calculated separatly for Form 1 and Form 2. Lastly, the pairs of partial subtest means were summed to yield pre-test class means on each factor.

The means data were submitted to two types of analyses: (1) repeated measures analysis of variance, and (2) analysis of covariance. The repeated measures analysis was used to investigate whether classes from either treatment group realized statistically significant gains in attitudes within grade levels of interest. Pre-test to post-test administration differences had to be interpreted with caution, however, because a few revisions in wording were made on the post-test form, and because the item sampling on the pre-test necessitated a different procedure for calculating class means from that used on the post-test.

The analysis of covariance method was applied to test the hypothesis that there were no statistically significant differences in post-test scores of the USMES and control groups once adequate statistical allowances were made relative to pre-test differences between the treatment groups.

Results for Part I of the Attitude Scale

The item means and standard deviations for the statements on Part I of the attitude scale are presented in Table 7.3.

Factor analysis of the data from Part I yielded seven significant factors. The discussion of pre-to-post treatment, and grade differences on each factor is accompanied by tables for the repeated measures and covariance analyses. A table of means, by group and by grade, and a graph of these data are also presented for each factor.

The evaluators note that one must exercise caution in interpreting these class means because the attitude factor scores were derived from items which were not truly scaled. While we sought to include both positive and negative statements in the instrument, we did not determine the position on the latent attitude variable reflected by the item. Thus one should examine a group's or grade's factor mean in relation to the means for other groups or grades, and with consideration of the possible range and midpoint for

TABLE 7.3

Pre-test and Post-test Means and Standard Deviations for Items on Part I of the Attitude Scale Given by Factor for USMES and Control Groups

Item No.[a]	Item	USMES Pre	USMES Post	Control Pre	Control Post
	Factor 1: Science Appreciation				
7.	I wish I had more science in school.	3.28[b] (1.52)	3.41 (1.43	3.02 (1.54)	3.13 (1.53)
8.	I am interested in learning more about science.	3.63 (1.31)	3.74 (1.31)	3.59 (1.41)	3.47 (1.44)
9.	Science is not useful for children.*	3.59 (1.45)	4.32 (1.11)	3.59 (1.50)	4.34 (1.11)
10.	Studying science bores me.*	3.36 (1.51)	3.54 (1.41)	3.29 (1.53)	3.34 (1.50)
11.	Science helps to improve the world.	4.10 (1.19)	4.38 (0.96)	4.15 (1.25)	4.35 (1.07)
12.	What you learn in science is often the basis of a good hobby.	3.52 (1.26)	3.33 (1.15)	3.46 (1.33)	3.65 (1.21)
	Factor 2: Academic Insecurity				
14.	I don't like to talk to the whole class about my ideas.	2.82 (1.46)	3.17 (1.47)	2.88 (1.53)	3.29 (1.45)
22.	I have a lot of questions I never get a chance to ask.	3.15 (1.41)	3.41 (1.40)	3.22 (1.52)	3.43 (1.41)
29.	I get confused when I don't know why I'm studying some things.	3.69 (1.21)	3.58 (1.25)	3.75 (1.23)	3.55 (1.25)
31.	I hate to make a mistake in class.	4.12 (1.21)	3.98 (1.26)	4.17 (1.21)	4.09 (1.30)

TABLE 7.3 (Cont.)

Item No.[a]	Item	USMES Pre	USMES Post	Control Pre	Control Post
	Factor 3: Non-mastery				
19.	I like to study lots of things, even if I don't learn them well.	**	3.77 (1.30)	**	3.81 (1.27)
28.	I like to go on to new topics, even if I haven't learned much about the topics I studied before.	3.14 (1.33)	3.26 (1.33)	3.09 (1.36)	3.21 (1.36)
	Factor 4: Preference for Group Learning Activites				
14.	I don't like to talk to the whole class about my ideas.*	2.82 (1.46)	3.68 (1.36)	2.88 (1.53)	3.60 (1.40)
15.	Talking with other students in small groups is fun.	4.17 (1.17)	4.28 (1.18)	4.32 (1.09)	4.27 (1.[illegible]7)
16.	I like to help other students learn.	4.16 (1.08)	4.[illegible] (0.99)	4.18 (1.16)	4.[illegible]8 (1.[illegible]7)
24.	I like to talk in a small group about my ideas.	[illegible] (1.39)	3.6[illegible] (1.32)	3.80 (1.29)	3.70 (1.10)
26.	I enjoy talking with other people in large groups.	3.36 (1.39)	3.14 (1.36)	3.46 (1.37)	3.08 (1.38)
27.	I like to learn by working with other students.	3.99 (1.21)	4.01 (1.15)	4.13 (1.12)	4.04 (1.18)
	Factor 5: Arithmetic Enjoyment				
1.	I try not to do much arithmetic because I am not good with numbers.*	3.[illegible]9 (1.29)	4.[illegible] (1.[illegible]9)	3.77 (1.35)	4.06 (1.[illegible]9)
2.	I enjoy arithmetic.	3.77 (1.27)	3.77 (1.32)	3.60 (1.46)	3.66 (1.37)
4.	When I hear the word arithmetic, I have an unpleasant feeling.*	2.75 (1.42)	3.68 (1.36)	2.90 (1.44)	3.60 (1.40)

TABLE 7.3 (Cont.)

Item No.[a]	Item	USMES Pre	USMES Post	Control Pre	Control Post
	Factor 6: Self-directed Learning				
13.	I like to choose what I want to learn.	3.78 (1.42)	3.95 (1.29)	3.67 (1.55)	3.78 (1.42)
18.	I like my teachers to tell me what I'm supposed to learn.*	3.66 (1.39)	2.52 (1.43)	3.76 (1.42)	2.41 (1.45)
21.	I like to decide for myself what I study in school.	3.38 (1.36)	3.43 (1.39)	3.26 (1.51)	3.44 (1.44)
23.	I prefer to choose the people I want to work with in class.	4.08 (1.28)	4.09 (1.21)	4.04 (1.34)	4.00 (1.32)
25.	I like to listen to the teacher talk to the whole class.*	3.61 (1.38)	2.15 (1.22)	3.82 (1.33)	2.13 (1.26)
	Factor 7: Arithmetic Value				
3.	Arithmetic is not useful to children.*	4.31 (1.25)	4.47 (1.13)	4.32 (1.26)	4.51 (1.06)
5.	Arithmetic is as important as any other subject.	4.09 (1.24)	4.28 (1.18)	4.17 (1.26)	4.27 (1.17)
6.	I won't need arithmetic when I grow up.*	4.35 (1.20)	4.53 (0.98)	4.36 (1.17)	4.52 (1.01)

[a]Item numbers were taken from the post-test form.

[b]Standard deviations are given in parentheses.

*Items with an asterisk were reverse scores.

**This item was not included on the pre-test.

a given scale. Graphs of the means data are presented to aid this interpretation.

The results for Part I are given below by factor.

A. Factor 1: Science Appreciation

Table 7.4 contains the results of the treatment by grade analysis of variance with repeated measures on Factor 1. Grade level differences in science appreciation scores were statistically significant at $p < .05$, but these differences required qualification by treatment group, as indicated by the significant treatment X grade interaction effect ($p < .05$). This repeated measures analysis of variance also revealed a highly significant ($p < .0001$) pre-to-post test administration difference which did not depend on treatment or grade level.

When post-test scores were adjusted for pre-test differences in science appreciation scores, the analysis of covariance computed on factor 1 scores resulted in grade level differences significant at $p < .015$, and treatment group differences which approached statistical significance at $p \doteq .07$. These results are shown in Table 7.5.

The Factor 1 means printed in Table 7.6 and graphed in Figure 7.1 reveal the nature of these differences. The highly significant pre-to-post increase in science appreciation for both treatment groups at all grade levels is portrayed vividly by the graph. Initially positive expressions of science appreciation became more positive over the course of the school year.

The graph of Factor 1 means also highlights the grade level differences and the grade by treatment interaction effect which were significant in the repeated measures analysis of variance. Overall, the younger grades reported

TABLE 7.4

Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 1: Science Appreciation

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	378.96	6.77	
Grade (G)	2	54.95	27.48	5.357**
Treatment (T)	1	12.59	12.59	2.455
T x G	2	49.82	24.91	4.855*
Error	51	261.59	5.13	
Within Ss	57	116.93	2.05	
Tests (A)	1	60.61	60.61	61.072**
G x A	2	2.70	1.35	1.358
T x A	1	1.21	1.21	1.201
T x G x A	2	1.79	0.90	0.901
Error	51	50.62	0.99	
Total	113	495.88	4.39	

*$p < .05$
**$p < .01$

TABLE 7.5

Analysis of Covariance for Attitude Scores on Part I, Factor 1: Science Appreciation

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment (T)	1	4.99	4.99	3.406
Grade (G)	2	13.38	6.69	4.564*
T x G	2	2.80	1.40	0.956
Error	50	73.50	1.47	
Total	55	94.67	1.72	

*$p < .05$

TABLE 7.6

Means and Standard Deviations for Attitude Scores on Part I, Factor 1: Science Appreciation

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	21.93 (1.28)	21.25 (2.00)	21.80 (2.91)	21.55
	Post	23.85 (1.44)	22.98 (2.02)	22.94 (1.82)	23.19
	Adjusted	23.43	22.96	22.59	22.99
Control	N	7	13	6	26
	Pre	21.94 (1.93)	21.75 (1.59)	18.66 (0.29)	21.09
	Post	23.95 (1.42)	22.57 (1.45)	19.88 (0.86)	22.32
	Adjusted	23.52	22.25	21.41	22.39
Total for Grades	N	15	29	13	57
	Pre	21.93	21.47	20.35	
	Post	23.90	22.80	21.53	
	Adjusted	23.47	22.61	22.00	

Note—Possible range of Factor 1 scale score means is 6 to 30; the midpoint of the scale is 18.

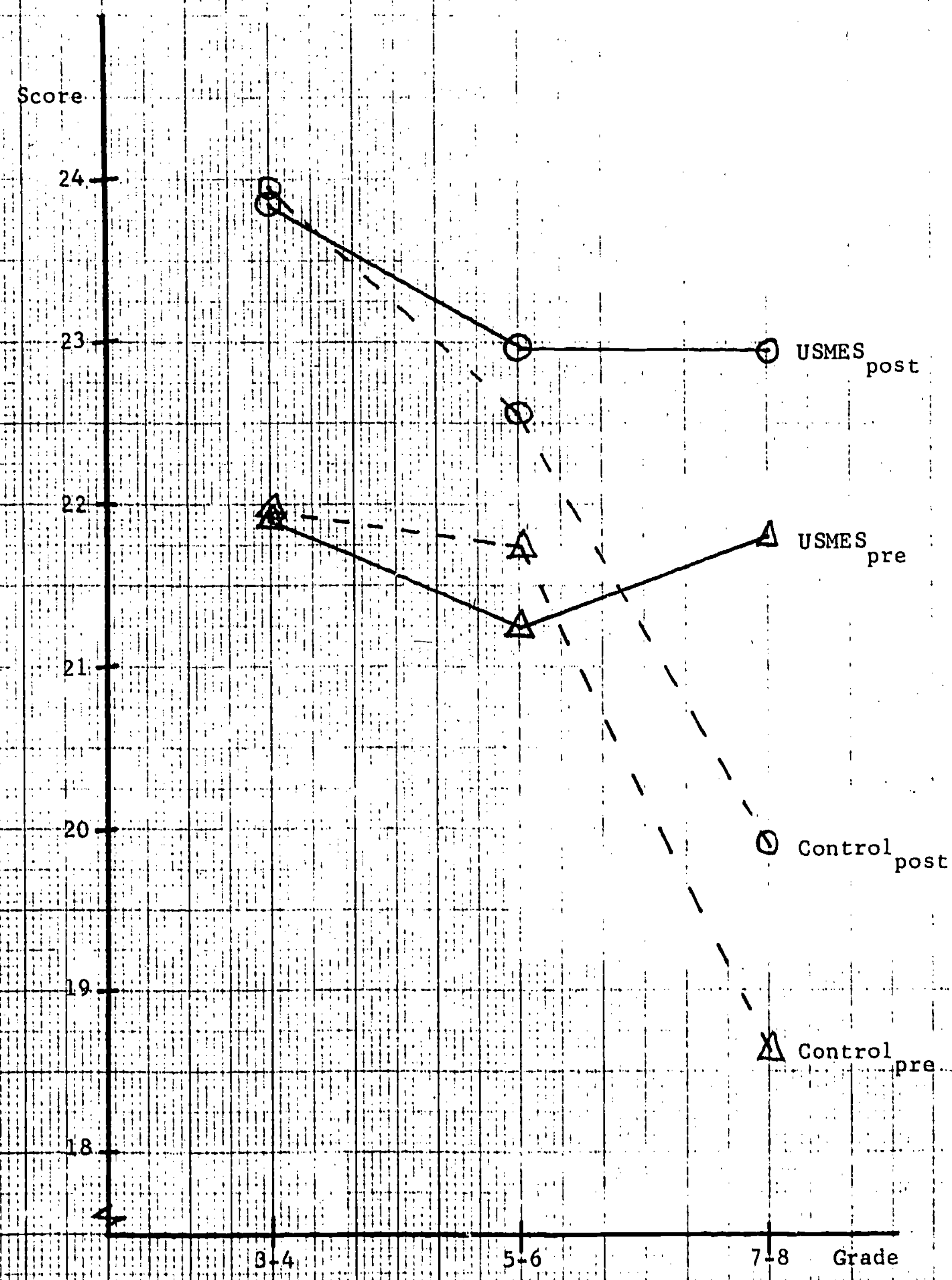

Figure 7.1-- Means on Part I, Factor 1: Science Appreciation; Scale range: 6 to 30; Scale midpoint: 18.

greater appreciation for science. However, when the grade level means collapsed across treatments were distinguished by treatment, a very different picture emerged for the USMES versus control groups at the 7th and 8th grade level. Here, the USMES group scored significantly higher than thercontrol group on both the pre- and post-test measures, though the changes for each group were in the same direction.

With the statistical adjustments for pre-test differences achieved by the analysis of covariance, the significant grade level effect persisted. but the interaction effect was no longer significant.

B. Factor 2: Academic Insecurity

Repeated measures analysis of variance and analysis of covariance programs were run on factor 2 scores. The results are shown in Tables 7.7 and 7.8 respectively. None of the F ratios in either table reached statistical significance at the .05 level.

The factor 2 means printed in Table 7.9 are presented graphically in Figure 7.2. Inspection of the graph reveals fairly flat profiles of pre- and post-test mean scores for both treatment groups at all grade levels. Only the seventh- and eighth-grade level USMES group showed any appreciable pre-to-post difference in Factor 2 scores, but this increase in academic insecurity was not statistically significant.

While none of the items were truly scaled, it is interesting to note that all cell means fell approximately two points above the supposed point of neutrality (12) on this scale of academic insecurity. Were the majority of students somewhat insecure about their school work?

TABLE 7.7

Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 2: Academic Insecurity

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	42.26	0.75	
Grade(G)	2	2.05	1.02	1.384
Treatment(T)	1	0.43	0.43	0.581
T x G	2	2.07	1.04	1.400
Error	51	37.71	0.74	
Within Ss	57	37.29	0.65	
Tests(A)	1	1.92	1.92	3.032
G x A	2	1.86	0.93	1.464
T x A	1	0.17	0.16	0.247
T x G x A	2	1.03	0.52	0.814
Error	51	32.32	0.63	
Total	113	79.55	0.70	

TABLE 7.8

Analysis of Covariance for Attitude Scores on Part I, Factor 2: Academic Insecurity

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	0.02	0.02	0.024
Grade(G)	2	3.88	1.94	2.482
T x G	2	2.88	1.44	1.843
Error	50	39.00	0.78	
Total	55	45.78	0.83	

TABLE 7.9

Means and Standard Deviations for Attitude Scores on Part I, Factor 2: Academic Insecurity

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	13.82 (0.65)	13.79 (0.73)	13.92 (0.57)	13.83
	Post	13.79 (1.01)	13.97 (0.68)	14.98 (0.64)	14.15
	Adjusted	13.80	13.98	14.98	14.25
Control	N	7	13	6	26
	Pre	14.03 (0.86)	14.06 (0.77)	13.93 (1.12)	14.02
	Post	14.38 (0.62)	14.08 (1.23)	14.28 (0.70)	14.20
	Adjusted	14.37	14.07	14.27	14.24
Total for Grades	N	15	29	13	57
	Pre	13.92	13.91	13.93	
	Post	14.06	14.02	14.65	
	Adjusted	14.08	14.02	14.63	

Note--Possible range of factor 2 scale score means is 4 to 20; the midpoint of the scale is 12.

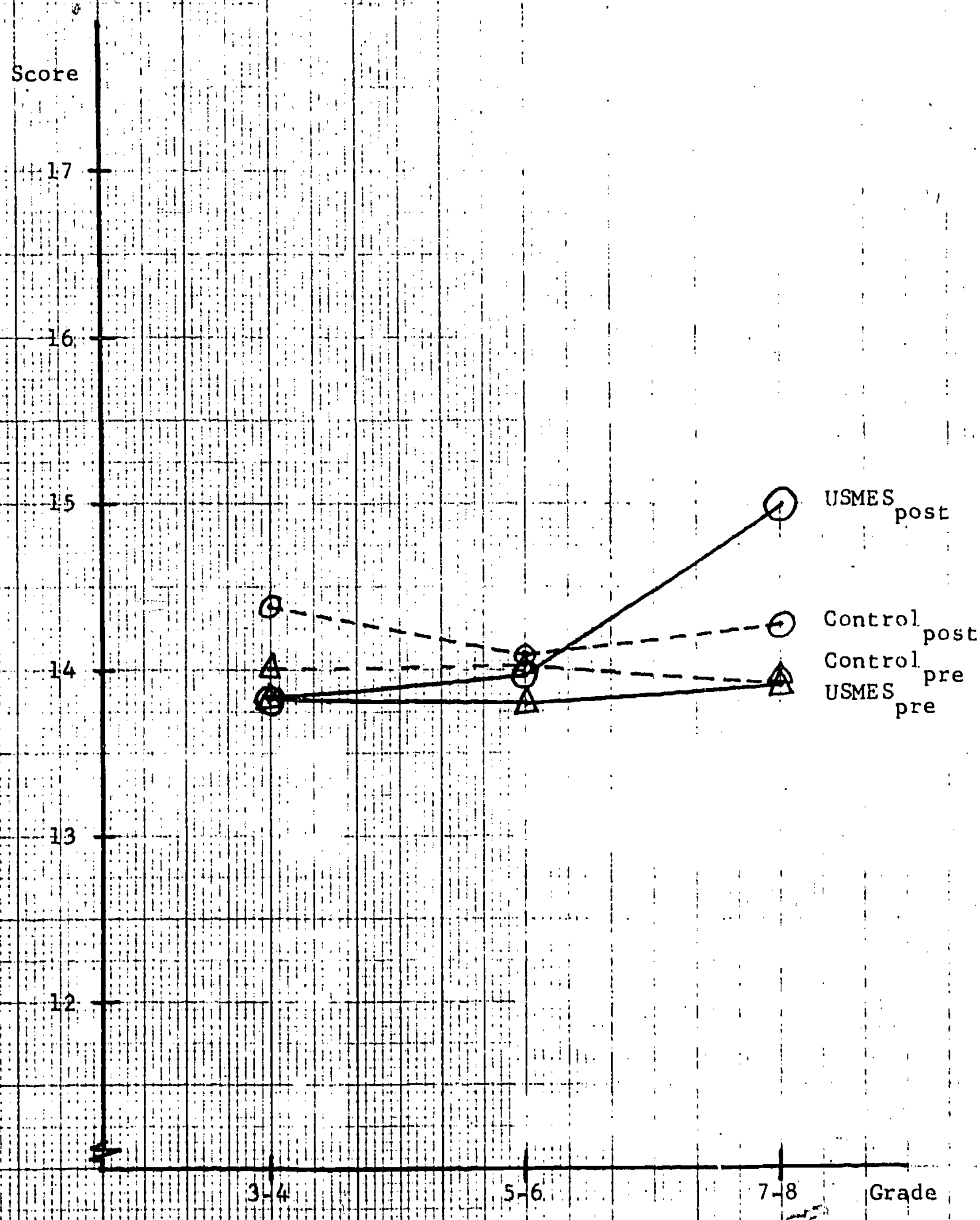

Figure 7.2-- Means on Part I, Factor 2: Academic Insecurity; Scale range: 4 to 20; Scale midpoint: 12.

C. Factor 3: Non-mastery

Post-test scores on factor 3 were based on two-items, but one of these items loading on factor 3 was inadvertently omitted from the pre-test forms. Consequently, the "pre-test" scores for non-mastery should be regarded only as estimates of pre-test position. (The pre-test estimates were determined by multipling the one item score by two.) In light of this error, the repeated measures analysis of variance to investigate the significance of pre-to-post test administration differences was misleading and is not presented here.

Since the pre-test estimates were deemed satisfactory as a covariate, an analysis of covariance for factor 3 scores was computed. The results are shown in Table 7.10. The F-ratio for grade differences was highly significant at $p < .0001$. No other effect was significant.

Factor 3 means are presented numerically in Table 7.11 and graphically in Figure 7.3. Ignoring the fairly flat profile of pre-test estimates, one should observe a significant decrease in non-mastery post-test scores from the lower to upper grade levels in both USMES and control groups. This outcome seems to indicate that the lesser need for variety and change and the increased attention spans are simply a function of maturation.

While this result does not distinguish between treatments, it does have implications for USMES usage. It has been documented elsewhere (Shann, August, 1975, Chapter III) that extended, less intensive applications of USMES units over the school year have been unsatisfactory. Students, particularly younger students, lost interest in units presented that way. Since that result was documented during 1973-74, the USMES developers began to

TABLE 7.10

Analysis of Covariance for Attitude Scores on Part I, Factor 3: Non-mastery

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	0.01	0.01	0.015
Grade(G)	2	12.92	6.46	18.548**
T x G	2	0.50	0.25	0.255
Error	50	17.50	0.35	
Total	55	30.93	0.56	

**p < .01

TABLE 7.11

Means and Standard Deviations for Attitude Scores on Part I, Factor 3: Non-mastery

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	8.37 (0.57)	8.41 (0.77)	8.02 (0.64)	8.31
	Post	7.70 (0.55)	7.04 (0.71)	6.21 (0.43)	7.02
	Adjusted	7.71	7.05	6.20	6.98
Control	N	7	13	6	26
	Pre	8.60 (0.77)	8.20 (0.30)	8.04 (0.41)	8.27
	Post	7.52 (0.57)	7.15 (0.54)	6.28 (0.49)	7.05
	Adjusted	7.54	7.14	6.26	6.98
Total for Grades	N	15	29	13	57
	Pre	8.48	8.32	8.03	
	Post	7.61	7.09	6.24	
	Adjusted	7.62	7.10	6.23	

Note--Possible range of factor 3 scale score is 2 to 10; the midpoint of the scale is 6.

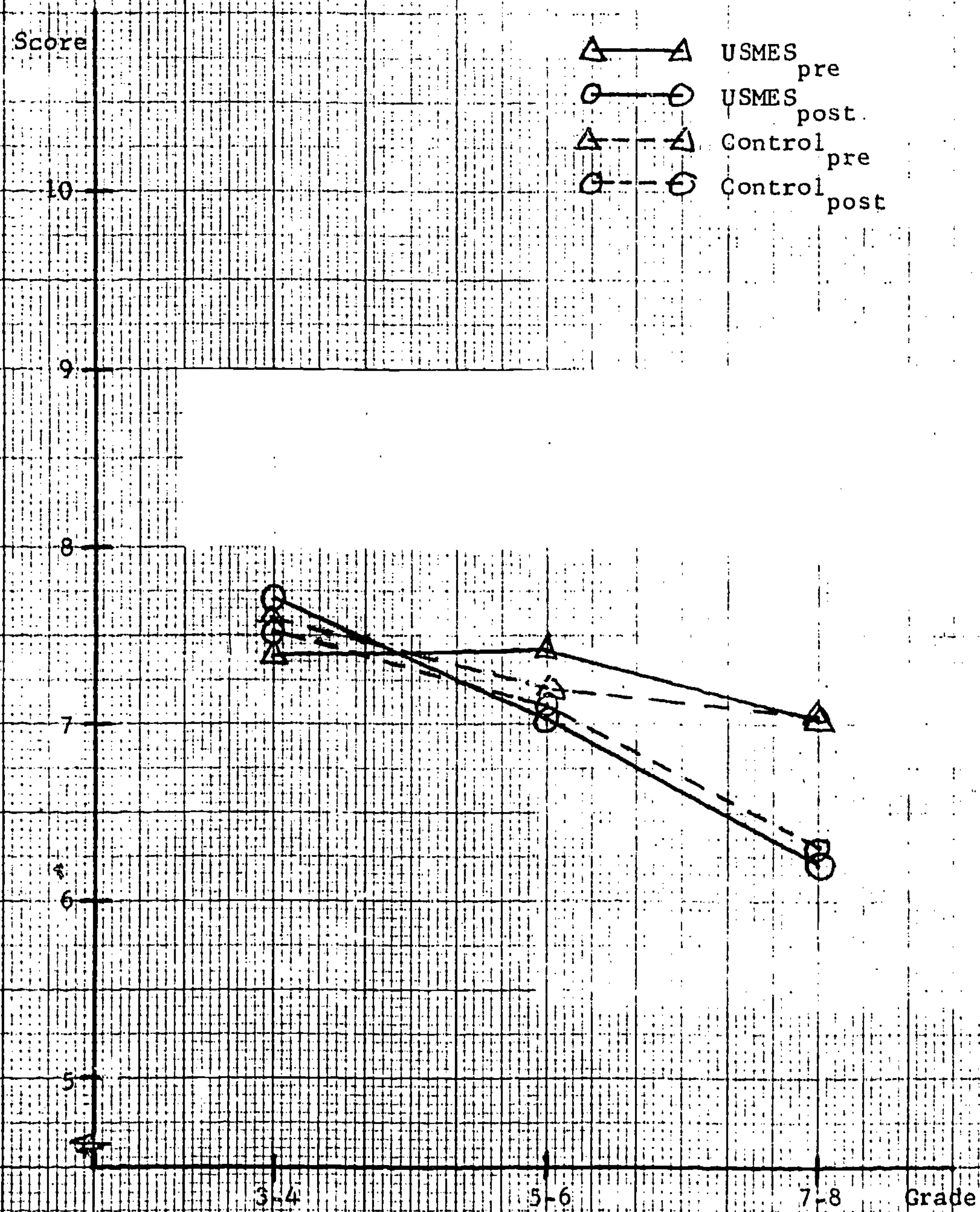

Figure 7.3-- Means on Part I, Factor 3: Non-mastery; Scale range: 2 to 10; Scale midpoint: 6.

encourage shorter, more intensive use of USMES challenges rather than encouraging that one unit should be pursued over an entire school year.

D. Factor 4: Preference for Group Learning Activities

Table 7.12 contains the results of the repeated measures analysis of variance for scores on factor 4, while Table 7.13 contains the analysis of covariance results for this factor. In neither analysis were treatment or grade level differences significant. However, the F-ratio for pre-to-post differences was significant at $p < .01$.

Looking at the Factor 4 means in Table 7.14 and the graph of these means in Figure 7.4, one can see that all groups, except USMES students at the fifth- and sixth-grade level decreased in their average preference for group learning activities. Nevertheless, all means were at least three points above the supposed point of neutrality on this attitude factor; the students generally liked to learn by working with other students.

E. Factor 5: Arithmetic Enjoyment

In Table 7.15, the results of the repeated measures analysis of variance are presented, while Table 7.16 contains the analysis of covariance results for scores on factor 5. The repeated measures analysis revealed a significant grade difference, but this grade effect did not reach statistical significance in the covariance analysis when post-test scores were adjusted for pre-test differences. The ANCOVA grade effect was significant only at $p < .11$.

It was the pre-to-post test differences in factor 5 scores which were highly significant at $p < .0001$. Interpretation of this result is aided by examination of the pre- and post-test means shown in Table 7.17 and graphed in Figure 7.5. Both treatment groups at all grade levels showed an

TABLE 7.12

Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 4: Preference for Group Learning Activities

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	136.92	2.45	
Grade(G)	2	2.55	1.28	0.522
Treatment(T)	1	1.21	1.21	0.494
T x G	2	8.46	4.23	1.731
Error	51	124.70	2.45	
Within Ss	57	68.88	1.21	
Tests(A)	1	8.34	8.34	7.636**
G x A	2	1.81	0.90	0.828
T x A	1	2.40	2.40	2.198
T x G x A	2	0.60	0.30	0.274
Error	51	55.73	1.09	
Total	113	205.80	1.82	

**p < .01

TABLE 7.13

Analysis of Covariance for Attitude Scores on Part I, Factor 4: Preference for Group Learning Activities

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	1.53	1.53	0.800
Grade(G)	2	3.08	1.54	0.809
T x G	2	1.28	0.64	0.334
Error	50	95.50	1.91	
Total	55	101.39	1.84	

TABLE 7.14

Means and Standard Deviations for Attitude Scores on Part I, Factor 4: Preference for Group Learning Activities

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	22.35 (1.62)	22.12 (1.13)	22.87 (1.06)	22.35
	Post	21.95 (1.03)	22.16 (1.88)	22.00 (1.21)	22.07
	Adjusted	22.09	22.43	21.88	22.13
Control	N	7	13	6	26
	Pre	23.30 (0.91)	22.87 (0.89)	22.26 (1.34)	22.85
	Post	22.60 (1.43)	22.04 (1.36)	21.16 (1.37)	21.99
	Adjusted	22.26	21.92	21.34	21.84
Total for Grades	N	15	29	13	57
	Pre	22.80	22.46	22.56	
	Post	22.25	22.11	21.61	
	Adjusted	22.17	22.17	21.61	

Note--Possible range of factor 4 scale score means is 6 to 30; the midpoint of the scale is 18.

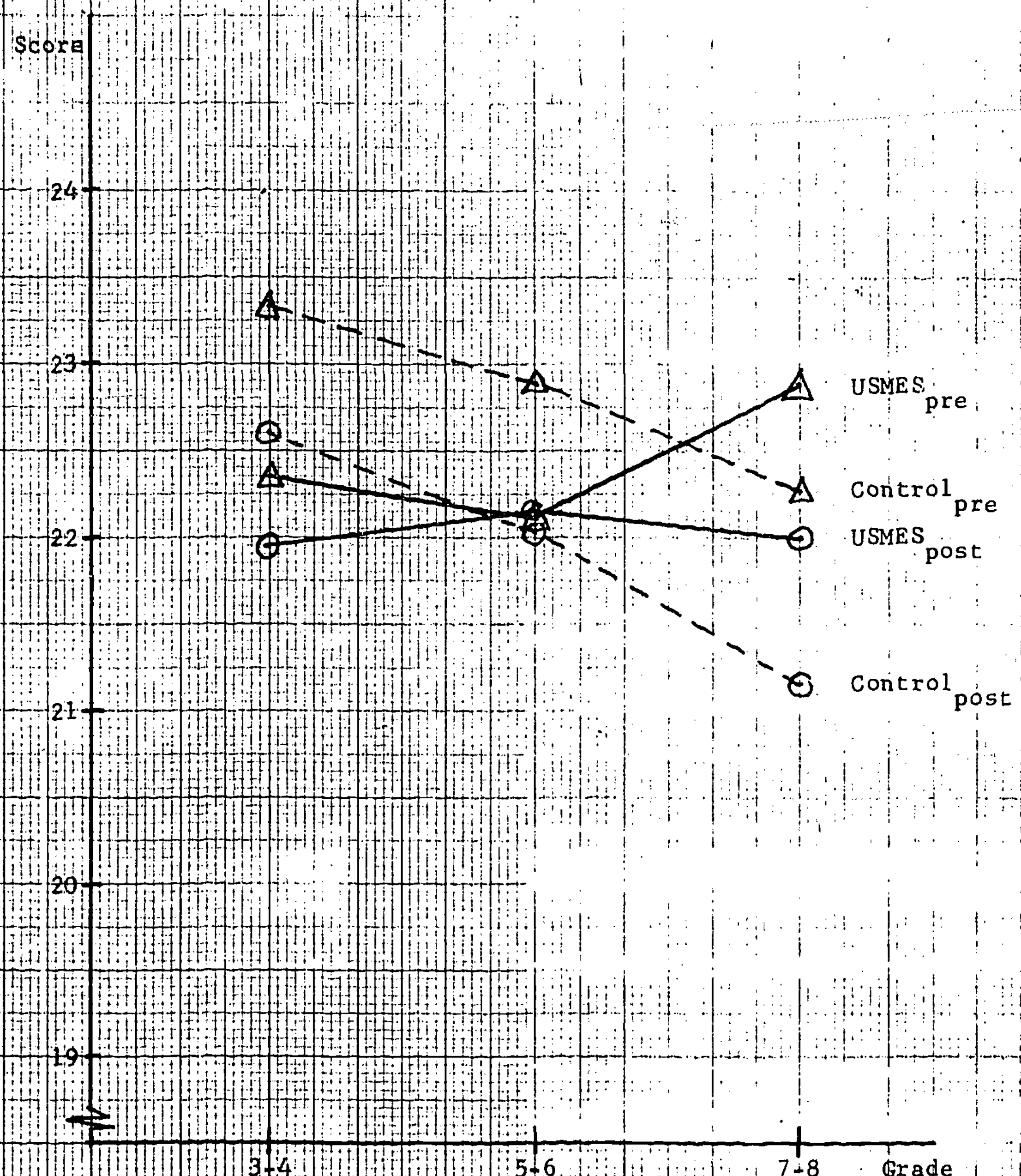

Figure 7.4--Means on Part I, Factor 4: Preference for Group Learning; Scale range: 6 to 30; Scale midpoint: 18.

TABLE 7.15

Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 5: Arithmetic Enjoyment

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	65.89	1.18	
Grade(G)	2	7.25	3.63	3.379*
Treatment(T)	1	0.63	0.63	0.586
T x G	2	3.30	1.65	1.537
Error	51	54.71	1.07	
Within Ss	57	63.57	1.12	
Tests(A)	1	30.98	30.98	49.967**
G x A	2	0.86	0.43	0.690
T x A	1	0.07	0.07	0.120
T x G x A	2	0.05	0.02	0.038
Error	51	31.62	0.62	
Total	113	129.46	1.15	

*$p < .05$
**$p < .01$

TABLE 7.16

Analysis of Covariance for Attitude Scores on Part I, Factor 5: Arithmetic Enjoyment

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	0.38	0.38	0.409
Grade(G)	2	4.36	2.18	2.238
T x G	2	0.74	0.37	0.400
Error	50	47.00	0.94	
Total	55	49.93	0.91	

TABLE 7.17

Means and Standard Deviations for Attitude Scores on Part I, Factor 5: Arithmetic Enjoyment

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	10.21 (1.05)	10.54 (0.56)	10.38 (0.48)	10.42
	Post	11.35 (0.93)	11.76 (0.86)	11.09 (1.10)	11.50
	Adjusted	11.38	11.69	11.06	11.38
Control	N	7	13	6	26
	Pre	10.55 (0.68)	10.46 (1.14)	9.72 (0.86)	10.31
	Post	11.68 (1.14)	11.49 (1.02)	10.47 (1.09)	11.30
	Adjusted	11.60	11.44	10.66	11.23
Total for Grades	N	15	29	13	57
	Pre	10.37	10.50	10.08	
	Post	11.50	11.64	10.80	
	Adjusted	11.49	11.56	10.86	

Note--Possible range of factor 5 scale score means is 3 to 15; the midpoint of the scale is 9.

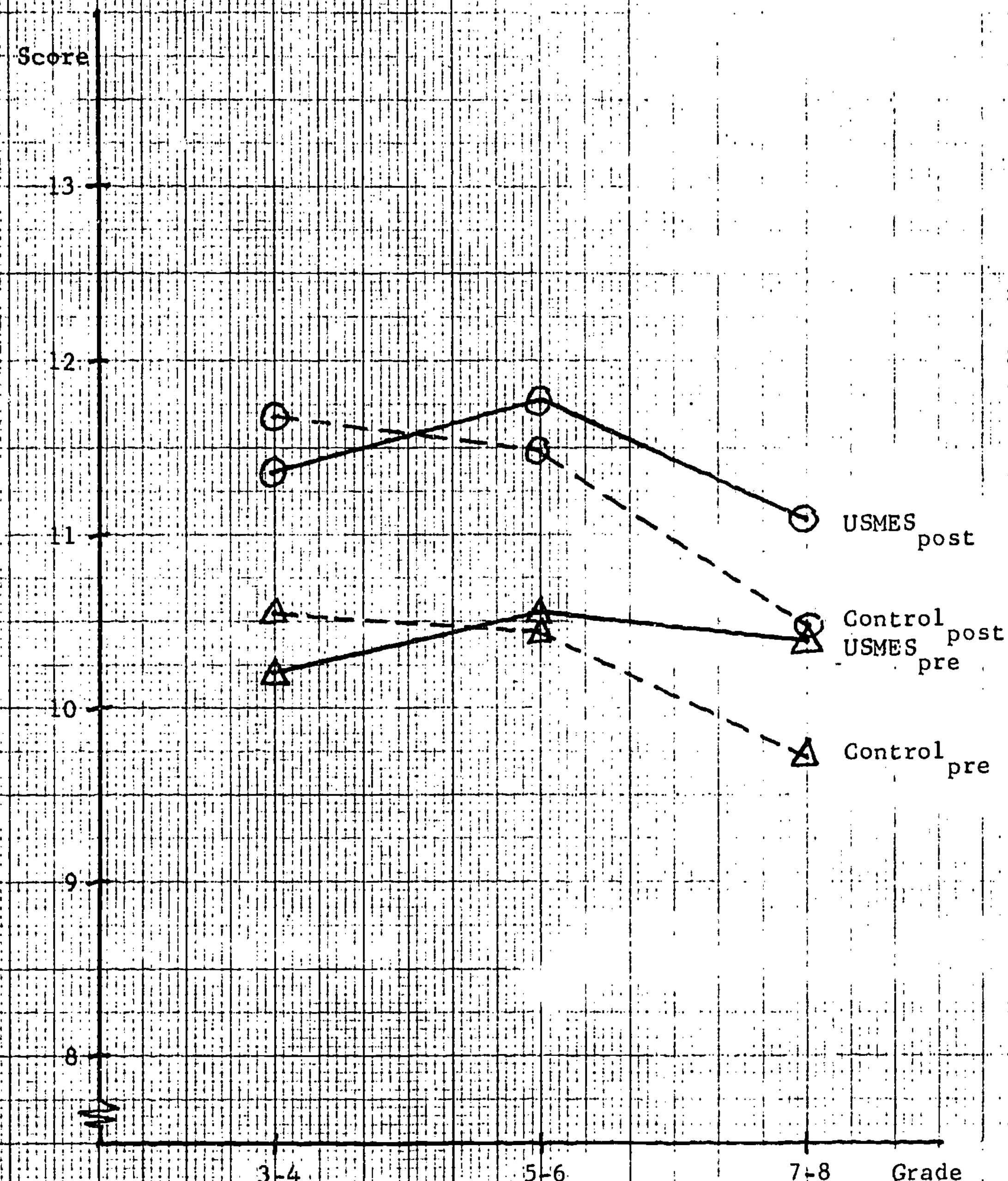

Figure 7.5-- Means on Part I, Factor 5: Arithmetic Enjoyment; Scale range: 3 to 15; Scale midpoint: 9.

increase in arithmetic enjoyment from Fall to Spring. The increase was of approximately the same magnitude for all groups of students.

F. Factor 6: Self-Directed Learning

The results of the repeated measures anlysis of variance for scores on factor 6 are presented in Table 7.18. The F-ratio for grade differences was significant at $p < .0001$, and there was even less doubt that pre-to-post differences could not be attributed to chance. However, there was also a significant interaction between these two effects: grade and time of test administration.

As shown in Table 7.19, highly significant grade differences ($p < .0001$) in factor 6 scores also emerged from the covariance analysis, when post-test scores were adjusted for pre-test differences. However, for factor 6, this adjustment was minimal. Table 7.20 and Figure 7.6 show that the pre-test means for both groups at all grade levels were very similar. On the other hand, grade level differences in post-test scores were striking.

At the beginning of the school year, the tendency was for students in both treatment groups at all grade levels to express a uniformly high preference for self-directed learning: deciding each for himself what to study; what to learn; whom to work with in class. However, in the Spring a vastly different picture emerged. At post-test time, the younger students in grades B and C found substantially less appeal in self-directed learning. Scores for fifth- and sixth-graders in both treatment groups also decreased on factor 6, but the decreases were not as marked as those for third- and forth-graders. The oldest students exhibited the smallest decrease in preference for self-directed learning over the period from Fall to Spring.

TABLE 7.18

Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 6: Self-Directed Learning

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	140.54	2.51	
Grade(G)	2	50.45	25.22	14.480**
Treatment(T)	1	1.18	1.18	0.679
T x G	2	0.07	0.03	0.0.9
Error	51	88.84	1.74	
Within Ss	57	325.37	5.71	
Tests(A)	1	229.44	229.44	230.732**
G x A	2	43.32	21.66	21.780**
T x A	1	1.73	1.73	1.744
T x G x A	2	0.16	0.08	0.083
Error	51	50.71	0.99	
Total	113	465.91	4.12	

**$p < .01$

TABLE 7.19

Analysis of Covariance for Attitude Scores on Part I, Factor 6: Self-Directed Learning

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	3.11	3.11	1.921
Grade(G)	2	90.34	45.17	27.910**
T x G	2	0.14	0.07	0.045
Error	50	31.00	1.62	
Total	55	124.59	2.27	

**$p < .01$

TABLE 7.20

Means and Standard Deviations for Attitude
Scores on Part I, Factor 6:
Self-Directed Learning

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	18.56 (1.38)	18.70 (0.81)	18.54 (0.62)	18.62
	Post	14.30 (0.69)	16.02 (1.43)	17.96 (1.27)	16.01
	Adjusted	14.32	15.99	17.99	16.10
Control	N	7	13	6	26
	Pre	18.42 (1.47)	18.76 (0.92)	18.75 (0.79)	18.67
	Post	13.94 (1.56)	15.50 (1.42)	17.59 (1.06)	15.56
	Adjusted	14.02	15.44	17.54	15.67
Total for Grades	N	15	29	13	57
	Pre	18.49	18.73	18.64	
	Post	14.13	15.79	17.79	
	Adjusted	14.17	15.72	17.77	

Note--Possible range of factor 6 scale score means is 5 to 25; the midpoint of the scale is 15.

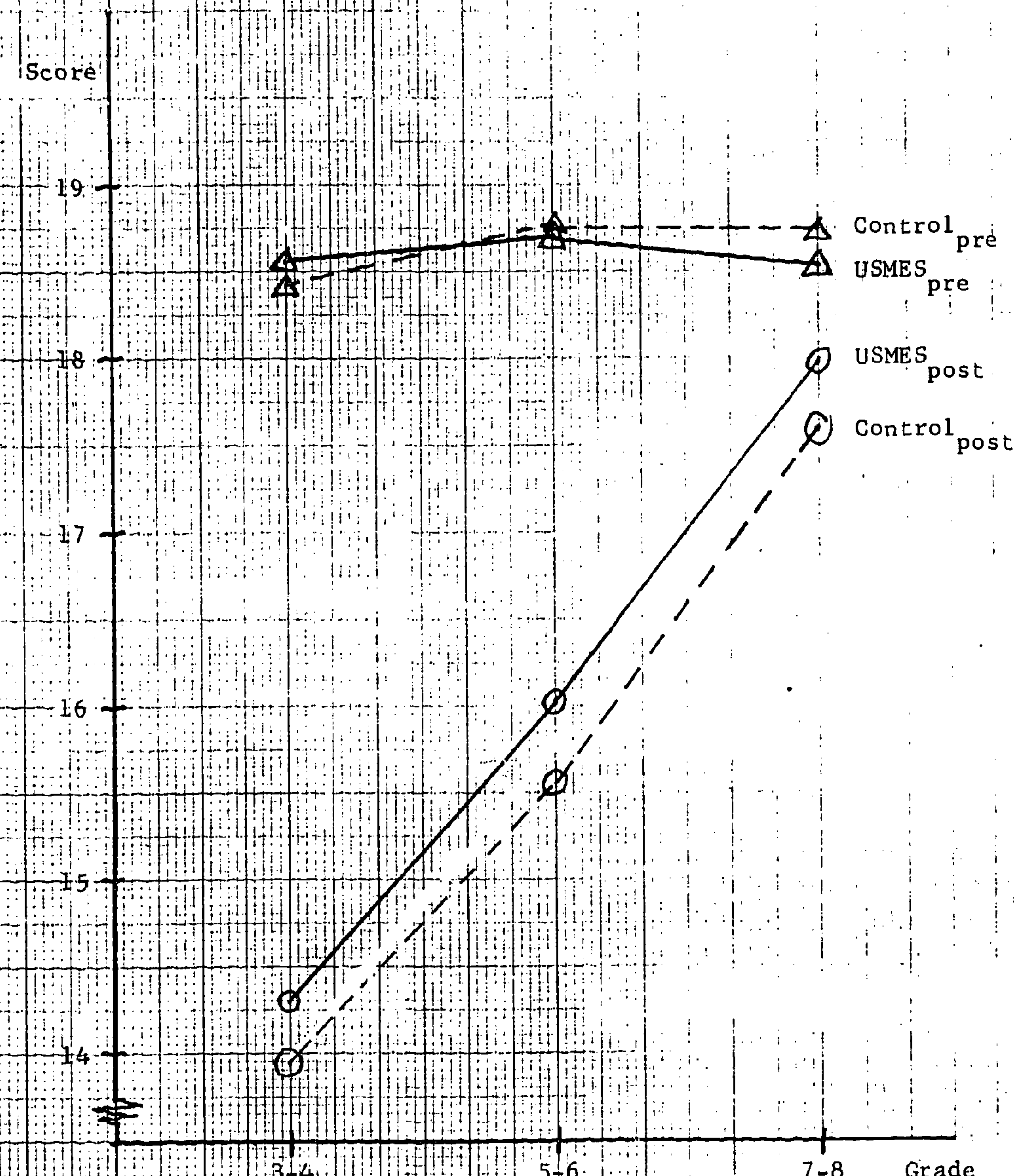

Figure 7.6--Means on Part I, Factor 6: Self Directed Learning; Scale range: 5 to 25; Scale midpoint: 15.

One would expect that older students would be, in fact, more self-directed, more independent. One would also expect that the more mature students should be more consistent and more accurate in appraising their own preferences.

G. Factor 7: Arithmetic Value

Table 7.21 contains the results of the repeated measures analysis of variance for scores on factor 7, and Table 7.22 contains the results of covariance analysis for these scores. Neither the grade nor the treatment effect was significant. However, as shown in Table 7.21, the difference between pre- and post-test administration was highly significant at $p < .0001$.

The nature of the pre-to-post differences are revealed in Table 7.23 and in the graph of these means, Figure 7.7. Both treatment groups at all three grade levels showed highly significant increases in arithmetic value scores between the Fall and Spring testing periods. The position of the post-test means toward the upper limit of the scale is especially noteworthy.

These positive attitudes toward arithmetic, from factor 5 and factor 7, were corroborated by the results of our interviews with USMES children. They do like math, very much; they find it useful and enjoyable.

Results from Part II of the Attitude Scale

Table 7.24 contains the item means and standard deviations for the statements on Part II of the attitude scale. These items were factor analyzed separately from Part I items, because all of the Part II items were adapted from a scale by Covington and all of the items were designed to measure attitudes toward complex problem solving, in the context of a specific problem facing a group of engineers.

TABLE 7.21

Repeated Measures Analysis of Variance for Attitude Scores on Part I, Factor 7: Arithmetic Value

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	28.00	0.50	
Grade (G)	2	2.77	1.38	3.878
Treatment (T)	1	0.16	0.16	0.325
T x G	2	0.57	0.29	0.598
Error	51	24.50	0.48	
Within Ss	57	317.13	5.56	
Tests (A)	1	295.05	295.05	779.153**
G x A	2	0.73	0.36	0.959
T x A	1	0.17	0.17	0.454
T x G x A	2	1.88	0.93	2.476
Error	51	19.31	0.37	
Total	113	345.13	3.05	

**p < .01

TABLE 7.22

Analysis of Covariance for Attitude Scores on Part I, Factor 7: Arithmetic Value

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment (T)	1	0.00	0.00(38)	0.009
Grade (G)	2	2.64	1.32	3.032
T x G	2	1.52	0.76	1.747
Error	50	22.00	0.44	
Total	55	26.16	0.48	

TABLE 7.23

Means and Standard Deviations for Attitude Scores on Part I, Factor 7: Arithmetic Value

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	9.82	10.27	9.82	10.05
	Post	12.88	13.44	13.65	13.34
	Adjusted	12.91	13.41	13.68	13.34
Control	N	7	13	6	26
	Pre	10.24	10.16	10.27	10.20
	Post	13.08	13.60	13.08	13.34
	Adjusted	13.06	13.60	13.06	13.24
Total for Grades	N	15	29	13	57
	Pre	10.02	10.22	10.03	
	Post	12.97	13.51	13.39	
	Adjusted	12.38	13.51	13.37	

Note--Possible range of factor 7 scale score means is 3 to 15; the midpoint of the scale is 9.

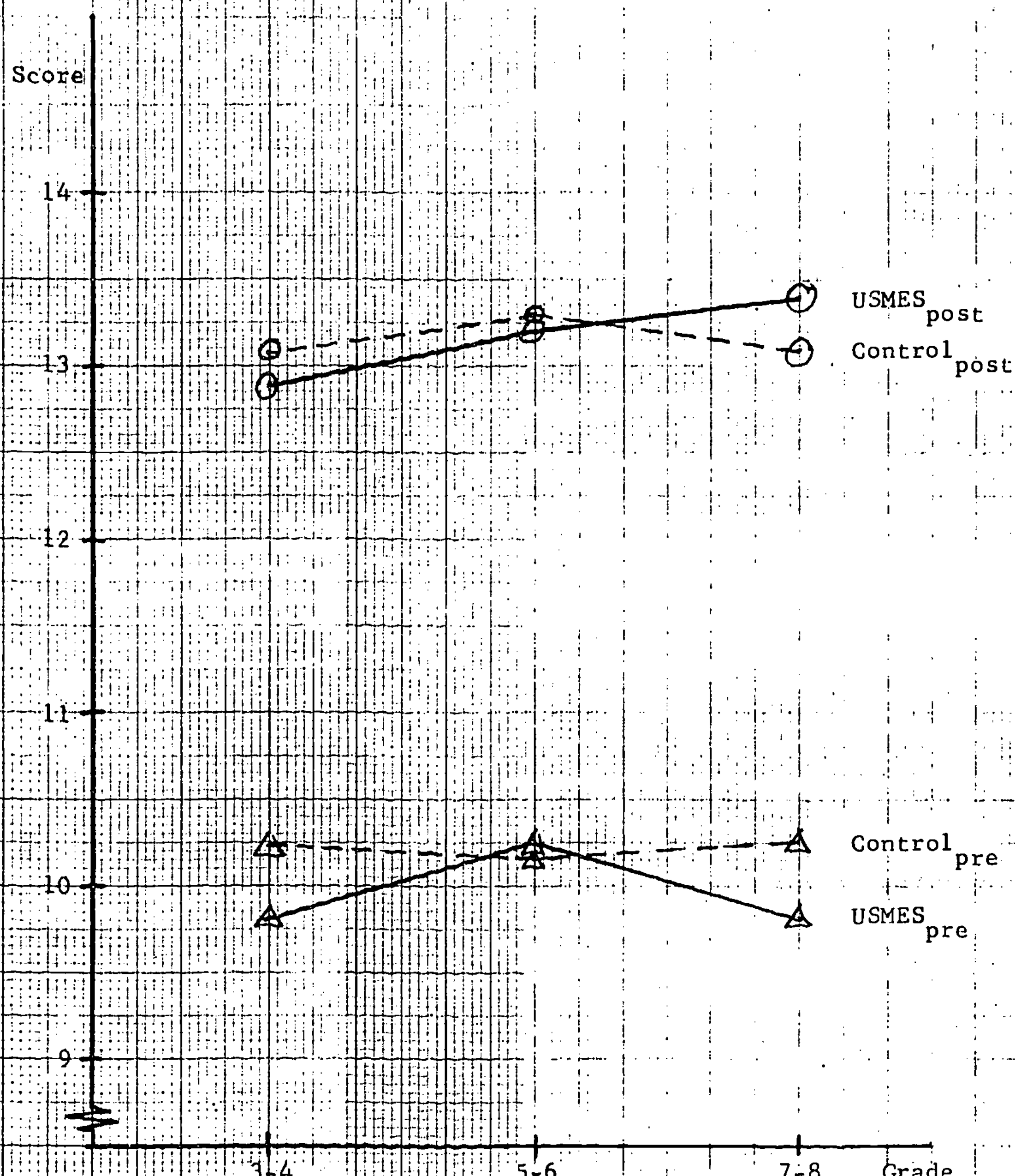

Figure 7.7--Means on Part I, Factor 7: Arithmetic Value; Scale range: 3 to 15; Scale midpoint: 9.

TABLE 7.24

Pre-test and Post-test Means and Standard Deviations for Items on Part II of the Attitude Scale Given by Factor for USMES and Control Groups

Item No.[a]	Item	USMES		Control	
		Pre	Post	Pre	Post
	Factor 1: Need to Please--Acquiesce				
3.	In a problem like this one, the best answer will be the one that most of the class decides is right.	3.47 (1.29)[b]	3.35 (1.34)	3.46 (1.31)	3.41 (1.34)
5.	The best answer is the one that the teacher thinks is right.	2.98 (1.47)	2.35 (1.41)	3.14 (1.47)	2.45 (1.43)
6.	It is best to make sure that an idea is a good one before sharing it with the class.	4.09 (1.18)	3.53 (1.31)	4.22 (1.12)	3.79 (1.25)
10.	I don't think I should ask too many questions about problems in class.	2.82 (1.46)	2.73 (1.37)	2.83 (1.45)	2.85 (1.40)
11.	Other students know more about problems like this than I do.	3.04 (1.26)	2.93 (1.20)	2.90 (1.33)	3.01 (1.29)
12.	If I already have one good idea, I would rather stick with it than look for more ideas.	3.07 (1.53)	2.70 (1.40)	2.91 (1.55)	2.81 (1.40)
	Factor 2: Divergent Thinking				
4.	An idea for solving a problem could lead to a wrong answer but still be a good idea.	3.79 (1.07)	3.90 (1.02)	3.78 (1.13)	3.91 (0.96)
8.	I would like to work on a problem like this one.	3.52 (1.30)	3.28 (1.34)	3.45 (1.32)	3.18 (1.38)
9.	If I worked on this problem, I would get a lot of good ideas.	3.68 (1.31)	3.44 (1.10)	3.84 (1.13)	3.32 (1.13)
13.	I would like to work on a problem like this, even though I might not be able to solve it.	3.52 (1.30)	2.93 (1.20)	3.51 (1.32)	3.01 (1.29)
15.	I am very good at thinking and solving problems.	3.31 (1.16)	3.19 (1.17)	3.37 (1.21)	3.19 (1.18)

TABLE 7.24 (Cont.)

Item No.[a]	Item	USMES Pre	USMES Post	Control Pre	Control Post
	Factor 3: Academic Competition				
2.	If someone gets an idea that no one else has thought of, he should keep it to himself.	2.25 (1.53)	1.92 (1.36)	2.16 (1.49)	1.98 (1.35)
14.	I think that my ideas for solving this problem would be better than ideas given by other students.	2.62 (1.29)	2.50 (1.19)	2.52 (1.26)	2.57 (1.28)
	Factor 4: Convergent Thinking				
1.	There is probably only one answer to a problem like this one.	3.10 (1.28)	2.87 (1.28)	3.16 (1.30)	2.81 (1.26)
7.	There is probably only one best way to solve a problem like this one.	3.81 (1.29)	3.08 (1.26)	3.80 (1.29)	3.10 (1.28)

[a]Item numbers were taken from the post-test form.

[b]Standard deviations are given in parentheses.

Factor analysis of the data from Part II yielded four significant factors. As was the case with Part I factors, the discussion of pre-to-post, treatment, and grade differences on each Part II factor is accompanied by tables for the repeated measures and covariance analyses. Tables of means by group and by grade level, and graphs of these data are also presented.

These results for Part II are given below by factor.

A. Factor 1: Need to Please--Acquiesce

Differences between pre-test and post-test scores on this first factor from Part II of the attitude scale were highly significant, as shown by the results of the repeated measures analysis of variance in Table 7.25. The F-ratio for grade effects in this analysis was also significant at $p < .0001$. However, the results in Table 7.26 show that when post-test means were statistically adjusted to account for pre-test differences, the grade effect was no longer statistically significant.

On the other hand, the covariance analysis produced an F-ratio for the treatment difference which approaches statistical significance at $p < .14$. The outcome is not statistically significant, but it merits further investigation, especially since relatively stable class means were used as the unit of analysis, and the degrees of freedom are much lower that they would be in the less conservative use of individual's scores as the unit of analysis.

One should look at the means in Table 7.27 to examine the nature of this possible treatment group difference. The adjusted mean of 17.74 for USMES students suggests that they are somewhat less inclined to want to please, conform, or consent without protest, than students in the control classes,

TABLE 7.25

Repeated Measures Analysis of Variance for Attitude Scores on Part II, Factor 1: Need to Please--Acquiesce

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	576.99	10.30	
Grade (G)	2	233.39	116.70	18.435
Treatment (T)	1	9.74	9.74	1.539
T x G	2	11.04	5.52	0.872
Error	51	322.83	6.33	
Within Ss	57	120.34	2.11	
Tests (A)	1	63.42	63.42	60.980
G x A	2	2.29	1.14	1.101
T x A	1	1.24	1.24	1.944
T x G x A	2	0.35	0.17	0.167
Error	51	53.04	1.04	
Total	113	697.33	6.17	

TABLE 7.26

Analysis of Covariance for Attitude Scores on Part II, Factor 1: Need to Please--Acquiesce

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment (T)	1	3.94	3.94	2.191
Grade (G)	2	1.30	0.65	0.392
T x G	2	.78	0.38	0.214
Error	50	90.00	1.80	
Total	55	96.02	1.75	

TABLE 7.27

Means and Standard Deviations for Attitude Scores on Part II, Factor 1: Need to Please--Acquiesce

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	21.50 (0.95)	19.25 (2.60)	16.82 (2.17)	19.28
	Post	19.31 (1.43)	17.56 (2.49)	15.72 (2.10)	17.60
	Adjusted	17.89	17.73	17.61	17.74
Control	N	7	13	6	26
	Pre	22.20 (1.42)	18.98 (1.57)	18.17 (1.62)	19.66
	Post	20.71 (1.03)	17.72 (1.20)	17.16 (2.35)	18.40
	Adjusted	18.79	18.08	18.10	18.32
Total for Grades	N	15	29	13	57
	Pre	21.83	19.13	17.44	
	Post	19.96	17.63	16.39	
	Adjusted	18.34	17.90	17.85	

Note--Possible range of factor 1 scale score means is 6 to 30; the midpoint of the scale is 18.

whose adjusted mean is 18.32

The graph of means on factor 1 in Figure 8 presents a vivid picture of the substantial pre-to-post and grade differences. All groups decrease in their need to please to conform, to acquiesce, over the period of a school year. The sizes of the decreases are similar for all grade levels, but the higher grade students tend less and less to need to please--to acquiesce. It is more typical for the younger students to want to satisfy the teacher's expectations and to conform to the viewpoints of classmates. USMES students at the seventh- and eighth-grade levels expressed the least need to please or to consent without protest.

B. Factor 2: Divergent Thinking

Results from the repeated measures analysis of variance for scores on attitudes toward divergent thinking are presented in Table 7.28. The grade differences were significant at $p < .001$, and the pre-to-post differences were significant at $p < .01$.

Even with statistical adjustment of post-test means for pre-test differences, the grade differences remained significant at $p < .02$. This result from the analysis of covariance of factor 2 scores is shown in Table 7.29.

The means on attitudes toward divergent thinking are presented by treatment and by grade in Table 7.30. These pre- and post-test means are portrayed graphically in Figure 9. In general, the means decrease from pre-test to post-test. It seems that overall, the younger students expressed more confidence in their divergent thinking abilities and greater preference for divergent production activities. The direction of this grade difference seems to contradict the results, from the analysis of Part II, Factor 1,

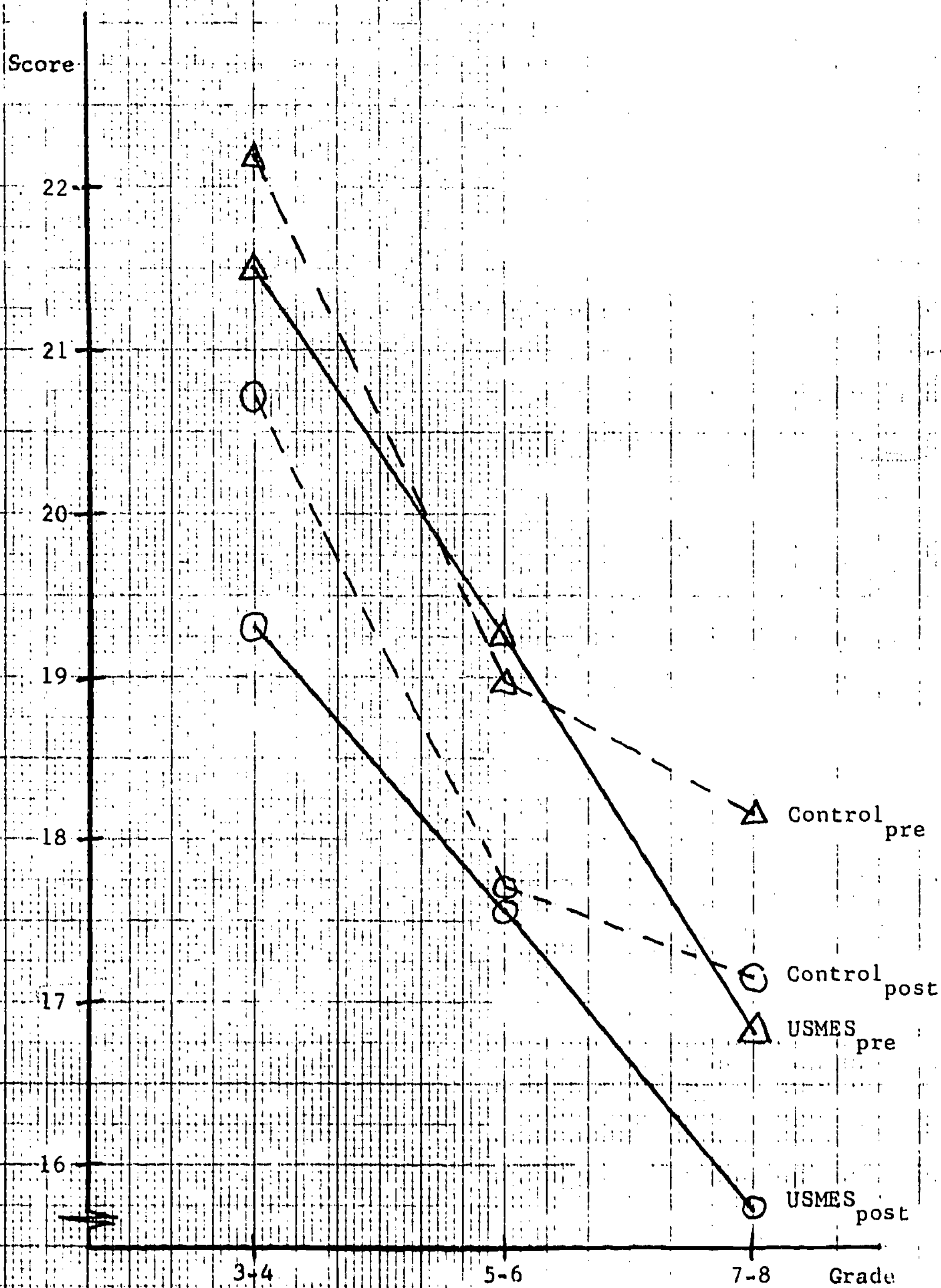

Figure 7.8.- Means on Part II, Factor 1: Need to Please--Acquiesce; Scale range: 6 to 30; Scale mid-point: 18.

TABLE 7.28

Repeated Measures Analysis of Variance for Attitude Scores on Part II, Factor 2: Divergent Thinking

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	75.78	1.35	
Grade(G)	2	19.45	9.72	9.270
Treatment(T)	1	0.01	0.01	0.007
T x G	2	2.82	1.41	1.344
Error	51	53.50	1.05	
Within Ss	57	43.97	0.77	
Tests(A)	1	5.44	5.44	8.529
G x A	2	1.24	0.62	0.971
T x A	1	1.22	1.22	1.912
T x G x A	2	3.56	1.78	2.794
Error	51	32.52	0.64	
Total	113	119.75	1.06	

TABLE 7.29

Analysis of Covariance for Attitude Scores on Part II, Factor 2: Divergent Thinking

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	0.77	0.77	1.037
Grade(G)	2	6.12	3.06	4.118*
T x G	2	0.60	0.30	0.398
Error	50	37.00	0.74	
Total	55	44.49	0.81	

*$p < .05$

TABLE 7.30

Means and Standard Deviations for Attitude Scores on Part II, Factor 2: Divergent Thinking

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	17.67 (0.61)	18.21 (1.02)	17.08 (1.28)	17.82
	Post	18.19 (0.54)	17.49 (1.07)	17.04 (1.02)	17.57
	Adjusted	18.23	17.41	17.21	17.62
Control	N	7	13	6	26
	Pre	18.96 (1.20)	17.92 (0.74)	17.24 (0.81)	18.05
	Post	17.98 (0.74)	17.39 (0.90)	16.67 (0.49)	17.38
	Adjusted	17.73	17.38	16.80	17.30
Total for Grades	N	15	29	13	57
	Pre	18.27	18.08	17.16	
	Post	18.09	17.45	16.87	
	Adjusted	17.98	17.39	17.01	

Note--Possible range of factor 2 scale score means is 5 to 25; the midpoint of the scale is 15.

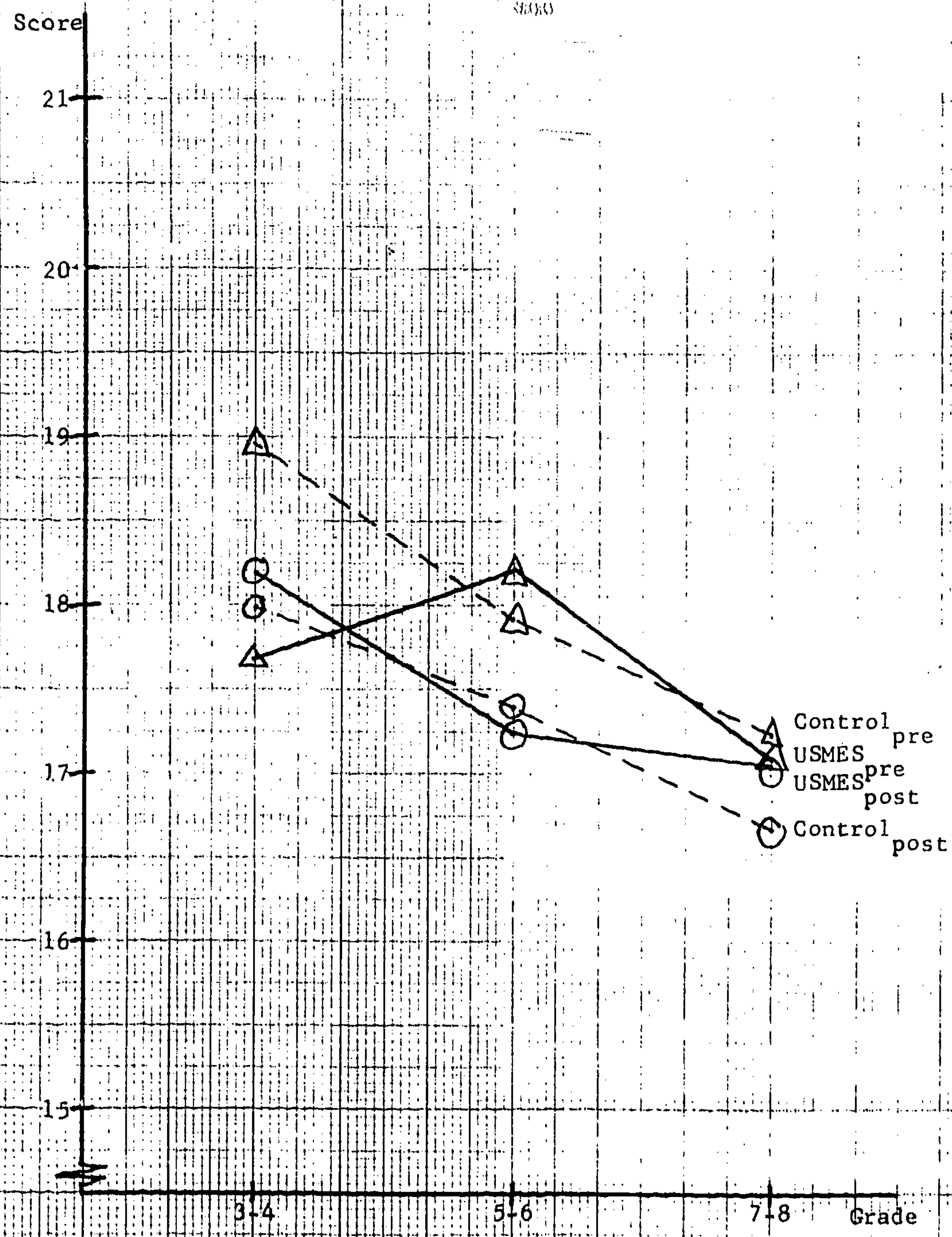

Figure 7.9.--Means on Part II, Factor 2: Divergent Thinking; Scale range: 5 to 25; Scale midpoint: 15.

which showed that the students in the higher grades tended to score lower on the need to please--acquiesce, or the need to conform. However, statements on this factor dealt, for the most part, with how one would respond in a group. The factor 2 items focused on one's individual efforts.

C. Factor 3: Academic Competition

Table 7.31 contains the results of the repeated measures analysis of variance on factor 3 scores from Part II. Grade level differences were found to be significant at $p < .01$, and the interaction effect between grade and test administration was significant at $p < .05$.

Analysis of covariance results for factor 3 are reported in Table 7.32. When the statistical adjustments were made for pre-test differences, only the treatment effect approached significance, at $p < .14$. An F-ratio with this p value is not statistically significant, but the effect merits further investigation.

The means for factor 3 scores are presented in Table 7.33. They are displayed graphically in Figure 10. The overall grade difference revealed in the repeated measures analysis of variance is one involving a decrease in factor 3 scores from the lower to higher grade levels. The grade by test administration interaction effect reported in the repeated measures analysis table can also be seen in the graph. At the third- and forth-grade level, both USMES and control groups scored higher on the pre-test; at the fifth- and sixth-grade level, both groups scored at approximately the same position on both pre-test and post-test; but an inversion occurred at the seventh- and eighth-grade level, where pre-test scores are lower for both USMES and control groups.

TABLE 7.31

Repeated Measures Analysis of Variance for Attitude Scores on Part II, Factor 3: Academic Competition

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	57.88	1.03	
Treatment(T)	1	0.01	0.01	0.01
Grade(G)	2	21.91	10.96	15.59**
T x G	2	0.12	0.07	0.08
Error	51	35.84	0.70	
Within Ss	57	16.44	0.29	
Tests(A)	1	0.48	0.48	1.90
T x A	1	0.77	0.77	3.06
G x A	2	1.68	0.84	3.33
T x G x A	2	0.69	0.34	1.37
Error	51	12.82	0.25	
Total	113	74.32	0.66	

*p < .05
**p < .01

TABLE 7.32

Analysis of Covariance for Attitude Scores on Part II, Factor 3: Academic Competition

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	0.61	0.61	2.177
Grade(G)	2	0.58	0.29	1.011
T x G	2	0.52	0.26	0.931
Error	50	14.00	0.28	
Total	55	15.71	0.29	

TABLE 7.33

Means and Standard Deviations for Attitude Scores on Part II, Factor 3: Academic Competition

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	5.69 (0.79)	4.61 (0.81)	4.00 (0.57)	4.75
	Post	4.85 (0.40)	4.38 (0.75)	4.13 (0.43)	4.45
	Adjusted	4.46	4.41	4.39	4.42
Control	N	7	13	6	26
	Pre	5.22 (0.61)	4.56 (0.81)	4.03 (0.49)	4.62
	Post	5.20 (0.56	4.43 (0.58)	4.30 (0.62)	4.60
	Adjusted	4.99	4.47	4.55	4.67
Total for Grades	N	15	29	13	57
	Pre	5.47	4.59	4.02	
	Post	5.01	4.40	4.21	
	Adjusted	4.73	4.44	4.47	

Note--Possible range of factor 3 scale score means is 2 to 10; the midpoint of the scale is 6.

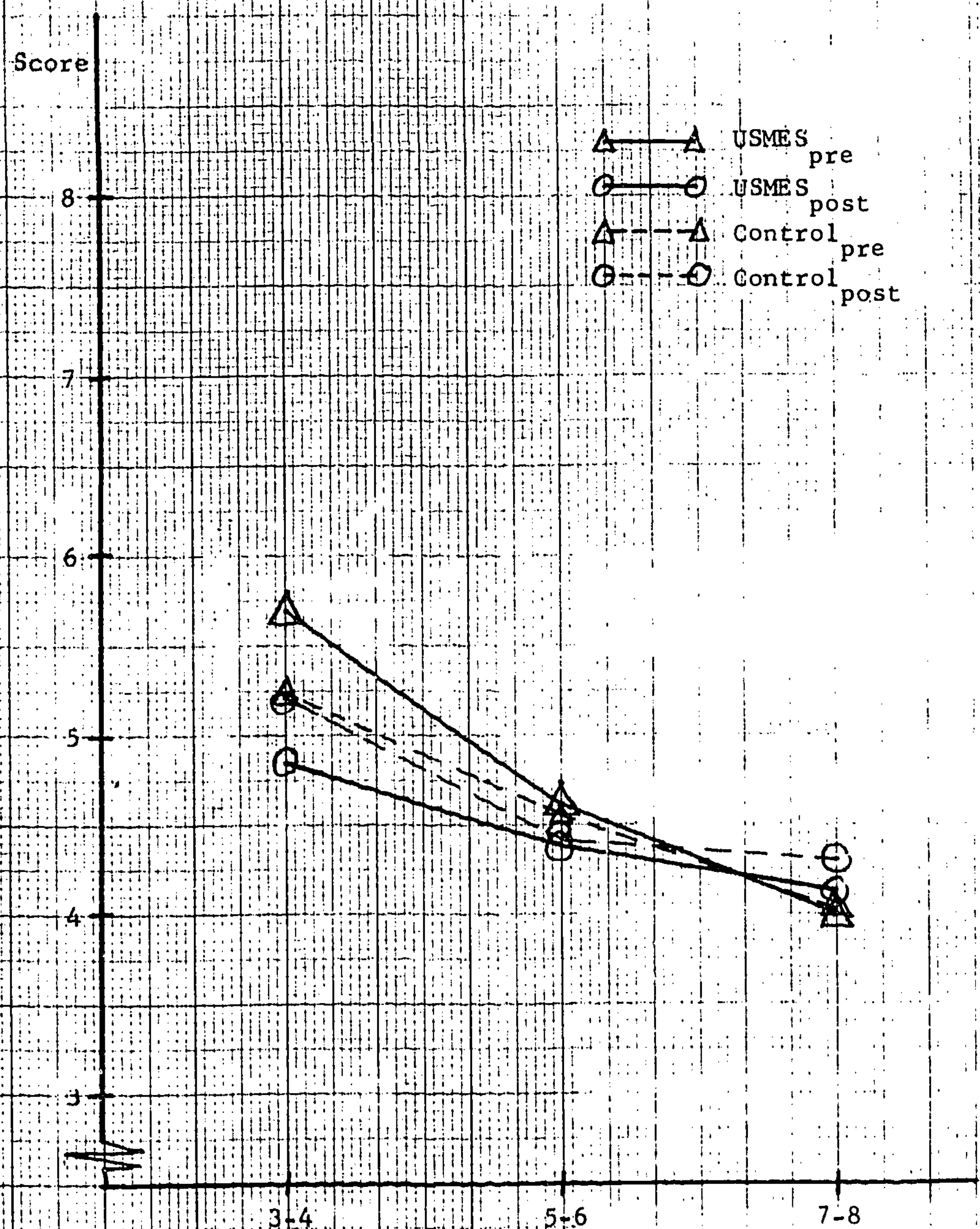

Figure 7.10--Means on Part II, Factor 3: Academic Competition; Scale range: 2 to 10; Scale midpoint: 6.

The meaning one can attach to these results for factor 3 is questionable, however, because the content of the factor itself is not clear. The combination of items loading on factor 3 is not logically satisfying; the assigned name for the factor was not arrived at easily; and it is not consistent with our experience that younger students should express more academic competition than older students.

D. Factor 4: Convergent Thinking

Results of the repeated measures analysis of variance for scores on Factor 4 are shown in Table 7.34. The pre-to-post differences were highly significant at $p < .0001$, as were differences attributable to grade level.

The significant grade level effect persisted even when post-test scores were adjusted for pre-test differences. This result is shown in Table 7.35 which contains the analysis of covariance results for scores on factor 4.

The factor 4 means presented in Table 7.36 are portrayed graphically in Figure 11. These data reveal that students at higher grade levels are less inclined to believe that only one solution is appropriate for a complex problem. Further, the graph shows a marked decrease in this belief over the school year, from pre-test to post-test administration.

Summary and Discussion

An attitude scale was developed and pilot tested by the evaluation team especially for this USMES evaluation. The scale consisted of two parts. Part I contained items designed to measure attitudes toward math and science and toward various teaching strategies and learning activities which are embodied by the USMES approach. The statement in Part II of the scale were selected and adapted from items developed by Covington (1967) and presented in the context of a real-life, complex problem facing a group of engineers.

TABLE 7.34

Repeated Measures Analysis of Variance for Attitude Scores on Part II, Factor 4: Convergent Thinking

Source	df	Sum of Square	Mean Square	F-Ratio
Between Ss	56	35.75	0.64	
Grade(G)	2	11.19	5.59	11.705
Treatment(T)	1	0.00	0.00	0.001
T x G	2	0.19	0.09	0.198
Error	51	24.37	0.48	
Within Ss	57	45.06	0.79	
Tests(A)	1	30.49	30.49	119.304
G x A	2	0.59	0.30	1.604
T x A	1	0.13	0.13	0.494
T x G x A	2	0.83	0.41	1.615
Error	51	13.03	0.26	
Total	113	80.81	0.72	

TABLE 7.35

Analysis of Covariance for Attitude Scores on Part II, Factor 4: Convergent Thinking

Source	df	Sum of Square	Mean Square	F-Ratio
Treatment(T)	1	0.12	0.12	0.315
Grade(G)	2	3.98	1.99	5.189**
T x G	2	0.98	0.49	1.274
Error	50	19.00	0.38	
Total	55	24.08	0.44	

**p $<$ .01

TABLE 7.36

Means and Standard Deviations for Attitude Scores on Part II, Factor 4: Convergent Thinking

Treatment Groups		Grade Levels 3-4	5-6	7-8	Total for Treatments
USMES	N	8	16	7	31
	Pre	7.30 (0.58)	6.81 (0.48)	6.95 (0.61)	6.96
	Post	6.67 (0.84)	5.88 (0.66)	5.47 (0.48)	5.99
	Adjusted	6.58	5.96	5.50	6.02
Control	N	7	13	6	26
	Pre	7.56 (0.60)	6.84 (0.48)	6.81 (0.79)	7.03
	Post	6.47 (0.58)	5.67 (0.61)	5.81 (0.60)	5.92
	Adjusted	6.29	5.75	5.89	5.98
Total for Grades	N	15	29	13	57
	Pre	7.42	6.82	6.88	
	Post	6.58	5.79	5.63	
	Adjusted	6.43	5.85	5.70	

Note--Possible range of factor 4 scale score means is 2 to 10; the midpoint of the scale is 6.

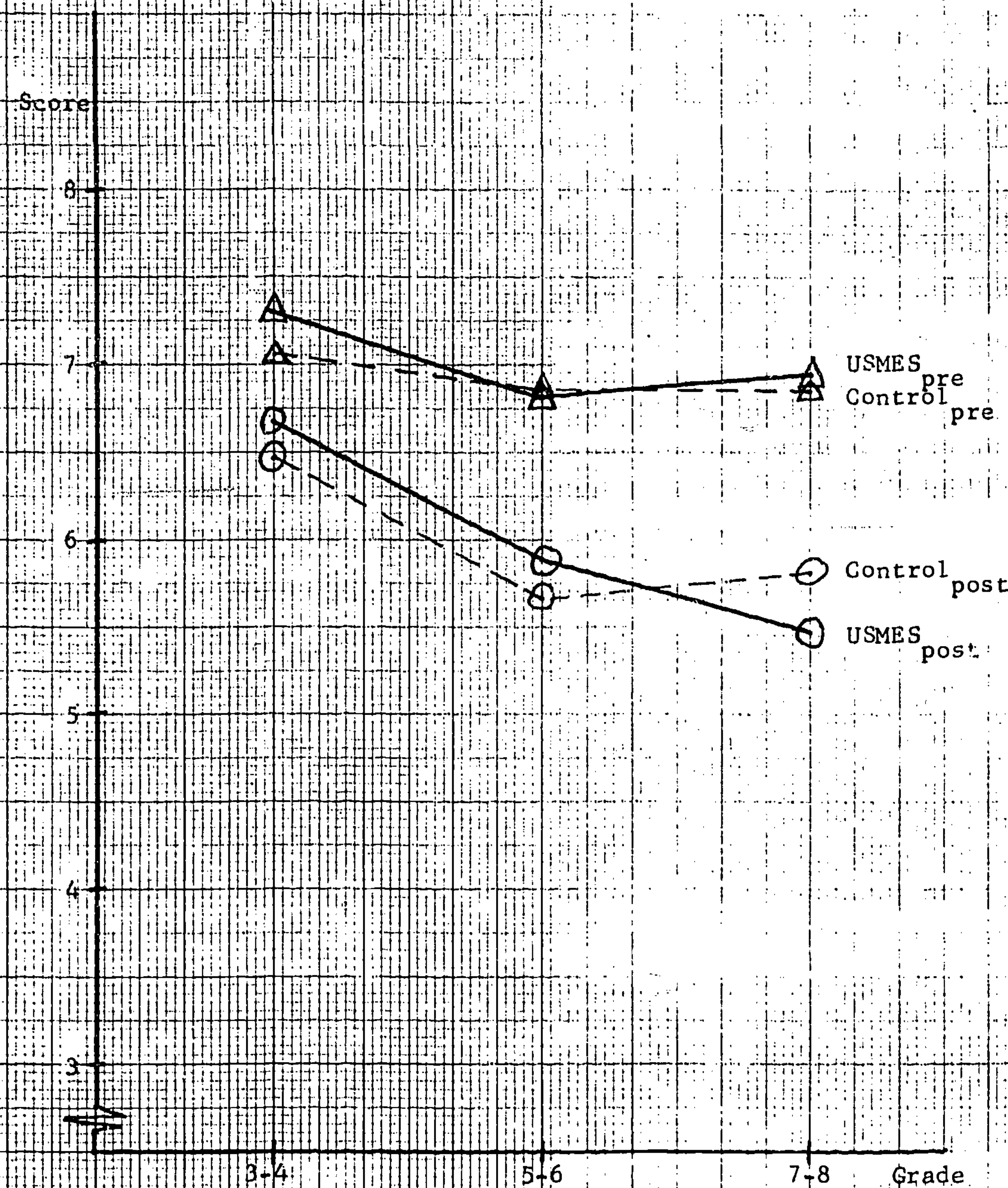

Figure 7.11-- Means on Part II, Factor 4: Convergent Thinking; Scale range: 2 to 11; Scale midpoint: 6.

Factor analyses of the largest data base, the post-test attitude results for 1491 students, yielded 7 factors on Part I and 4 factors on Part II. Hoyt estimates for internal consistency reliability of the 11 factors ranged from .35 to .77.

Scale scores were computed for each factor by adding the raw scores for items loading .3 or greater on the factor. Class means on scale scores were used as the unit of analysis in repeated measures analysis of variance and in analysis of covariance for each factor.

The repeated measures analyses revealed highly significant ($p < .0001$) pre-to-post test administration differences for most of the 11 factors. Only Only the academic insecurity scale from Part I and the academic competition scale from Part II produced no significant pre- post-test differences. These results are summarized in Table 7.37.

These pre-to-post differences may be indicative only of "time-of-year" effects, as discussed by Ahlgren and his associates with reference to their Minnesota School Affect Assessment:

> There is a class of items, mostly "pure school" types like PRINCIPAL, GETTING GOOD MARKS AND LISTENING TO THE TEACHER, which generally show a distinct drop in ratings between fall and spring. There is usually an almost complete "recovery" over the summer. This may be entirely a general human reaction and does not necessarily point to deleterious aspects of schooling.
>
> When evaluating programs, it is especially important to be aware of this kind of time-of-year effect. A pre-post evaluation of a new program for example, could show zero or negative results that were in fact an imporvement over the usual drop. For program evaluation these effects should be accounted for by the use of a control group, or be based on fall-fall or spring-spring differences rather than fall-spring differences. (Ahlgren, Christensen, & Lun, 1973, p. 27-28.)

TABLE 7.37

Summary of Significant[a] Results of Repeated Measures Analyses of Variance and of Analyses of Covariance on Eleven Attitude Factors

Factor Number	Factor Name	Repeated Measures Analysis				Covariance Analysis		
		Treatment	Grade	Pre-Post	Interactions	Treatment	Grade	Interactions
PART I								
1.	Science Appreciation	N.S.[b]	$p < .05$	$p < .0001$	T x G[c]	$p < .07$	$p < .015$	none
2.	Academic Insecurity	N.S.	N.S.	N.S.	none	N.S.	N.S.	none
3.	Non-mastery[d]	———	———	———	———	N.S.	$p < .0001$	none
4.	Group Learning	N.S.	N.S.	$p < .01$	none	N.S.	N.S.	none
5.	Arithmetic Enjoyment	N.S.	$p < .05$	$p < .0001$	none	N.S.	$p < .11$	none
6.	Self-Directed Learning	N.S.	$p < .0001$	$p < .0001$	G x A[e]	N.S.	$p < .0001$	none
7.	Arithmetic Value	N.S.	N.S.	$p < .0001$	none	N.S.	$p < .06$	none

[a]Only those F-ratios with p values $\leq .05$ have been regarded as statistically significant. However, effects with $.15 > p > .05$ have been included in this table because they merit further investigation.

[b]Not significant.

[c]Treatment by grade interaction.

[d]No repeated measures analysis of variance was computed for Part I, factor 3 scores.

[e]Interaction of grade with test administration.

TABLE 7.37 (Cont.)

Factor Number	Factor Name	Repeated Measures Analysis				Covariance Analysis		
		Treatment	Grade	Pre-Post	Interactions	Treatment	Grade	Interactions
	PART II							
1.	Need to Please Acquiesce	N.S.	$p < .0001$	$p < .0001$	none	$p < .145$	N.S.	none
2.	Divergent Thinking	N.S.	$p < .001$	$p < .01$	none	N.S.	$p < .05$	none
3.	Academic Competition	N.S.	$p < .01$	N.S.	G x A[e]	$p < .14$	N.S.	none
4.	Convergent Thinking	N.S.	$p < .0001$	$p < .0001$	none	N.S.	$p < .01$	none

[a] Only those F-ratios with p values $\leq .05$ have been regarded as statistically significant. However, effects with $.15 > p > .05$ have been included in this table because they merit further investigation.

[b] Not significant.

[c] Treatment by grade interaction.

[d] No repeated measures analysis of variance was computed for Part I, factor 3 scores.

[e] Interaction of grade with test administration.

This evaluation did employ a control group design but fall-fall testings were not feasible because samples would not have been accessible the second fall. There were no treatment by test administration interactions resulting from our analyses. Both treatment groups moved in the same direction. But in light of Ahlgren's and his associates' observations, our results are very interesting because on most of the scales, the groups' attitudes moved toward the more socially or academically desirable direction up or down, over the course of the school year.

The time-of-year effect observed by Ahlgren et al. is probably evidenced in the overall decline of students' scores on factor 4, Group Learning, and on factor 6, Self-Directed Learning, over the period from Fall to Spring. At the end of the school year, students, perhaps, are tired or less motivated,aand would prefer to play a more passive role in learning, while the teacher "runs the show."

There were no significant treatment differences revealed by repeated measures analyses, as the summary of results in Table 7.37 indicates. However, there were a number of significant grade differences in attitude factors which have implications for USMES development and USMES usage, even though the grade differences did not interact with treatment when covariance adjustments were made. Guidelines for curriculum development and implementation which come from research in developmental psychology are reinforced by some of the grade differences observed in this affective evaluation.

Not surprisingly then, students in the older grades expressed a greater preference for self-directed learning activities. Older students were less concerned with pleasing the teacher and consenting to the answer held by the class. And, students at the higher grade levels were less inclined to believe that only one solution is best for a complex problem.

Some of the interesting results of this attitude assessment were not the statistically significant results from our inferential statistical tests. The descriptive statistics indicating the overall attitude positions, across groups and across grade levels suggest some very heartening affective responses to math and science from students which may be, in part, the result of almost two decades, of the intensive math-science curriculum development efforts.

The students we tested, both USMES and control, were very positive in their attitudes toward arithmetic. They enjoyed it very much, and this enjoyment increased over the school year and perhaps further over grade levels. (Statements about increased enjoyment in higher grades must be tenuous because the research was cross sectional and not longitudinal.)

The factor structure emerging from the factor analysis of our scale suggested that students could distinguish between the attributes of enjoyment and value with respect to arithmetic. Not only did the students in our sample enjoy arithmetic very much, they also valued it highly, and their average description of value to arithmetic was heightened almost to the limit of our measurement scale on the post-test.

These expressions of positive regard for arithmetic were corroborated, in part, by the results of our interviews with 120 USMES children. When asked what they had done in school that year that they particularly enjoyed, approximately half of the student interviewers responded "math!" or "arithmetic!" without prompting.

The descriptive statistics summarizing the treatment groups' pre- and post-test positions, at each grade level, on the scale of science appreciation were also noteworthy. For all categories, initially positive expressions of science appreciation became more positive over the course of the school year. This result may be indicative of a slowing-down or a reversal in the trend of older students' generally negative attitudes toward science and scientists which was observed in the 1950's (Heath, Maier, Remmers, & Rogers, 1957). Almost two decades of intensive science curriculum development activity have followed. Many studies of the Cognitive outcomes of new methods of science instruction have appeared since 1957. However, it is difficult to obtain a picture of the effects of these curricula on the development of affective behaviors of students; the cognitive studies proliferate, but the research on affective responses to science curricula is disproportionately smaller (Kahn & Weiss, 1973, p. 784).

Only one treatment difference approached statistical significance closely, but that difference may have special practical significance, especially in light of the measure yielding that treatment difference at $p < .07$, as shown in Table 7.37. When post-test scores on Factor 1 were adjusted for pre-test difference, the analysis revealed that a significant treatment difference may exist. USMES students tended to express greater appreciation for science than control students.

CHAPTER VIII

MORALE PROBLEM: POSSIBLE FOCUS FOR FUTURE EVALUATIONS

This chapter will focus briefly on an issue which is clearly outside the design for this evaluative study, but which has been voluntarily and forcefully brought to the attention of the evaluators at almost every field site that was visited. The issue holds critical potential for the future maintainence and development of the USMES program. This chapter summarizes the issue and recommends a more structured, objective investigation of it in future evaluative projects.

This 1974-75 report, like that submitted for the previous year, documents strong support for the USMES project, both in its conceptual design and in its classroom implementation. The sources of these evaluations have included teachers and principals engaged in the use of the USMES program.

In apparent contradiction, several of these same subjects engaged us in confidential interviews at almost every site we visited and communicated to us attitudes of disappointment and disenchantment with the USMES program. We estimate that these confidential interviews and their negative content represent about 50% of the teachers and administrators using USMES on the local level. Their comments are representatively paraphrased as follows:

> At this school, only one-third of the original USMES teachers are still using USMES. Some of our "drop-outs" are quite bitter.
>
> (a principal): While I am very supportive of the USMES program, I can't get any of my teachers to use it.
>
> Other USMES teachers in my school are so clearly dishonest in their abuse of the USMES training incentives, I no longer want to be identified with the program.
>
> "X" is an exemplary USMES teacher, on paper, but in fact he never uses USMES in his classrooms and his logs are pure fiction.

> The training program, flight trips, etc., are a total rip-off of the American taxpayer.
>
> Our team leader is too autocratic and, thus far, has turned off three USMES teachers.
>
> Don't say anything bad about me back in Boston, Tell EDC what a good job I am doing so I can go back to Boston again next summer.

If these complaints are interpreted as critical of the program itself, they clearly conflict with the more objective and carefully analyzed information reported throughout this study. Nor did we, as participating interviewers, understand their thrust to be in this direction. What we did hear were indications of a serious morale problem among the USMES trained teachers and principals on the local level. And this on two counts: (1) Principals supporting USMES in their schools feel they are not in communication with the program's officers, especially in regard to the changes in direction which have occurred over the past three years. (2) The pivotal teacher representing USMES in each local school frequently is a personality who says all the right things about USMES but does not do USMES in the classroom. His leadership is seen as more persuasive than honest. In these cases, he is not viewed either as a teacher of integrity or as a genuine implementer of the USMES philosophy. This representative figure, however, "personifies" USMES to his colleagues on the local level; to all those within his sphere of influence, his failings become representative of the value-estimate of the entire USMES program.

Because of our assurances of confidentiality, and because we designed no instruments which would test these injections against objective evidence, the observations presented here must remain on the level of hearsay. It must remain the task of some future study to determine if, in fact, such a morale problem does exist; to examine the extend of its spread; to estimate the depth of its effect; to gauge the degree to which it endangers the future of the USMES development; to identify its causes; and to suggest possible methods of correction.

Our research indicates only that the USMES program, in comparison to other NSF curricular programs, has a distinct philosophy which relies heavily on the personal quality of its teachers. "USMES is more a philosophy than a set of materials." This factor makes the USMES method of selecting, motivating and training personnel more critical and more vulnerable than is true of other NSF programs.

CHAPTER IX

REPORT SUMMARY[a]

Focus for the 1974-75 Evaluation Project

This document reports on the 1974-75 USMES evaluation investigating the cognitive and affective responses of USMES students to this interdisciplinary, process curriculum. It includes the re[illegible]ts of a pre-post control group design to assess the curriculum's effects on students' basic skill development, their att[illegible]de change, and their progress in complex problem solving. The results of interviews with USMES teachers and students, unstructured observations at field sites, and the field staffs' documentation of USMES usage are also included. A second report will document work on the development of new techniques for assessing student progress in complex problem solving.

The original proposal to the National Science Foundation for the continued evaluation of USMES d[illegible]ring 1974-75 was broader in scope than the plan which was funded. As amended, the 1974-75 USMES Evaluation focused on student effects of the program: their abilities in problem solving; their basic skill development; their attitudes toward math, science, problem solving, and toward various learning activities embodied by the USMES philosophy. Teacher training, support networks for USMES users, formative program monitoring, material resource usage, and program dissemination patterns were deleted as areas for investigation. Clearly, the Foundation's overriding concern for an evaluation of USMES was the pursuit of an investigation of its "proof of concept," i.e. the examination of the students' problem solving abilities and basic skills as they develop under the influence of the USMES program.

[a]Essentially a summary of chapter summaries, this chapter can be obtained separately as a summary of the report.

Sample Selection, Data Collection, and Method of Analysis

Purposive sampling of new and experienced USMES teachers from 15 geographic areas was used to achieve a sample of USMES classes representing a cross section of grade levels, socio-economic levels, and unit challenges. Control classes came from non-USMES schools which were located in the same or neighboring communities as the USMES schools. These control classes were selected to match the USMES sample classes, one-for-one, on the bases of grade level, socioeconomic level, geographic area, and general features of the schools' program.

Responsible field staff personnel were trained to serve as on-site evaluators for the test administration and for the observation of class activities in these USMES and control sample classes. Interviews were completed by the evaluators with all 40 USMES teachers and 120 students in the evaluation sample. However, the maximum sample size achieved for other areas of the data collection was 37 USMES classes and 34 control classes.

The sample attrition from the proposed complement of 40 USMES and 40 control teachers can be attributed to problems with two observers who did not meet their commitments to data collection, and to the very stringent requirements of one state's law for permissions for pupil testing.

Several indicators in addition to the interview technique were used to acquire data on the program's effects on student performance. The pre-test, post-test control group design governed data collection on students' basic skill development, their performance in problem solving and changes in their attitudes toward math, science, problem solving and various learning activities. Six subtests from the Stanford Achievement Test battery were selected to measure basic skills. Problem solving ability was assessed with the Picnic

Problem and the Playground Problem--two tests of small group performance in simulated, real-life problem situations. A Likert-type attitude scale was developed to investigate attitude change.

The student performance data were submitted to two-factor repeated measures analyses of variance to determine if the treatment groups at each grade level had realized statistically significant gains from pre-test to post-test administration on any of the measures of performance. Whenever the data warranted and assumptions could be met, covariance analyses were also used to test the hypotheses that there were no significant differences between the treatment groups' adjusted post-test performances.

Characteristics of Experimental and Control Classes

One essential component of this evaluation is a description of what the USMES experimental program is like in practice, and how it differs from the treatment being applied to the comparative control groups. This documentation was necessary to make meaningful comparisons between the performance scores of students in the experimental program and the measures for students who did not receive the innovative curriculum.

USMES classes in the sample received diverse applications of the USMES program. Some USMES classes experienced brief applications of the program throughout the entire school year, while others had their USMES time concentrated in intensive periods over a few weeks only. Many combinations of levels of intensity and duration of usage were reported by the sample USMES teachers, but, on the average, classes spent 1½ hours a day, three days each week, for 12 weeks on an USMES unit. Most classes worked on only one unit during the year.

For most USMES classes, the time for USMES came primarily form science time. The statistically "average USMES class" reportedly borrowed some

Sample Selection, Data Collection, and Method of Analysis

Purposive sampling of new and experienced USMES teachers from 15 geographic areas was used to achieve a sample of USMES classes representing a cross section of grade levels, socio-economic levels, and unit challenges. Control classes came from non-USMES schools which were located in the same or neighboring communities as the USMES schools. These control classes were selected to match the USMES sample classes, one-for-one, on the bases of grade level, socioeconomic level, geographic area, and general features of the schools' program.

Responsible field staff personnel were trained to serve as on-site evaluators for the test administration and for the observation of class activities in these USMES and control sample classes. Interviews were completed by the evaluators with all 40 USMES teachers and 120 students in the evaluation sample. However, the maximum sample size achieved for other areas of the data collection was 37 USMES classes and 34 control classes.

The sample attrition from the proposed complement of 40 USMES and 40 control teachers can be attributed to problems [illegible]

Problem and the Playground Problem
simulated, real-life problem situa
developed to investigate attitude

The student performance data w
measures analyses of variance to d
grade level had realized statistic
post-test administration on any of
data warranted and assumptions cou
used to test the hypotheses that t
tween the treatment groups' adjust

Characteristics of Experimental an

One essential component of thi
USMES experimental program is like
treatment being applied to the com
was necessary to make meaningful c
of students in the experimental pr

two tests of small group performance in
ons. A Likert-type attitude scale was
ange.

e submitted to two-factor repeated
ermine if the treatment groups at each
ly significant gains from pre-test to
he measures of performance. Whenever the
be met, covariance analyses were also
re were no significant differences be-
post-test performances.

Control Classes

evaluation is a description of what the
n practice, and how it differs from the
rative control groups. This documentation
parisons between the performance scores
ram and the measures for students who did

control classes were using or had used USMES.

Despite this variety in the nature and intensity of the treatments whi the groups received, the results from the 1974-75 Class Activity Analysis dicated that there were clearly distinguishable differences observed in th kinds of activities pursued by USMES versus control students. Teachers co tinued to dominate class activity 16% to 20% of the time in both treatment groups. However, during the remaining observed class time, USMES students exhibited a wider repertoire of behaviors, and they spent larger amounts o time in more active, self-directed, and creative behavior than the control students. When the control students were not focusing on their teachers, they were spending much of the balance of the observed class time in very structured activities--prestructured reading, prestructured writing, and calculating--probably on worksheets or in workbooks for mathematics and/or science.

Interview Results: Teachers, Students, Administrators

The interview was retained as a data collection strategy for the continued evaluation of USMES during 1974-75 because of the valuable insights gained from the previous year's interview results. A sample of 40 USMES teachers was selected to be interviewed by the two senior members of the evaluation team during the first three months of 1975.

For each USMES teacher interviewed three of his/her students were also interviewed. In addition, the evaluation staff members spoke informally with the administrators in the schools of both the USMES and control teachers.

The interviews focused on the effects of USMES or student and teacher behavior. Other issues arose in the course of the interviews with teachers and administrators, but these were discussed in a separate section of the report, since they do not relate directly to the questions in the interview schedules.

Most of the points in the chapter on the interview data came from pairs of sources: teachers and students; teachers and administrators. This built-in system of checks helped to establish the validity of the information. Another source of information, the observers at each site, served as an additional check-point.

There was no disagreement on the subject of children's enjoyment of USMES. The children did enjoy USMES and they looked forward to using it. All agreed that each child derived something from the program: increased knowledge in content areas, or ability to solve problems, or socialization skills, or increased feeling of self-worth, or a confirmation of all three. What each child derived from USMES appeared to be a function of the teacher, the "challenge," and the child.

The philosophy of USMES received complete support from teachers and administrators. There was not a single instance of anyone in either group questioning the value of a problem-solving approach in education. Since these largely self-selected USMES users favored a real problem-solving approach, it follows that they also favored an integrated approach to teaching the disciplines, in order to solve the problems. And in theory, they did. In practice, there were difficulties. Departmentalized programs, rigid time schedules, and most teachers' limited content background (especially science) made the integrated approach difficult.

The nature of the USMES challenge was another factor which made the problem-solving approach and the integration of the disciplines difficult to implement. Some challenges simply did not lend themselves to a problem solving approach. Very often, the challenge was not perceived as a problem by the children, who simply saw what they did as a series of unrelated activities. In some instances, even the teacher did not perceive the USMES unit as a problem.

Administrators and teachers supported this perception by asking whether USMES was teaching problem solving or was just a series of activities, often seen as "gimmicks." For those students and teachers who saw a challenge as a problem, there was some feeling that the method of solution was generalizing to other areas.

While USMES appeared to be teaching new skills, it was seen mainly as reinforcing old learning. Teachers and students had no difficulty identifying the specific aspects of mathematics being learned, but neither students nor teachers could identify very much science involved in the program. Other content areas, e.g., language arts, social science, were identified by some teachers as being heavily involved in the program.

Although the content emphasis was a function of the particular challenge, it was also a function of a specific teacher's likes and dislikes. Teachers still tended to stress those areas which interested them or which they felt most comfortable teaching. And so, they tended also to choose those challenges with which they felt most comfortable. As a result, those units which emphasized the social science contents of mathematical applications were most often used.

Teachers continued to learn to use the program through workshops or by word-of-mouth rather than by using the manual. Other materials developed specifically for USMES, e.g., how-to-cards and technical papers, were also getting minimal usage, both by students and teachers. Even the Design Lab usage declined noticeably from the previous year.

All-in-all, however, the interviews revealed that USMES appeared to be fulfilling some of its promises. There were indications that children felt capable of dealing with their environment, and that teachers, through less directive teaching, were encouraging children to solve their own problems. USMES seemed to be changing the behavior of both teachers and students, in what the developers could view as a positive way.

Basic Skill Development

Development of problem solving abilities and basic skill development are seen as two interdependent tasks for the USMES program. To fully evaluate the first, an examination of the second must also be made.

A pre-test, post-test control group design was used to investigate whether USMES students maintain the same level of basic skill development as control students, even though USMES usage may detract from the amount of basic skills instruction which USMES students can receive. Basic skill development was measured with Fall and Spring administrations of selected subtests of the Stanford Achievement Test Battery: (a) Reading Comprehension, (b) Mathematics Computation, (c) Mathematics Application, (d) Mathematics Concepts, (e) Science, and (f) Social Science.

Several problems were encountered in the collection, analysis and interpretation of the SAT data. However, none of these problems prevented an unequivocal response to the expressions of concern about accountability for the communication of basic skills. Clearly, USMES students do not fall behind their control counterparts in their performance on tests of basic skills. On all six post-test measures, the overall USMES mean was higher than the overall control mean, but the differences were not statistically significant.

The USMES program purports to enhance the problem solving ability of elementary school students without impairing their basic skill development. Indeed, the results of our analyses of basic skills data suggest that fears about impairment of basic skills of USMES students are unwarranted. The question which may merit further investigation is not whether USMES detracts from basic skill development, but whether USMES enhances basic skill development, especially for students in the higher grades of elementary school.

The largest increases in basic skill development were observed between the lower grade blocks. Of course, this pattern of development refelcts the growth curve found for many areas of intellectual and physical development. While not statistically significant, there was a noteworthy trend for the growth rates of the control classes to fall behind those for the USMES classes at the upper grade levels.

Both USMES groups and control groups experienced similar rates of development in the basic skills in the early and middle elementary grades. Furthermore, both these control and USMES classes were generally close in average scores at pre-test time and at post-test time. However, the pattern changed for USMES and control groups in the higher elementary grades. While the USMES group exhibited continued growth in each of the six subtest areas, the control group revealed less growth or even showed a decline in performance.

This observation may be indicative of the effect of the USMES program on basic skill development at the higher elementary grade levels, but it may also be a function of sampling bias. Control classes were matched with USMES sample classes on the basis of grade level, socioeconomic composition and type of school program (traditional, "open," "non-graded," etc.), and the pairs of USMES and control classes came from neighboring schools. Nevertheless, careful matching on the most salient criteria is no assurance of comparability of treatment groups on all relevant factors other than the treatment.

Another point was raised in the discussion of basic skill development of USMES versus control students. It was a point of information which the evaluators uncovered during their interviews with USMES teachers. They found that USMES students have not been deprived of instruction in the basic skills. In some cases, they may have been getting more than the non-USMES students. Without exception, in all the USMES classes we interviewed, mathematics continues to be taught as a separate content area. These students were getting their ordinary math instruction plus "USMES math." One might hypothesize, therefore, that USMES students should exceed the control group in mathematics performance.

As for science, in the majority of cases, USMES was the science program, for a portion of the school year, or for the entire school year, and one wonders if there would be any science if USMES was not presented. Therefore, in the math and science skills areas, USMES should not be interferring with basic skill development, but rather, adding to it.

In other skill areas, i.e. reading, language arts, and social science, our other sources of data support the conclusion that USMES is not taking much time from these areas. Some of the classes are not self-contained, and for these classes, schedules and amounts of time are mandated for basic skills instruction.

Clearly, USMES usage, as practiced by sample classes representing a wide distribution of geographic areas and socioeconomic levels, did not affect basic skill development adversely. Previous investigations on this issue yielded similar results. The measurement of basic skills has been a costly and time consuming activity and sample teachers, principals, and their students have become increasingly resentful that this kind of test administration is disruptive of the school day, and sometimes is threatening to students.

The evaluators recommend that the resources devoted to comparing the basic skill development of USMES and control students should not be expended in the future. Moreover, the issue of basic skill development should be of diminished importance in light of the patterns of USMES usage in most schools. Most frequently, the time for USMES comes from regularly scheduled science time, and to a lesser extent from project time. Hence, one should not expect

USMES children to fall behind in the basic skills areas of reading, language arts, and mathematics which are of primary concern to most elementary school personnel. Stated simply, the issue is not an issue.

Proof of Concept Assessment

An objective assessment of proof of concept of the USMES curriculum was limited by the primitive state of the art of measuring problem solving abilities in elementary school children. As the evaluation team pursued a two-fold thrust of program evaluation and new instrument development, we applied the most satisfactory existing measures of problem solving to answer immediate needs shared by the developers and the funding agency about the progress of USMES students in real, complex problem solving.

These measures were the Playground Problem and the Picnic Problem. The conceptual bases for these simulated, real-life-relevant problem tasks reflected John Dewey's conceptualization of the problem solving process, whose "five logically distinct steps" permeate much of the literature about USMES prepared by the USMES Central Staff.

Designed as parallel forms of one another, both problem tests are accompanied by manuals for trained administrators' presentation of the tests to groups of five children.

The scoring protocols developed for the tests offer both cognitive and affective assessments. The cognitive scores provide indices of the students' abilities to identify, measure, calculate, and record data on factors which they think are salient to the solution of the problems. The behavioral assessments include ratings on motivation to accept the problem, commitment to task, efficiency of manpower, and the nature of group leadership. Additionally, the protocol for the Playground Problem afforded an assessment of the students' product: their drawing of the play area design.

Neither the Playground Problem nor the Picnic Problem satisfied the program developers' concerns that these tests meet all of their criteria for "realness." Therefore, rigorous investigation of these tests' reliability and statistical validity did not seem to be warranted. Content validation of the tests as simulated measures of life-like, complex problem solving was established.

No differences between USMES and control students were noted in the behavioral aspects of their work on the problems. The four cognitive scores were subjected to repeated measures analyses of variance and to analyses of covariance. Consistently, significant differences among grade levels were observed for all four cognitive aspects of the students performance. As one might expect, the older students in both treatment groups outperformed the younger students. They identified more factors and progressed to more frequent, higher level measuring, calculating, and recording on these factors. However, no significant differences between treatment groups were found on any of the ratings derived from the scoring protocol.

Attitude Changes in USMES and Control Students

Having analyzed the cognitive effects of the USMES program, on its students, the evaluation team then truned its focus on the affective dimension. What is the impact of USMES on the students' attitudes?

An attitude scale was developed and pilot tested by the evaluation team especially for this USMES evaluation. The scale consisted of two parts. Part I contained items designed to measure attitudes toward math and science and toward various teaching strategies and learning activities which are embodied in the USMES approach. Part II began with a statement of a real-life complex problem facing a group of engineers, and then followed with a series of items.

Factor analyses of the largest data base, the post-test attitude results for 1491 students, yielded 7 factors on Part I and 4 factors on Part II. The repeated measures analyses revealed highly significant ($p < .0001$) pre-to-post test administration differences for most of the 11 factors. Only the academic insecurity scale from Part II produced no significant pre- post-test differences.

However, these pre-to-post differences may be indicative only of "time-of-year" effects, as discussed by Ahlgren. This evaluation did employ a control group design but fall-fall testings were not feasible because samples would not have been accessible the second fall. There were no treatment by test administration interactions resulting from our analyses. Both treatment groups moved in the same direction. But in light of Ahlgren's and his associates' observations, our results are very interesting because on most of the scales, the groups attitudes moved toward the more socially or academically desirable direction, up or down, over the course of the school year.

The time-of-year effect observed by Ahlgren et al. is probably evidenced in the overall decline of students' scores on factor 4, Group Learning, and on factor 6, Self-Directed Learning, over the period from Fall to Spring. At the end of the school year, students, perhaps, are tired or less motivated, and would prefer to play a more passive role in learning, as the teacher "runs the show."

While there were no significant treatment differences revealed by repeated measures analyses, there were a number of significant grade differences in attitude factors which have implications for USMES development and USMES usage, even though the grade differences did not interact with treatment when covariance adjustments were made. Guidelines for curriculum development and implementation which come from research in developmental psychology are reinforced by some of the grade differences observed in this affective evaluation.

Students in the upper grades expressed a greater preference for self-directed learning activities, were less concerned with pleasing the teacher and consenting to the answer held by the class, and were less inclined to believe that only one solution is best for a complex problem.

The students we tested, both USMES and control, were very positive in their attitudes toward arithmetic. They enjoyed it very much, and this enjoyment increased over the school year and perhaps further over grade levels. Statements about increased enjoyment in higher grades must be tenuous because the research was cross sectional and not longitudinal.

The factor structure emerging from the factor analysis of our scale suggested that students could distinguish between the attributes of enjoyment and value with respect to arithmetic. Not only did the students in our sample enjoy arithmetic very much, they also valued it highly, and their average description of value to arithmetic was hightened almost to limit of our measurement scale on the post-test.

These expressions of positive regard for arithmetic were corroborated, in part, by the results of our interviews with 120 USMES children: When asked what they had done in school that year that they particularly enjoyed, approximately half of the student interviewers responded "math!" or "arithmetic!" without prompting.

The descriptive statistics summarizing the treatment groups' pre- and post-test positions, at each grade level, on the scale of science appreciation were also noteworthy. For all categories, initially positive expressions of science appreciation became more positive over the course of the school year. This result may be indicative of a slowing-down or a reversal in the trend of older students' generally negative attitudes toward science and scientists which was observed in the 1950's (Heath, Maier, Remmers, & Rogers, 1957). Almost two decades of intensive science curriculum development activity have followed. Many studies of the cognitive outcomes of new methods of science instruction have appeared since 1957. However, it is difficult to obtain a picture of the effects of these curricula on the development of affective behaviors of students; the cognitive studies proliferate, but the research on affective responses to science curricula is disproportionately smaller (Kahn & Weiss, 1973, p. 784).

Only one treatment difference closely approached statistical significance, but that difference may have special practical significance: USMES students tended to express greater appreciation for science than did control students.

Recommendation for a Future Study

Considerable information was volunteered to the senior evaluators by teachers and administrators at almost all of the USMES sites on a singular theme. In effect, the "feedback" indicated a serious morale problem growing among the implementation and developmental teachers and their sponsoring principals, threatening a possible movement to disengage from the program. The evaluators were not charged with the investigation of this issue, nor did they have the necessary instruments to document of "objectivize" its content. However, because of the serious possible implications of this issue for the future dissemination of the USMES program, a recommendation for its future investig[illegible]ication was advanced.

Conc[illegible]

In sum, this evaluation documents the decided perceptions of the USMES teachers that the USMES program does teach problem solving skills to its students, while [illegible]e objective instruments to measure problem solving skills are still too unsophisticated to give an accurate reading of this same question. Basic skills of USMES students, according to both teacher perceptions and objective tests, have not suffered. Additionally, results from the interviews with teachers and students documented the "excitement" for learning self-initiation and social interaction skills acquired by students in USMES classes. In their work on real problems, USMES students sensed that their efforts can make a difference.

REFERENCES

Ahlgren, A., Christensen, D.J., & Kim-Shing Lun. Minnesota school affect assessment manual. Minneapolis Center for Educational Development, University of Minnesota, 1973.

Aiken, L.R., Jr. Personality correlates of attitudes toward mathematics. Journal of Educational Research, 1963, 56, 476-480.

Aiken, L.R., Jr. Attitudes toward mathematics. Review of Educational Research, 1970, 40, 551-596.

Bender, Hilary. A synopsis of the evaluator's report on unified science and mathematics for elementary schools (USMES) during the 1973-74 school year. Report to the National Science Foundation, Materials and Instruction Development Section. Boston University, September, 1975.

Catalog of tests and related services. Test Department, Harcourt, Brace, Jovanovich, Inc. 1973-1974.

Charters, W.W., Jr., & Jones, J.E. On the risk of appraising non-events in program evaluation. Educational Researcher, Vol. 2, No. 11. November, 1973, pp. 5-7.

Cochran, W.G. Analysis of covariance: Its nature and uses. In J.A. Steger (Ed.), Readings in statistics. New York: Holt, Rinehart and Winston, 1971. Pp. 179-194.

Cochran, W.G. Some consequences when the assumption for analysis of variance are not satisfied. In J.A. Steger (Ed.), Readings in statistics. New York: Holt, Rinehart and Winston, 1971. Pp. 330-346.

Covington, M.V. "A childhood attitude inventory for problem solving." Berkeley: University of California, unpub. mimeo, 1967.

Cronbach, L.J., & Furby, L. How we should measure "change:" Or should we? Psychological Bulletin, 1970, 74, 68-80.

Dutton, W.H. & Blum, M.P. The measurement of attitudes toward arithmetic with a Likert-type test. Elementary School Journal, 1968, 68, 259-264.

Dutton, W.H., & Stephens, L. Measuring attitudes toward science. School Science and Mathematics, 1963, 63, 43-49.

Easton, D., & Dennis, J. Political socialization of the elementary school child. National Council for the Social Studies Yearbook, 1966, 36, 216-235.

Education Development Center. Unified science and mathematics for elementary schools, Newton, Mass.: September, 1972.

Education Development Center. USMES systems approach to development, widespread implementation and maintainence of a real problem solving program in elementary schools. Newton, Mass.: March, 1975.

Education Development Center. The USMES guide. Newton, Mass.: May, 1974.

Edu[illegible]ion Development Center. Curriculum correlation guide (preliminary [illegible]tion). Newton, Mass.: April, 1975.

[illegible]odon [illegible].P. The role of attitude in learning arithmetic. Arithmetic Teacher, [illegible]58, 5, 304-310.

Ferguson, G.A. Statistical analysis in psychology and education, (3rd ed.) New York: McGraw-Hill, 1971.

Friendenberg, E.Z. Coming of age in America: Growth and acquiescence. New York: Random House, 1965.

Guilford, J.P. "Intellectual factors in productive thinking. Productive thinking in education. Mary Jane Aschner and Charles Bish (Ed.). Washington, D.C.: National Education Association, 1965.

Heath, R., Maier, M., Remmers, H., & Rogers, D. High school students look at science. The Purdue Opinion Panel (Report of Poll No. 50), Purdue University, November, 1957.

Holt, J. How children fail. New York: Pitman, 1964.

Kahn, S.B., & Weiss, J. The teaching of affective responses. In R. Travers (Ed.) Second handbook of research on teaching. Chicago: Rand McNally, 1973. Pp. 759-804.

Kehoe, J. Students' interests in the social studies. Canadian Journal of History and Social Science, 1970, 6, 34-39.

Kerlinger, F.N. Foundations of behavioral research. (2nd ed.) New York: Holt, Rinehart and Winston, 1973.

Madden, R. et al. Stanford Achievement Test Series. New York: Harcourt, Brace, Jovanovich, 1972.

Messick, S. The criterion problem in the evaluation of instruction: Assessing possible outcomes, not just intended outcomes. In M.C. Wittrock, & D.E. Wiley (Eds.), The evaluation of instruction: Issues and problems. New York: Holt, Rinehart & Winston, 1970. Pp. 183-202.

Nie, N.H., Bent, D.H., & Hull, C.H. Statistical package for the social sciences. New York: McGraw-Hill, 1970.

Scriven, M. Student values as educational objectives. In _Proceedings of 1965 invitational conference on testing problems_. Princeton, N.J.: Educational Testing Service, 1966. Pp. 33-49.

Shann, M.H. _An evaluation of unified science and mathematics for elementary schools during the 1973-74 school year_. Report to the National Science Foundation, Materials and Instruction Development Section. Boston University, August, 1975.

Shapiro, B.J., & Aiello, T. _Report on basic skill development_. Report to the National Science Foundation. Boston University, June, 1974.

Shaw, M.E., & Wright, J.M. _Scales for the measurement of attitudes_. New York: McGraw-Hill, 1967.

Silberman, C.E. _Crisis in the classroom_. New York: Random House, 1970.

Stake, R.E. Objectives, priorities, and other judgment data. _Review of Educational Research_, 1970, 40, 181-212.

USMES News. Newsletter for Unified Science and Mathematics for Elementary Schools. Newton, Mass.: October, 1975, Vol. 3, No. 1.

APPENDIX A

School Information Form
1974-75

School Information Form
1974-75

Instructions: Please fill out one form for each USMES and Control school in the evaluation sample.

1. Name of School: ______________________________

2. Address of School: ______________________________
no. street

____
city state zip

3. Setting (check one): _____rural _____suburban _____city

4. Approximate socioeconomic level (check one): _____high
_____medium
_____low

Complete the following questions for USMES schools ONLY

5. Design Lab
   a. Does the school have a Design Lab? _____Yes _____No

   (If no, skip the rest of question #5)

   b. Is there a Design Lab manager? _____Yes _____No

   c. If "yes," what is the position of the manager?

   _____Teacher _____High school student
   _____Teacher aide _____Principal
   _____Volunteer _____Other (please explain)

   d. How many hours per week is the manager in the Lab?

   _____Hours every week

   _____He is not assigned for regular hours each week but usually comes in for about ______________ each week.

6. How many classes in the school did at least one complete USMES unit this year? ______________

7. How many classes in the school did at least one complete USMES unit last year? ______________

APPENDIX B

Class Information Form

1974-75

USMES Evaluation
Class Information Form
1974-75

Instructions: Please fill out one form for each USMES and Control class in the evaluation sample:

1. Observer: ______________________________

2. School: ______________________________

3. Address of School: ______________________________
number street

____
city state zip

4. Teacher: ______________________________

5. Grade(s): ________ 6. USMES class____ 7. # of Children ____
Control class____

8. a. If USMES: name of unit(s) completed this year: ______________

____

____

____

b. If Control: _____self-contained classroom

_____if not self-contained, subject(s) taught by teacher:

____

____

____

____

307

To Be Answered By USMES and Control Teachers:

1. We are interested in learning what materials or programs are used in the teaching of math, science and social studies. Please check which categories are appropriate (many checks may be appropriate). It is not necessary to give names of materials (except under "other"), but simply to check appropriate categories.

	Science	Math	Social Studies
1. Uses a single text			
2. Uses many books			
3. USMES			
4. Teacher made materials only			
5. Some teacher made materials			
6. SCIS			
7. ESS			
8. SAPA			
9. An individualized program			

2. List the approximate number of hours spent per week, on the following subjects and activities, for weeks when USMES was and was not done.

A Weeks When USMES IS Done		B Weeks When USMES is NOT Done
a. ______	a. Math	a. ______
b. ______	b. Science	b. ______
c. ______	c. Social Science	c. ______
d. ______	d. Language Arts	d. ______
e. ______	e. Music	e. ______
f. ______	f. Art	f. ______
g. ______	g. Physical Education	g. ______
h. ______	h. Special Projects	h. ______
i. ______	i. Other (specify: ______	i. ______
______	______)	______

3. How many years has this teacher been teaching?: ______

If the above checklist (for question #1 only) is not appropriate for your program, briefly describe your program.

To Be Answered By USMES and Control Teachers:

The following question pertains to the math and science background of each teacher. Please check appropriate items. (That is, if a teacher had an undergraduate math minor, put a check (✓) in that box. If a teacher has a graduate science major, put a check (✓) in that box. Many categories may be checked.)

	Science	Math	Social Science
A. Undergraduate Program			
1. Major			
2. Minor			
3. Education Methods Course			
4. If not major or minor: non-education courses			
B. Graduate Program			
1. Major			
2. Minor			
3. Education Methods Course			
4. a. Inservice Courses			
4. b. If not major or minor: non-education courses			

To Be Answered By Control Teachers ONLY

1. Has the control teacher used any of the USMES materials this year? Unit Resource Books, How-to-cards, Technical Papers, Design Lab? ______

______

2. Have the control children used any of the USMES materials or been exposed to USMES activities? ______

______

______

If yes, please explain: ______

______

______

______

______

______

3. Does the control teacher use the USMES philosophy in her class? ______

______

______

______

______

To Be Answered By USMES Teachers ONLY

1. How many years has the teacher been using USMES? ______________

2. What units, other than the present one, has the teacher ever used? ______

____

____

3. We are trying to determine how various USMES units are used! Please respond to the following questions for each unit used this year. (If this format seems completely inappropriate to the unit you've done, write a brief description on the back of this page describing the amount of time spent on the unit.)

	Unit 1	Unit 2	Unit 3
Name of unit			
About how many weeks was USMES used?			
About how many days per week?			
About how many hours per day was it used?			

4. Does the teacher express an interest in using USMES next year?

Yes ____ No ____ Other ____

APPENDIX C

USMES Classroom Activity Analysis

1974-75

1974 - 1975

USMES CLASSROOM ACTIVITY ANALYSIS

ss: ______________________ Date: ______________________

ool: ______________________ USMES/Control: ______________________

t/Subject: ______________________ Observer: ______________________

Classroom Organization

Large Group	+	+	+	+	+	+	+	+
Small Group	+	+	+	+	+	+	+	+
Individual	+	+	+	+	+	+	+	+

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

			1	2	3	4	5	
Activities		Measures						(16-17)
		Counts						(18-19)
		Constructs/Assembles						(20-21)
		Graphs						(22-23)
		Tests/Experiments						(24-25)
		Calculates						(26-27)
		Records Data						(28-29)
		Writes composition/illustrates						(30-31)
		Writes (prestructured)						(32-33)
		Reads How-to cards; Plays tapes						(34-35)
		Reads (prestructured)						(36-37)
		Free Reading						(38-39)
Interactions	Children	Talks to another -task						(40-41)
		Talks to another- social						(42-43)
		Takes part in small group discussion - task						(44-45)
		Takes part in small group discussion - social						(46-47)
	Teacher	Gives pre-structured information to teacher						(48-49)
		Gives original information to teacher						(50-51)
		Seeks information from teacher						(52-53)
		Talks to teacher social						(54-55)
Look/Listen		Takes part in class discussion or presentation						(56-57)
		Listen/look at child						(58-59)
		Listen/look at small group						(60-61)
		Listen/look at class						(62-63)
		Listen/Look at teacher						(64-65)
		Listen/look at film or AV Materials						(66-67)
		Collecting material/maintence						(68-69)
		Resting/Waiting						(70-71)
		Fooling Around						(72-73)

APPENDIX D

ADMINISTRATOR'S MANUAL

for

THE PLAYGROUND PROBLEM

A Measure of Problem Solving Ability for

Use in the Evaluation of USMES

Prepared by

The USMES Evaluation Staff
Boston University

Mary H. Shann, Ph.D.

USMES Evaluation Project Director

TO THE OBSERVER:

This Manual and the accompanying materials consist of the following:

1. Instructions to guide you in the administration of the Playground Problem
2. A catalog of playground equipment
3. A form on which to record your observations of the children's behaviors
4. A cassette tape for recording various segments of the sessions.

GENERAL INSTRUCTIONS

The problem solving behaviors of elementary school children constitute one of the most important areas for evaluation of the USMES program. The Playground Problem is to be used as one means of assessing the success of the USMES program in reaching its goals. This test is designed to enable the observer to collect data on both verbal and non-verbal behaviors involved in problem solving.

The Playground Problem should be administered to designated USMES classes and control classes. Five children are to be selected randomly from each USMES class and similarly from each control class in the evaluation sample. The test is to be given to each group of five children rather than to individuals.

Each group of children should be taken to an open area near the school and asked to plan a playground. The materials the children are to use in solving the problem, the instructions you are to give them, and the role you are to play as an observer will be explained in detail shortly.

We are interested in assessing the degree of cooperation and self- or group-motivated interest the children demonstrate during the entire problem solving period and the follow-up question period. We are equally interested in the degree to which the children employ practical considerations in solving the problem.

Our analysis of the Playground Problem test results will be based on three kinds or records: (a) a tape recording of the children's verbal presentation during the follow-up question period; (b) your observations of the children's behaviors as recorded on the observation form accompanying this Manual; and (c) a layout of the proposed playground which the children will be asked to draw on a large sheet of paper.

In general, your role as an observer will be to organize the test session, to instruct the children on what to do, and to observe and record their behavior. Specific instructions for administration of the Playground Problem are given in the following sections of this Manual.

ORGANIZATION

1. Selection of Children

A random sample of five children should be picked from each control class and each USMES class in your school. In the past, children have not always been picked randomly, and this is not acceptable. When children are picked on the basis of good academic performance on the one hand, or on the basis of "getting rid of the troublemaker" on the other, the entire session will have to be disregarded.

It would be best for you to pick the children yourself, but the teacher can also make the selections if correct procedures are used. The easiest appropriate method is to write the names of each child on a piece of paper, throw each piece in a hat, and then select five.

2. When to Administer the Playground Problem Test

This can be a critical factor. Oftentimes, children are more restless and less attentive at certain times of the day, and especially at certain times of the year--for example, the day before Christman Vacation.

Try to run your test sessions at approximately the same time of day--that includes the control classes as well as the USMES classes. The recommended time of day is as close to the beginning of the day as possible. Avoid extremely cold or rainy days, since the Playground Problem is to be administered outside.

Second, do not run your test sessions on the day before or after vacation periods, or on the days when special school events are to take place. In the past, some sessions have had to be discounted because of confounding factors of this nature.

In all of these considerations, use your own good judgement. A test administered under somewhat less than ideal conditions is probably better than no test returns at all for a class.

3. Where to Administer the Playground Problem Test

In preparation for the test, you should locate a suitable open area near the school. An empty lot would be ideal. However, if one is not available, a playing field or clear black topped area would be appropriate. This area should be the same for all groups of children in the same schools on your sample list, both USMES groups and control groups.

4. Materials to Accompany Test Administration

Prior to the testing session, you will need to gather together the following items:

Observation Equipment

Observation form

Tape recorder and blank cassette

Watch

Tools (in a cardboard box)

50 foot tape measure
Yard stick
Ball of string
Large piece of paper
Tri-wall (to use as hard surface for drawing plan)
Felt tip pens
Pencils
12" rulers
Catalog of playground equipment
Scrap paper
Scissors

INSTRUCTIONS TO THE CHILDREN

Soon after arriving at the open area, you should give the children the following instructions and you should record them on tape:

"Let's suppose this area was going to be made into a new playground for the children in your school." (Indicate clearly the limits of the area). "How would you plan this playground?"

"Here is a catalog of playground equipment which could be bought. If you had $2,000 to spend, which equipment would you choose?"

"Please work together to decide which equipment should be bought. Draw a plan of the playground on this piece of paper showing where the equipment would be placed."

"You have forty minutes to work together to make your plan. Here are some things you may use if you want to." (Hand one child the box containing the tape measure, pencils, etc.) "Remember, you can spend up to $2,000 on equipment."

DO NOT GIVE THE CHILDREN ANY SUGGESTIONS AS TO WHAT OTHER CONSIDERATIONS THEY SHOULD KEEP IN MIND. In the past, some test results have had to be invalidated because of suggestions and clues which observers had given to the children in the instructions. The instructions should be as similar as possible for the USMES groups and for the control groups. Any evidence of intentional or unintentional bias unfortunately results in invalidation of the test session.

Let the children know that they will have forty minutes to figure out their plan and draw it on paper. Tell them that at the end of this period, you will ask them questions about their plan, and that their answers will be recorded on tape (more about taping later).

OBSERVATIONS

During the forty-minute problem solving period, stay in the area in view of the children. You can repeat the instructions, if necessary. However, you should not participate in the problem solution by answering other questions or suggesting possible strategies. It is up to the children to decide whether or not to use the measuring equipment. Do not demand that any particular child help out in planning the playground if he or she does not want to.

After thirty minutes of the problem solving period have expired, tell the children that they have ten minutes to complete drawing their plan if they have not already done so.

During the forty minute problem solving period, the observer should make notes on the observation form describing the children's activities. Please write clearly. Each activity should be noted under the appropriate category heading. These notes should be specific and numbered sequentially. For example, under the heading "Measuring" the observer might note:

"5. Two kids measured the width of the lot with the 50' tape." The number "5" indicates that this is the fifth note the observer has made on the observation form. The next note might be:

"6. One child recorded the width of the lot as 45 feet." This observation would be placed under the heading "Recording Data."

You will have received intensive training in the use of this observation form at the Observers' Training Workshop.

PREPARATION FOR TAPING

After the forty minute problem solving period is completed, you should call the children together to prepare for tape recording the ten minute question period.

Children are often shy or giggly when they first speak into a microphone. Inaudiable responses make our work of analysis very difficult. To get around this problem, please ask each child to recite a sentence into the microphone, such as: "This is our plan," or "My name is ...." Tell the children that they must speak one at a time, and ask them to speak slowly and clearly.

Play the tape back to the children. This will give them some chance to get used to recording their voices, and it will give you a chance to see how well their voices are being picked up. (Note: this part of the recording is not important to us and can be erased).

When the entire session is over, we would like to have the following recordings returned to us:

Part 1: the instructions as you gave them originally to the children

Part 2: the ten minute question period given after the thirty minute problem solving period and after the practice taping.

QUESTION PERIOD

This period during which the children explain their plan and outline their reasoning should be tape recorded in its entirety. The children's presentation may be up to ten minutes long. You should record the data and group at the beginning of each question period taping. If you wish, you may take the children back into the school to make the recording.

It is very important to remember that the questions you ask the children and the procedures you use in soliciting their answers MUST be as similar as possible for the USMES groups and for the control groups. Again, any evidence of bias may invalidate the results.

Although you may have to use your imagination and various strategies to encourage the children to respond or to explain what they mean in greater detail, use the following "script" as a guide to the specific questions you should ask. It is very helpful, we are sure you know, if you show interest and enthusiasm in what the children have done. Remind the children to speak slowly and clearly so that other people can understand what they have said later. Do not rush the children but rather gently encourage them to say what they want.

FIRST QUESTION SERIES (Directed to the entire group*)

-- "How did you do?"

-- "Was it fun?"

SECOND QUESTION SERIES (Directed to the entire group*)

-- "Explain your playground plan."

-- "Why did you decide to but (4) pieces of equipment?"

-- "Do you know how much the equipment you have chosen will cost?"

-- "Why did you decide to put the swings over here? The slide over here?"

-- "What kinds of information did you need to help you make your decisions?"

THIRD QUESTION SERIES (Directed first to the entire group, and then to each child in turn who has not yet responded)

-- "Were there any other important factors you had to consider in making your decisions?"

-- "Is there anything anyone would like to say before we finish?"

While it may be necessary to structure the children's report by asking questions, you as the observer should not suggest rationale to the children by means of your questioning. For example, if there has been no mention of safety factors or indications that the issue of safety has been taken into consideration, the observer should not bring it up during the tape recording.

The playground problem does not have one solution. However, in the playground problem, a certain approach to problem solving is valued. An excellent response to the playground problem would include:

1. Measurement or calculation of available space.
2. Meaningful use of measuring equipment
3. Careful consideration of types of playground equipment chosen.
4. Comparisons between size of equipment as listed in catalog and space available on playground area.
5. Consideration of budget limitations.
6. Accuracy in drawing lay-out of proposed playground.
7. Consideration of human elements such as safety and aesthetic appeal.
8. Logical and clear presentation of rationale.

---

* When the question is directed to the entire group make sure that everyone talks who wants to, not only the "spokesman" for the group. Be sure they talk one at a time so that it is easy to understand what is being said.

However, particularly on the pre-test, the children may not respond in this manner. This in itself is interesting and important data and should not be interpreted as resulting from the format of the problem.

After the testing session is over, review the tape on your own. If you think any part of the conversation will be difficult for us to understand, please make a note of what was said and attach it to the observation form. Please be sure to return to us all tapings, observation sheets, scrap papers the students wrote on, and the playground layouts. The pre-test results should be sent to us soon after they have been completed. The Playground Manual and Catalog should be retained by you after administration of the pre-tests. They should be used again for administration of the posttests. Upon completion of the post-tests, please return to us the Manual and Catalog along with the testing results for the post-test.

Instructions for administration of this Playground Problem will have been reviewed in detail at your Observers' Training Workshop. However, if you have any further questions when you are ready to administer the test, please call the USMES Evaluation Team, collect, at (617) 353-3312.

Dr. Mary H. Shann

APPENDIX E

ADMINISTRATOR'S MANUAL
for
THE PICNIC PROBLEM

A Measure of Problem Solving Ability for
Use in the Evaluation of USMES

Prepared by

The USMES Evaluation Staff
Boston University

Mary H. Shann, Ph.D.
USMES Evaluation Project Director

TO THE OBSERVER:

This Manual and the accompanying materials consist of the following:

1. General instructions to help guide you in the implementation of the picnic problem.

2. Observation sheets upon which all of your observations and notes should be made.

3. Park Map and Photograph of Picnic Foods for use by the children.

4. Cassette Tape for recording various segments of the session.

GENERAL INSTRUCTIONS

The problem solving behaviors of elementary school children constitute one of the most important areas for evaluation of the USMES program. The Picnic Problem is to be used as one means of assessing the success of the USMES program in reaching its goals. This test is designed to enable the observer to collect data on both verbal and non-verbal behaviors involved in problem solving.

The Picnic Problem should be administered to designated USMES classes and control classes. Five children are to be selected randomly from each USMES class and similarly from each control class in the evaluation sample. The test is to be given to each group of five children rather than to individuals.

Each group of children should be brought to a separate room if possible, or some other quiet location, where they are to be given a common problem to be solved, in this case, the Picnic Problem. The materials the children are to use in solving the problem, the instructions you are to give them, and the role you are to play as an observer will be explained in detail shortly.

We are interested in assessing the degree of co-operation and self or group-motivated interest the children demonstrated during the entire problem-solving period and the follow-up question period. We are equally interested in the degree to which the children employ practical considerations in solving the problem.

Our analysis of the Picnic Problem Test results will be based on three kinds of records: (a) a tape recording of the children's verbal presentation during the follow-up question period; (b) your observations of the children's behaviors as recorded on the observation form accompanying this Manual; and (c) the pieces of scrap paper on which the children recorded measurements or made calculations.

Your role as an observer will be to organize the test session, to instruct the children on what to do, and to observe and record their behavior. Specific instructions for administration of the Picnic Problem are given in the following sections of this Manual.

ORGANIZATION

1. Selection of Children

A random sample of five children should be picked from each control class and each USMES class in your school. In the past, children have not always been picked randomly, and this is not acceptable. When children are picked on the basis of good academic performance on the one hand, or on the basis of "getting rid of the troublemaker" on the other, the entire session will have to be disregarded.

It would be best for you to pick the children yourself, but the teacher can also make the selections if the correct procedures are used. The easiest appropriate method is to write the names of each child of a piece of paper, throw each piece in a hat, and then select five.

2. When to administer the Picnic Problem

This can be a critical factor. Oftentimes, children are more restless and less attentive at certain times of the day, and especially at different times of the year--for example, the day before Christmas vacation.

Try to run your test sessions at approximately the same time of day--that includes the control classes as well as the USMES classes. The recommended time of day is as close to the beginning of the day as possible.

Secondly do not run your sessions on the day before or after vacation periods, or on the days when special school events are to take place. In the past some sessions have had to be discounted bacause of confounding factors of this nature.

In all of these considerations, use your own good judgement. A test administered under somewhat less than ideal conditions is probably better then no test returns at all for a class.

3. Where to administer the Picnic Problem

The instructions given to the children, the actual problem solving period and the follow-up question period should all take place in the same area and it should be the same area for all groups of children (i.e. both USMES and control groups).

The ideal location for the sessions would be a quiet room where there is minimal possibility for distractions.

4. Materials to Accompany Test Administration

Prior to the testing session, you will need to gather together the following items:

Observation Equipment

Observation form
Tape recorder and blank cassette
Watch

Tools (In a cardboard box)

Yard stick
12" rulers
Ball of string
Scissors
Pencils
Scrap paper
50 foot tape measure

Other Materials

Map of parks
Photograph of foods

INSTRUCTIONS TO THE CHILDREN

After the children are in the test area and you have their attention, you should give the children the following instructions and you should record your instructions on tape.

"You have been picked to take part in a game to see how well you can plan a picnic, as a group." (Try to get the children's enthusiasm and interest by asking them a few questions about their own experiences, if they went on any picnics last summer, etc).

"Let's suppose that you are asked to plan a picnic for 25 children and that you will have $50 to spend."

"Let's suppose that none of the parks allows Bar-B-Qing, but that you can order food for your picnic from a food service which has stands at the picnic areas i each park." (Point out the picnic areas on the map). "You must place your order 2 (two) days ahead of time so that they will have enough food on hand."

"Here is a picture showing the foods you may order and the price of each item: Hamburgers are 50¢ each; hotdogs are 30¢ each; soda is 20¢ a can; potato chips are 10¢ a bag; and ice cream cones or ice cream sandwiches are 20¢ each."

"This map shows the areas you can choose for the picnic. Each park charges admission." (Review the map of the picnic areas with the children. Point out the admission charges per person for each park, and explain the various symbols on the map). For example, "This symbol indicates a playground, and here are the playgrounds in each park." (Do likewise for all the other symbols "Notice that the map is drawn to scale, and 1" on the map equals 10 miles."

"Your transportation will be provided via school bus free of charge. You may spend from 10:00a.m. to 4:00 p.m., from the time you must board the bus until the time you must be back at the school."

"Please work together to decide where you would choose to go for this picnic, and what foods you would buy."

"You have forty minutes to work together to make your plan. Here are some things you may use if you want. (Hand one child the box containing the rulers, pencils, etc.) "Remember, you can spend up to $50 and that your time is from 10:00a.m. to 4:00p.m. including time spent traveling in the bus."

DO NOT GIVE THE CHILDREN ANY SUGGESTIONS AS TO WHAT OTHER CONSIDERATIONS THEY SHOULD KEEP IN MIND. In prior years, some of the test results had to be invalidated because of suggestions or clues given to the children. Any evidence of intentional or unintentional bias unfortunately results in invalidation of the test session. The instructions should be <u>as similar as possible</u> for

USMES groups and for control groups.

Let the children know that they will have forty minutes to figure out their plan. Tell them that at the end of this period, you will ask them questions about their plan, and that their answers will be recorded on tape (more about taping later).

OBSERVATIONS

During the forty minute problem solving period, stay in the area in view of the children. You can repeat the instructions, if necessary. However, you should not participate in the problem solution by answering other questions or suggesting possible strategies. It is up to the children to decide whether or not to use the measuring equipment. Do not demand that any particular child help out in planning the picnic if he or she does not want to.

After thirty minutes of the problem solving period have expired, tell the children that they have ten minutes to complete their plan if they have not already done so.

During the forty minute problem solving period, the observer should make notes on the observation form describing the children's activities. Please write clearly. Each activity should be noted under the appropriate category heading. These notes should be specific and numbered sequentially. For example, under the heading "Measuring" the observer might note:

"4. Two kids measured the distance to each park with string.'

"5. Two kids measured the string distances against a ruler."

The numbers "4" and "5" indicate that these are the fourth and fifth notes the observer has made on the observation form. The next note might be:

"6. One child converted string lengths to distances in miles."

This observation would be placed under the heading "Calculating."

You will have received intensive training in the use of the observation form for the Picnic Problem at the Observers' Training Workshop.

PREPARATION FOR TAPING

After the forty minute problem solving period is completed, you should call the children together to prepare for tape recording the ten minute question period.

Children are often shy or giggly when they first speak into a microphone. Inaudible responses make our work of analysis very difficult. To get around this problem, please ask each child to recite a sentence into the microphone, such as: "This is our plan," or "My name is..." Tell the children that they must speak one at a time, and ask them to speak slowly and clearly.

Play the tape back to the children. This will give them some chance to get used to recording their voices, and it will give you a chance to see how well their voices are being picked up. (Note: this part of the recording is not important to us and can be erased).

When the entire session is over, we would like to have the following recordings returned to us:

Part 1: the instructions as you gave them originally to the children.

Part 2: the ten minute question period given after the forty minute problem solving period and after the practice taping.

QUESTION PERIOD

This period during which the children explain their plan and outline their reasoning should be tape recorded in its entirety. The children's presentation may be up to ten minutes long. You should record the date and the group at the beginning of each question period taping.

It is very important to remember that the questions you ask the children and the procedures you use in soliciting their answers MUST be as similar as possible for the USMES groups and for the control groups. Again, any evidence of bias may invalidate the results.

Although you may have to use your imagination and various strategies to encourage the children to respond, or to explain what they mean in greater detail, use the following "script" as a guide to the specific questions you should ask. It is very helpful, we are sure you know, if you show interest and enthusiasm in what the children have done. Remind the children to speak slowly and clearly so that other people can understand what they have said later. Do not rush the children but rather gently encourage them to say what they want.

FIRST QUESTION SERIES (Directed to the entire group*)

-- "How did you do?"

-- "Was it fun?"

SECOND QUESTION SERIES (Directed to the entire group*)

-- "Explain your plans for the picnic."

-- "Which park did you choose? Why?"

-- "Which foods did you choose? Why?"

-- "Do you know how much the picnic will cost?"

-- "What kinds of information did you need to help you make your decisions?"

THIRD QUESTION SERIES (Directed first to the entire group, and then to each child in turn who has not yet responded)

-- "Were there any other important factors you had to consider in making your decisions?"

-- "Is there anything anyone would like to say before we finish?"

While it may be necessary to structure the children's report by asking questions, you as the observer should not suggest rationale to the children by means of your questioning. For example, if there has been no mention of distance factors or indications that the traveling time has been taken into consideration, the observer should not bring it up during the tape recording.

- -

* When the question is directed to the entire group make sure that everyone talks who wants to, not only the "spokesman" for the group. Be sure they talk one at a time so that it is easy to understand what is being said.

The Picnic Problem does not have one solution. However, in the Picnic Problem, a certain approach to problem solving is valued. An excellent response to the Picnic Problem would include:

1. Measurement and calculation of the distances to each park.
2. Meaningful, efficient use of measuring equipment.
3. Careful consideration of the advantages of each park.
4. Consideration of reasonable quantities and the variety of foods chosen.
5. Weighing the admission costs to parks against the costs of the foods desired.
6. Consideration of budget limitations
7. Consideration of human elements such as taste preferences and activity preferences
8. Logical and clear presentation of rationale.

However, particularly on the pre-test, the children may not respond in this manner. This in itself is interesting and important data and should not be interpreted as resulting from the format of the problem.

After the testing session is over, review the tape on your own. If you think any part of the conversation will be difficult for us to understand, please make a note of what was said and attach it to the observation form. Please besure to return to us all tapings, observation sheets, and scrap papers the students wrote on. The pre-test results should be sent to us soon after they have been completed. The Picnic Problem Manual, map and photograph should be retained by you after administration of the pretests. They should be used again for administration of the posttests. Upon completion of the posttests, please return to us the Manual, map and photograph along with the testing results for the posttest.

Instructions for administration of this Picnic Problem will have been reviewed in detail at your Observers' Training Workshop. However, if you have any further questions when you are ready to administer the test, please call the USMES Evaluation Team, collect, at (617) 353-3312.

Dr. Mary H. Shann

APPENDIX F

Scoring Protocol for the Playground Problem: A Manual for Rating and Coding Students' Performance On a Test of Complex Problem Solving

Prepared by

Mary H. Shann, Ph.D.

USMES Evaluation Project Director

Boston University

1974

Section I.--IDENTIFICATION (Columns 1-20)

I.D. code records the teacher grade level, unit and other descriptive information related to reliability and validity issues.

Column 1: identifies form of the problem-solving test.

6 = Playground

7 = Picnic

Column 2: identifies time of testing.

1 = Pre-test

2 = Post-test

Column 3: identifies treatment.

1 = USMES

2 = Control

Columns 4,5: identify teacher.

(See master list for teacher codes)

Column 6,7,8: identify grade level.

(See master list for grade level codes)

In columns 9 and 10 enter the unit code as follows:

Advertising	01
Bicycle Transportation	02
Burglar Alarm Design (now called <u>Protecting Property</u>), (may also be called <u>Security</u> by some teachers)	03
Classroom Design	04
Classroom Management	05

Community Gardening	06
Consumer Research	07
Describing People	08
Designing for Human Proportions	09
Design Lab Design	10
Dice Design	11
Eating in School	12
Getting in Shape	13
Getting There (formerly Finding Your Way, Getting From Place to Place)	14
Growing Plants	15
Lunch Lines	16
Making School Safer	17
Manufacturing	18
Mass Communications (formerly Mass Media)	19
Nature Trails	20
Orientation (formerly Student Migration)	21
Pedestrian Crossings	22
Planning Special Occasions	23
Play Area Design and Use	24
School Rules (formerly School Rules and Decision Making)	25
School Supplies (formerly Managing and Conserving School Resources), (or Recycling)	26
School Zoo (formerly Outgrowth of Animal Behavior, and Ecosystems which are no longer units)	27
Soft Drink Design	28
Sound in the Environment (formerly Outgrowth of Music which is no longer a separate unit)	29

Traffic Flow 30

Using Free Time (formerly Designing Indoor/Outdoor Games) 31

Using Free Time After School (After School Activities) 32

Ways to Learn 33

Weather Predictions 34

										Blank
1	2	2	4	5	6	7	8	9	10	11

Column 11: Leave Blank

Based on your review of the audio tape and observer's notes, indicate whether you think any of the following factors may render this testing session invalid. Code your response 0 = No, 1 = Yes in the appropriate column.

Problem	Column
Biased selection of students	12
Prompting by observer	13
Prior student experience with this test	14
Inclement weather (0 for picnic problem)	15
Noisy testing environment	16
Outside interference/interruptions	17
Observer deviated from standard procedure	18
Blank	19-20

Section II.--BEHAVIORAL ASPECTS (Columns 21-24)

There are four factors which are considered in this segment. The scoring of this group shall proceed as follows:

Factor: 1

Motivation: to accept the problem and attempt to solve the problem.

Scoring:

0 No one accepts problem or trys to solve problem.

1 1 Student accepts/trys to solve problem.

2 2 Students accepts/trys to solve problem.

3 3 Students accepts/trys to solve problem.

4 4 Students accepts/trys to solve problem.

5 5 Students accepts/trys to solve problem.

Enter the proper score in column 21.

Factor: 2

Commitment to task: the level of itensity of the group to continue working toward a solution.

Scoring:

0 No effort.

1 Disinterested, fooling around, little input.

2 Some positive input (one or two interested in problem and working with little progress).

3 Group is interested but efforts are not organized, and time is being wasted.

4 Group is interested, working and not wasting time or effort.

Enter proper score in column 22.

Factor: 3

Organization: allocation of responsibilities for efficiency of manpower.

Scoring:

0 No effort.

1 Unplanned, haphazard, or chaotic (students do their own thing - do not allocate item or all work on the same thing).

2 No all students involved (either by choice or flat). Some are working on problem some are not - may be arguing among each other.

3 Students have allocated some tasks - may have some working on same item; or possibly 1 may not be involved.

4 Tasks are allocated and students working efficiently-however students may have trouble with their item and seek help.

5 Tasks allocated and all are working productively.

Enter proper score in column 23.

tor: 4

Structure: Group leadership

Scoring:

0 None

1 Autocratic--one person dominates who does not listen to other students' ideas.

2 Minority Leadership--one or two persons listen to others and then lead or direct.

3 Plurality--general agreement of several members leads to direction and leadership; most contributions are recognized and evaluated.

4 Democratic--all students contribute; no one's suggestions are ignored or ridiculed. One spokesman may arise but sources of ideas/efforts are recognized.

Enter proper score in column 24.

21	22	23	24

Section III.--COGNITIVE ASPECTS (Columns 25-56)

Data for this section can be derived primarily from the observer form and the tapes. It will be necessary to read the observer form and listen to the tapes to bridge any apparent gaps or vague statements found in either the form or the tape.

The cognitive aspects shall include variables considered in solving the problem and the level or method of measuring the variables. The implementation of the measurement in terms of calculation and the recording of the data will be collected and encoded.

A total of 10 variables can be accommodated by the scoring protocol. For each variable, its identification, measurement, calculation and recording will be scored.

IIIA. Factor: COST OF EQUIPMENT

Identification:

Scoring: 0 No

1 Yes

Enter in column 25.

Measurement:

Scoring

0 No measurement done.

1 Vague or very general estimates.

2 Estimations by imprecise methods or by eyeballing. It does not provide enough information to arrive at a decision.

3 Useful information which can be used to arrive at a decision but the data should be more accurate or precise.

4 Precise measurement or clearly appropriate data that can lead to solution.

Enter in column 26.

Calculations:

Scoring:

0 No calculations.

1 Vague or very general calculations that do little quantification.

2 Calculations are imprecise or guesses are arrived at by trial and error and are not sufficient to provide necessary data to arrive at a solution.

3 Useful calculations which can be used to arrive at a solution. It may not be accurate or have considered totals or balances. It should be more precise.

4 Calculations are appropriate, precise and can lead to a solution.

Enter in column 27.

Recording:

Scoring

0 records.

1 Very general or imprecise records.

2 Adequate records.

Enter in column 28.

25	26	27	28

IIIB. Factor: SIZE OF EQUIPMENT VS. SIZE OF CHILDREN (i.e., larger scale equipment for older children; smaller scale equipment for younger children)

Identification:

Scoring: 0 No

1 Yes

Enter in column 29.

Measurement:

Scoring: 0 No measurement.

1 Vague or general estimates, i.e., big equipment for big kids.

2 Express need to know proportion of big and small kids in their school.

Enter in column 30.

Calculations:

Scoring: 0 No calculations.

1 General or arbitrary assignment of equipment for size of children i.e., for example "lets get half big equip ent; half small."

2 More careful estimates on how many big and small kids attend their school and selections of equipment reflects distribution of size of students.

Enter in column 31.

Recording:

Scoring: 0 No records.

1 Very general or imprecise records.

Enter in column 32.

29	30	31	32

IIIC. Factor: SIZE OF EQUIPMENT VS. AREA AVAILABLE (e.g., a swing will use 100 sq. feet and we have 1000 sq. feet all together to use.)

<u>Identification:</u>

Scoring:

0 No

1 Yes

Enter in column 33.

<u>Measurement:</u>

Scoring:

0 No measurement done.

1 Vague or very general estimates.

2 Estimations by imprecise methods or by eyeballing. It does does not provide enough information to arrive at a decision.

3 Useful information which can be used to arrive at a decision but the data should be more accurate or precise.

4 Precise measurement or clearly appropriate data that can lead to solution.

Enter in column 34.

<u>Calculations:</u>

Scoring:

0 No calculations.

1 Vague or very general calculations that do little quantification.

2 Calculations are imprecise or guesses are arrived at by trial and error and are not sufficient to provide necessary data to arrive at a solution.

3 Useful calculations which can be used to arrive at a solution. It may not be accurate or have considered totals or balances. It should be more precise.

4 Calculations are appropriate, precise and can lead to a solution.

Enter in column 35.

<u>Recording:</u>

Scoring:

0 No records.

1 Very general or imprecise records.

2 Adequate records.

Enter in column 36.

33	34	35	36

IIID. Factor: CAPACITY OF EQUIPMENT
(e.g., 4 kids can use a swing set with four seats; more kids can use a big jungle jim.)

Identification:

Scoring: 0 No
1 Yes

Enter in column 37.

Measurement:

Scoring: 0 No measurement.
1 Vague or general estimates; i.e., big stuff can be used by more kids.
2 Express need to know specific number of children who can use each piece of equipment at one time.

Enter in column 38.

Calculations:

Scoring: 0 No calculation.
1 General estimates of capacity (e.g., most of the kids in a class could use something at the same time).
2 Precise figures on capacity (e.g., altogether, the equipment we choose will handle 25 kids at one time).

Enter in column 39.

Recording:

Scoring: 0 No records.
1 Very general or imprecise records.
2 Adequate records.

Enter in column 40.

37	38	39	40

IIIE. Factor: DURABILITY OF EQUIPMENT (i.e., stronger, lasts longer)

Identification:

Scoring: 0 No

1 Yes

Enter in column 41.

Measurement:

Scoring: 0 No measurement.

1 Vague statements, i.e., its better.

2 General/precise, i.e., stronger, lasts longer.

Enter in column 42.

Calculations:

Scoring: 0 No calculations.

1 Calculations in a general or vague sense.

Enter in column 43.

Recording:

Enter 0 in column 44.

41	42	43	44

IIIF. Factor: PLACEMENT OF EQUIPMENT FOR SAFETY CONSIDERATIONS

Identification:

Scoring: 0 No

1 Yes

Enter in column 45.

Measurement:

Scoring: 0 No measurement.

1 General or vague statements of more or less safety.

2 More precise measures of safety, i.e., more distance so kids do not run into the other stuff.

Enter in column 46.

Calculations:

Scoring: 0 No calculations.

1 Vague as to placement, i.e., that close enough.

2 Some concept of calculation, i.e., about 6 ft. or the like.

Enter in column 47.

Recording:

0 No records.

1 Records.

Enter in column 48.

45	46	47	48

111G. Factor: PLACEMENT OF EQUIPMENT FOR EFFICIENT UTILIZATION OF AREA

Identification:

Scoring: 0 No

1 Yes

Enter in column 49.

Measurement:

Scoring: 0 No measurement.

1 Vague or general statements, i.e., it fits.

2 More precise statements of placement based on size or shape of equipment or terrain.

Enter in column 50.

Calculations:

Scoring: 0 No calculations.

1 General or vague calculation based on placement and practical considerations, e.g., putting it there leaves us with more space for playing ball.

Enter in column 51.

Recording:

Scoring: 0 No records.

1 Very general or vague records.

Enter in column 52.

49	50	51	52

IIIH. Factor: OTHER CONSIDERATIONS

Column:

53 Number of additional factors mentioned.

54 "Fun" mentioned as consideration (0=no, 1=yes).

55 "Appeal of equipment for all ages" mentioned as consideration (0=no, 1=yes).

56 Blank

Enter in column 53-56.

53	54	55	56

Section IV.--PRODUCT ASPECTS (Columns 57-60)

Evaluation of four product aspects shall be based on the students' drawing of their playground design.

The Product - Plan

Scale:

Scoring:

- 0 No scale.
- 1 Approximate scale that indicated relative size of equipment; representations of distances are reasonable.
- 2 Scale is precise or is coded.

Enter in Column 57.

Labels:

Scoring:

- 0 No labels.
- 1 Labels are present and appropriate to equipment.

Enter in Column 58.

Landmarks:

Scoring:

- 0 No landmarks.
- 1 Landmarks are present.
- 2 Landmarks are present, appropriate and/or coded, i.e., enduring and relevant to playground area.

Enter in Column 59.

Area:

Scoring:

- 0 No area limitations.
- 1 Area is defined.

Enter in Column 60.

57	58	59	60

APPENDIX G

Scoring Protocol for the Picnic Problem: A Manual for Rating and Coding Students' Performance On a Test of Complex Problem Solving

Prepared by

Mary H. Shann, Ph.D.

USMES Evaluation Project Director

Boston Univeristy

1974

Section I.--IDENTIFICATION (Columns 1-20)

I.D. code records the teacher grade level, unit and other descriptive information related to reliability and validity issues.

Column 1: identifies form of the problem-solving test.

6 = Playground

7 = Picnic

Column 2: identifies time of testing.

1 = Pre-test

2 = Post-test

Column 3: identifies treatment.

1 = USMES

2 = Control

Columns 4,5: identify teacher.

(See master list for teacher codes)

Column 6,7,8: identify grade level.

(See master list for grade level codes)

In columns 9 and 10 enter the unit code as follows:

Advertising	01
Bicycle Transportation	02
Burglar Alarm Design (now called Protecting Property), (may also be called Security by some teachers)	03
Classroom Design	04
Classroom Management	[illegible]

Community Gardening	06
Consumer Research	07
Describing People	08
Designing for Human Proportions	09
Design Lab Design	10
Dice Design	11
Eating in School	12
Getting in Shape	13
Getting There (formerly Finding Your Way, Getting From Place to Place)	14
Growing Plants	15
Lunch Lines	16
Making School Safer	17
Manufacturing	18
Mass Communications (formerly Mass Media)	19
Nature Trails	20
Orientation (formerly Student Migration)	21
Pedestrian Crossings	22
Planning Special Occasions	23
Play Area Design and Use	24
School Rules (formerly School Rules and Decision Making)	25
School Supplies (formerly Managing and Conserving School Resources), (or Recycling)	26
School Zoo (formerly Outgrowth of Animal Behavior, and Ecosystems which are no longer units)	27
Soft Drink Design	28
Sound in the Environment (formerly Outgrowth of Music which is no longer a separate unit)	29

Traffic Flow	30
Using Free Time (formerly <u>Designing Indoor/Outdoor Games</u>)	31
Using Free Time After School (<u>After School Activities</u>)	32
Ways to Learn	33
Weather Predictions	34

										Blank
1	2	2	4	5	6	7	8	9	10	11

Column 11: Leave Blank

Based on your review of the audio tape and observer's notes, indicate whether you think any of the following factors may render this testing session invalid. Code your response 0 = No, 1 = Yes in the appropriate column.

Problem	Column
Biased selection of students	12
Prompting by observer	13
Prior student experience with this test	14
Inclement weather (0 for picnic problem)	15
Noisy testing environment	16
Outside interference/interruptions	17
Observer deviated from standard procedure	18
Blank	19-20

Section II.--BEHAVIORAL ASPECTS (Columns 21-24)

There are four factors which are considered in this segment. The scoring of this group shall proceed as follows:

<u>Factor: 1</u>

Motivation: to accept the problem and attempt to solve the problem.

Scoring:

0 No one accepts problem or trys to solve problem.

1 1 Student accepts/trys to solve problem.

2 2 Students accepts/trys to solve problem.

3 3 Students accepts/trys to solve problem.

4 4 Students accepts/trys to solve problem.

5 5 Students accepts/trys to solve problem.

Enter the proper score in column 21.

<u>Factor: 2</u>

Commitment to task: the level of itensity of the group to continue working toward a solution.

Scoring:

0 No effort.

1 Disinterested, fooling around, little input.

2 Some positive input (one or two interested in problem and working with little progress).

3 Group is interested but efforts are not organized, and time is being wasted.

4 Group is interested, working and not wasting time or effort.

Enter proper score in column 22.

<u>Factor: 3</u>

Organization: allocation of responsibilities for efficiency of manpower.

Scoring:

0 No effort.

1 Unplanned, haphazard, or chaotic (students do their own thing - do not allocate item or all work on the same thing).

2 No all students involved (either by choice or flat). Some are working on problem some are not - may be arguing among each other.

3 Students have allocated some tasks - may have some working on same item; or possibly 1 may not be involved.

4 Tasks are allocated and students working efficiently-however students may have trouble with their item and seek help.

5 Tasks allocated and all are working productively.

Enter proper score in column 23.

Factor: 4

Structure: Group leadership

Scoring:

0 None

1 Autocratic--one person dominates who does not listen to other students' ideas.

2 Minority Leadership--one or two persons listen to others and then lead or direct.

3 Plurality--general agreement of several members leads to direction and leadership; most contributions are recognized and evaluated.

4 Democratic--all students contribute; no one's suggestions are ignored or ridiculed. One spokesman may arise but sources of ideas/efforts are recognized.

Enter proper score in column 24.

21	22	23	24

Section III.--COGNITIVE ASPECTS (Columns 25-68)

Data for this section can be derived primarily from the observer form and the tapes. It will be necessary to read the observer form and listen to the tapes to bridge any apparent gaps or vague statements found in either the form or the tape.

The cognitive aspects shall include variables considered in solving the problem and the level or method of measuring the variables. The implementation of the measurement in terms of calculation and the recording of the data will be collected and encoded.

A total of 13 variables can be accommodated by the scoring protocol. For each variable, its identification, measurement, calculation and recording will be scored.

III A Factor: COST OF ADMISSION

Identification:
Scoring

0 No
1 Yes

Enter in Column 25.

Measurement:
Scoring

0 No measurement.
1 Vague or very general estimates of the cost of admission at each park.
2 Estimations by imprecise methods or by eyeballing. It does not provide enough information to arrive at a decision.
3 Useful information which can be used to help select park but data should be more accurate or precise.
4 Precise measurement of cost of admission for whole class at each park.

Enter in Column 26.

Calculations:
Scoring

0 No calculations.
1 Vague or very general calculations of cost of admission to each park.
2 Calculations are imprecise or guesses used as an estimate of cost. This is not sufficient to provide necessary data to arrive at a solution.
3 Useful calculations which can be used to arrive at solution, but the data should be more accurate or precise.
4 Calculations are appropriate and precise. Correct calculation of price of admission to each park for entire class.

Enter in Column 27.

Recording:
Scoring

0 No records.
1 Very general or imprecise records.
2 Adequate records.

Enter in Column 28.

25	26	27	28

III B Factor: COST OF FOOD

Identification:
Scoring

0 No

1 Yes

Enter in Column 29.

Measurement:
Scoring

0 No measurement done.

1 Vague or very general estimates of cost of food per person or for entire class.

2 Estimations by imprecise methods or by eyeballing. No attempt to plan menu. It does not provide enough information to arrive at a decision.

3 Useful information which can be used to help select food but data should be more accurate or precise. There is an attempt to plan menu for the class.

4 Precise measurement of cost of food for the entire class is made, staying within budget limitations. A menu is planned.

Enter in Column 30.

Calculations:
Scoring

0 No calculations.

1 Vague or very general calculations of cost of food per person or for entire class.

2 Calculations are imprecise or guesses used as an estimate of cost. No considerations of menu for each person or for entire class. This is not sufficient to provide necessary data to arrive at a solution.

3 Useful calculations which can be used to arrive at solution, but the data should be more accurate or precise. Consideration of menu takes place.

4 Calculations are appropriate and precise. Correct calculation of cost of food for entire class. Menu well planned out, which can lead to a solution.

Enter in Column 31.

Recording:
Scoring

0 No records.

1 Very general or imprecise records.

2 Adequate records.

Enter in Column 32.

29	30	31	32

III C Factor: COST OF FOOD vs. COST OF ADMISSION

Identification:
Scoring

0 No

1 Yes

Enter in Column 33.

Measurement:
Scoring

0 No measurement.

1 Vague or general awareness that cost of food and cost of admission must not exceed the $50. budget limit.

2 Estimates relationship of cost of food to admission cost by imprecise methods or by eyeballing it does not provide enough information to arrive at a decision.

3 Useful information which can be used to judge what proportion of money should be allocated to the food and to the cost of admission respectively, but the data should be more accurate or precise.

4 Precise measurement of relationship between cost of food and cost of admission. Allocates certain proportion of $50. to food and certain proportion to admission fee.

Enter in Column 34.

Calculations:
Scoring

0 No calculations.

1 Vague or very general calculations that do little quantification.

2 Calculations are imprecise or guesses used as an estimate of cost of food and admission. Little awareness of relationship between cost of food and cost of admission. This is not sufficient to provide necessary data to arrive at a solution.

3 Useful calculations which can be used to arrive at solution, but the data should be more accurate or precise. Is aware that certain proportion of money should be allocated to food and a certain proportion to admission.

4 Calculations are appropriate and precise. Correct calculation of both food costs and admission costs, keeping within a budget of $50.

Enter in Column 35.

Recording:
Scoring

0 No records.

1 Very general or imprecise records.

2 Adequate records.

Enter in Column 36.

III D Factor: TIME AVAILABLE FOR PICNIC

<u>Identification</u>:

Scoring	0	No
	1	Yes

Enter in Column 37.

<u>Measurement</u>:

Scoring	0	No measurement.
	1	Vague or very general awareness of time limit.
	2	Acknowledges time limitation of 6 hours, including travel time and time at park, and makes plan according to this time limit.

Enter in Column 38.

<u>Calculations</u>:

Scoring	0	No calculations.
	1	Vague or very general calculations involving travel time to each park. General awareness of time limitation as a consideration in choosing a park.
	2	More precise calculations of relative times to get to each park, and then relating travel time to time limitation of 6 hours.

Enter in Column 39.

<u>Recording</u>:

Scoring	0	No records.
	1	Very general or imprecise records.
	2	Adequate records.

Enter in Column 40.

37	38	39	40

III E Factor: TRAVEL TIME vs. PLAYTIME

Identification:
Scoring

0 No

1 Yes

Enter in Column 41.

Measurement:
Scoring

0 No measurement.

1 Vague or very general awareness that both time factors should be taken into consideration in choosing park.

2 More precise measurements of travel time to each park and judging what proportion of time should be spent traveling and what proportion of time should be spent for playing in in the park.

Enter in Column 42.

Calculations:
Scoring

0 No calculations.

1 Vague or general estimates of relative travel times to each park by eyeballing or guessing and then consideration and general estimation of time left over for play at each park.

2 More precise calculations of relative travel times to each park, and time left over for play at each park.

Enter in Column 43.

Recording:
Scoring

0 No records.

1 Very general or imprecise records.

2 Adequate records.

Enter in Column 44.

41	42	43	44

III F Factor: CONSIDERATION OF FOOD COST, ADMISSION COST AND TIME RELATIONSHIPS

<u>Identification:</u>

<u>Scoring</u>

0 No

1 Yes

Enter in Column 45.

<u>Measurement:</u>

<u>Scoring</u>

0 No measurement.

1 Vague or general statements regarding the relationship of the 3 factors, which are used to help lead to a solution.

2 More precise statements and/or estimates of the relationship of the three factors, which can help lead to a solution.

Enter in Column 46.

<u>Calculations:</u>

<u>Scoring</u>

0 No calculations.

1 Very general estimates of the relationship of food costs, admission costs and time. Weighing of the pros and cons of different alternatives occurs.

2 More precise calculations of different alternative solutions (regarding selection of food and a specific park), recognition of the relationship of the 3 factors, and selection of one alternative (e.g., calculates travel time, and amount of money left for food at each of the 3 parks).

Enter in Column 47.

<u>Recording:</u>

<u>Scoring</u>

0 No records.

1 Very general or imprecise records.

2 Adequate records.

Enter in Column 48.

45	46	47	48

III G Factor: DISTANCES TO PARKS

Identification:

Scoring

0 No

1 Yes

Enter in Column 49.

Measurement:

Scoring

0 No measurement.

1 Awareness that distance to each park should be taken into consideration. Vague or very general estimates are made.

2 Estimates by imprecise methods or by eyeballing.

3 Useful information which can be used to arrive at a decision. Recognition of the use of the map scale, but measurement should be more accurate or precise.

4 Precise measurement of distance to each park made, and recognition that travel time within the park to particular facilities should be included in the total distance to each park.

Enter in Column 50.

Calculations:

Scoring

0 No calculations.

1 Vague or very general calculations that do little quatification (e.g. Forest Valley Park looks twice as far away as Pine Hill Park).

2 Calculations are imprecise or guessing occurs and are not sufficient to provide necessary data to arrive at a solution, (e.g., Pine Hill Park looks about 30 miles away).

3 Useful calculations using the map scale which can be used to arrive at a solution. It may not be accurate or have considered distances to be traveled within the park to the facilities in to the total distance to be traveled to each park.

4 Calculations are appropriate, precise and can lead to a solution.

Enter in Column 51.

Recording:

Scoring

0 No records.

1 Very general or imprecise records.

2 Adequate records.

Enter in Column 52.

49	50	51	52

III H Factor: SIZE OF FACILITIES

Identification:

Scoring
- 0 No
- 1 Yes

Enter in Column 53.

Measurement:

Scoring
- 0 No measurement.
- 1 Vague or general estimates (i.e., Greehill Park is much bigger than Pine Hill Park).
- 2 More precise measures of the size of each park (i.e., using map scales to roughly measure the area of each park).

Enter in Column 54.

Calculations:

Scoring
- 0 No calculations.
- 1 General estimates of the size of each park, mainly by eye-balling.
- 2 More careful calculations, using the map scale to figure out the approximate areas of each park.

Enter in Column 55.

Recording:

Scoring
- 0 No records.
- 1 Very general or imprecise records.
- 2 Adequate records.

Enter in Column 56.

53	54	55	56

III I Factor: PLAY EQUIPMENT (Brought along for children's use at playground, e.g., baseballs and bats)

Identification:
Scoring

0 No

1 Yes

Enter in Column 57.

Measurement:
Scoring

0 No measurement.

1 Vague or general estimates of type of amount of equipment that should be brought to park.

2 Express need to know specific number of different pieces of equipment to be brought to park, taking in to consideration the number of children who would be using each particular piece of equipment.

Enter in Column 58.

Calculations:
Scoring

0 No calculations.

1 General or arbitrary assignment of equipment for children participating in the picnic.

2 More careful estimates, with selection of equipment reflecting individual child preferences, abilities and whether or not the amount of equipment brought along is in proportion to the number of children utilizing it.

Enter in Column 59.

Recording:
Scoring

0 No records.

1 Very general or imprecise records.

2 Adequate records.

Enter in Column 60.

57	58	59	60

III J Factor: SAFETY CONSIDERATIONS FOR TRIP

Identification:
Scoring

0 No

1 Yes

Enter in Column 61.

Measurement:
Scoring

0 No measurement.

1 General or vague considerations of safety precautions, and more or less safety of each park, (e.g., in the large park, there is a greater possibility of someone getting lost).

2 More precise safety measures taken, (e.g., specific assignment of adults for supervision on the bus and at the park.

Enter in Column 62.

Calculations:
Scoring

0 No calculations.

1 Vague or general references to safety precautions that should be taken.

2 More careful or precise calculations made in order to have a safe trip, (e.g., number of supervisors needed).

Enter in Column 63.

Recording:
Scoring

0 No records.

1 Records.

Enter in Column 64.

61	62	63	64

III K Factor: OTHER CONSIDERATIONS

Column:

65 Number of additional factors mentioned.

66 "Fun" mentioned as consideration, (0=no, 1=yes).

67 Blank.

68 Blank.

Enter in Column 65-68.

65	66	67	68

APPENDIX H

Teacher Interview

Spring, 1975

371

TEACHER INTERVIEW
Spring, 1975

Name ______________________________

School ______________________________

Unit Just Finished (Or Currently Working On) ______________________________

____

Total Number of Units Done This Year ______________________________

Total Number of Different Units Ever Done ______________________________

1. "What do you see as the primary goal of the USMES program?"

   _____a. Increased ability to solve problems

   _____b. Teach math and science

   _____c. Teach children to think and act independently

   _____d. Other

2. If real life problem solving is not mentioned, "To what extend do you see real life problem solving as a goal?" (If teacher cannot respond to "real life problem solving" reask the question, "To what extent do you see the solution of problems which are meaningful to the child in his environment, as a goal of USMES?")

   _____a. It's a very important goal

   _____b. Not very important

   _____c. I don't know what real life problem solving is

   _____d. Other

3. "Have the problems you've solved (as part of USMES) come about naturally or were they contrived?" If contrived, "How did you get the children interested in the problems?"

_____a. Naturally

_____b. Contrived

_____c. Teacher perceives them as "natural" but her description is of a contrived situation.

_____d. Other

4. "Was a solution to the problem found?"

_____a. Yes

_____b. No

_____c. Unit still in progress

5. "Were the children satisfied with the solution?"

_____a. Yes

_____b. No

_____c. Unit still in progress

6. "Did they feel the solution made a difference to anyone, that is, was it implemented?"

_____a. Yes

_____b. No

_____c. Unit still in progress

7. "Have you used the USMES approach to solve problems for which USMES units don't exist?"

_____a. Yes

_____b. No

If "yes" briefly describe the nature of the problem.

8. "Who raised the problem?"

_____a. Teacher

_____b. Child

_____c. Other

9. "Was the problem solved successfully?"

_____a. Yes

_____b. No

10. "How many USMES units have you ever used?"

____

11. "How many had you seen presented before you used them?"

____

12. For those not seen presented, "How did you learn to do the others?"

_____a. Manual

_____b. Just followed the natural direction of the problem as I had with the others

_____c. Talked it over with someone who had done the problem

_____d. Other

13. Have the manuals (or has the manual) been helpful?

_____a. Yes 13a. If "yes," in what ways?

_____b. No 13b. If "no," why not?

What would you like to see in the manual to make it more useful?

_____a. Logs are too wordy

_____b. A list of activities that I can do

_____c. Other

14. "Are you currently supplementing USMES with other math, science?"

_____a. Yes _____ Math only

_____b. No _____ Science only

_____c. I have to: we don't have a self-contained classroom and someone teaches math

_____d. Same as "c" except for "science"

15. "What math are they learning from the current unit?"

_____a. Measuring

_____b. Graphing

_____c. Reinforces math skills

_____d. Makes them see that math is useful in real world

_____e. Scale drawing

_____f. Other

16. "What science are they learning from the current unit?"

_____a. This unit doesn't really have any science in it

_____b. Other

17. "Where is the time coming from that you use for USMES?"

_____a. Regular science class

_____b. Project time

_____c. Other

18. "What is happening to children as a result of USMES?"

_____a. Nothing

_____b. They're more independent

_____c. They're more responsible

_____d. They enjoy school more

_____e. Other

19. "Do you feel that USMES is appropriate for all children or do you see some students who probably do not benefit from it at all?"

_____a. They all benefit from it

_____b. Immature

_____c. Inner-city

_____d. Irresponsible

_____e. Lacking in self-direction

_____f. Not motivated

_____g. There are enough activities so that something is beneficial to each child

_____h. It's too soon to tell

_____i. Other

20. "Do you find that as you use more units you handle them diffenently?"

_____a. Yes (how?)

_____b. Used only one

_____c. No

_____d. Other

21. "Do you see a cumulative effect on children as they are exposed to more units?"

_____a. Yes (what?)

_____b. No

_____c. Question not appropriate

_____d. They know what to do: They're more independent and go ahead and do what they should

22. "Has your perception of the program changed as you continue to use it?"

_____a. I like it more

_____b. I like it less

_____c. No

_____d. I thought of it as very non-directive but now see it is necessary to give some guidance

_____e. Other

23. "Do you see yourself as becoming more or less directive <u>in the USMES</u> units as you continue to use the program?"

_____a. More directive

_____b. Less directive

_____c. Not any more or any less, but I see direction as crucial at certain points

APPENDIX I

Child Interview
Spring, 1975

Child Interview
Spring, 1975

NAME ______________________________

GRADE ______________________________

SCHOOL ______________________________

TEACHER ______________________________

SEX ______________________________

Units worked on this year (into from observer before interview) ______________________________

____

1. "Has your teacher told you why you've been asked to talk to me?"

   _____ Yes, you want to talk about USMES.

   "Yes, that's right, but I'd also like to ask you about other things you do in school."

OR

2. _____ No, I don't know why I'm here.

   "Well, I'd like to know what children like or don't like about school."

---

3. "What did you do in school this year that you enjoyed?"

_____a. Science	_____g. Gym
_____b. Math	_____h. Lunch
_____c. USMES	_____i. Nothing
_____d. Solve problems	_____j. No USMES activity mentioned
_____e. Work in groups	_____k. Other
_____f. Design Lab activities	

4. If no USMES activity is mentioned in response to question #3, ask if they did each USMES related activity, if they enjoyed it and if they didn't enjoy it, why not.

Not Done	tivity	Enjoyed	Reason Not Enjoyed
_____a.	Science	_____	_____
_____b.	Math	_____	_____
_____c.	USMES	_____	_____
_____d.	Solve problems	_____	_____
_____e.	Work in groups	_____	_____
_____f.	Design Lab activities	_____	_____

Terminate the interview for any student who claims he has not done any USMES.

5. Using the USMES unit mentioned by the child, "Tell me about the ____________ unit. "Tell me what you did."

If the child responds by telling what he did, ask him, "What were the other children doing while you were doing those things?"

Based on the above information, the interviewer must make the following judgements?

_____a. The child can see where his work fits into the group effort
_____ Yes _____ No

_____b. The child knows the problem and understands he is performing a series of activities
_____ Yes _____ No

Working in Groups

6. "When you work on USMES, do you work alone or with others?"

_____a. Alone

_____b. Others

_____c. Sometimes alone, sometimes in groups

7. "When you work with others, do you always work with the same people or with lots of different people?"

_____a. Same

_____b. Different

_____c. Work just with my friends

8. Who chooses the people you work with?

_____a. Teacher

_____b. Each child

_____c. A leader

9. a. "What kinds of things did you learn in Math when you were working with ______________________________?"

b. "In science?"

c. "About solving problems?"

10. a. "Would you enjoy spending more time on things like ______________________ and less time doing other kinds of things?"

_____Yes

_____No

_____Other

b. "What would you like to spend less time on?"

c. "Did you learn anything in doing ______________(unit)______________ that you could use in class in any other way."

_____No

_____Yes (what?) ______________________________

If the child has not mentioned the problem being solved to this point, ask, "Why did you do all of this; why are you doing all this work and studying about ____?"

If you cannot get the child to state a problem, do not ask questions 11-13 but go directly to question 14.

11. "Do you really think that ________________ (problem) needed to be changed?"

_____a. Yes

_____b. No

12. "After you worked on ________________ (problem) and solved the problem, what happened to your solution?"

_____a. It was used

_____b. Nothing

_____c. Other

13. "Do you think children can solve problems like the ________________ problem, or do you think it's best if parents and teachers solve those problems?"

_____a. Children can solve them.

_____b. Adults should solve them.

_____c. Other

Design Lab

14. If a Design Lab was used, "You mentioned using a Design Lab. How often was it used?"

a. _____ Times a week

b. _____ When we are scheduled to use it (how often) ______________________

c. _____ Whenever we need to use it

d. _____ Other

15. "What was the Design Lab used for?"

_____a. USMES only

_____b. Crafts

_____c. Other

16. "Do you use it as often as you'd like?"

_____a. Yes

_____b. No

_____c. Other

How-to-Cards

17. "Have you ever seen a How-to-Card?"

_____a. Yes

_____b. No

18. If "yes," "Do you know how to use them?"

_____a. Yes (if "yes," ask him to describe how)

_____b. No

19. "How helpful are they?"

_____a. Not at all

_____b. Very

_____c. Not very

_____d. Other

APPENDIX J

Attitude Scale - Pilot Test

Form 0

0									
1	2	3	4	5	6	7	8	9	10

Teacher's Name: ______________________________

Grade: ______________________________

DIRECTIONS:

Please write your teacher's name and your grade on the lines above.

The sentences on the next pages tell how some children feel about things you do in school.

1. Read each sentence.

2. Decide which phrase tells how <u>you</u> feel. Here are the phrases:

Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
( )	( )	( )	( )	( )

3. Check the phrase that tells how <u>you</u> feel about each sentence.

Here are some examples we'll try together:

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
a. Listening to music on the radio is fun.	( )	( )	( )	( )	( )
b. Recess is boring.	( )	( )	( )	( )	( )

If you have any questions, ask them now.

Start on the next page. When you get to page 4, you will see the word STOP. Put you pencil down and wait for the instructions for Part 2.

Now turn the page. You will see the number 2 in the upper right hand corner. Begin.

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
1. I avoid arithmetic because I am not very good with numbers.	( )	( )	( )	( )	( )
2. I enjoy arithmetic.	( )	( )	( )	( )	( )
3. Arithmetic is not useful to children.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
4. Studying science bores me.	( )	( )	( )	( )	( )
5. Science is necessary to improve the world.	( )	( )	( )	( )	( )
6. What you learn in science is often the basis of a good hobby.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
7. I like to read history books.	( )	( )	( )	( )	( )
8. Social studies is boring.	( )	( )	( )	( )	( )
9. I like to have meetings in school to discuss school problems.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
10. I like to chosse what I want to learn.	( )	( )	( )	( )	( )
11. I don't like to talk to the whole class about my ideas.	( )	( )	( )	( )	( )
12. Talking with other students in small groups is fun.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
13. I like to help other students learn.	( )	( )	( )	( )	( )
14. I don't like to listen to other students talk to the class.	( )	( )	( )	( )	( )
15. I like my teachers to tell me what I'm supposed to learn.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
16. I enjoy doing school work by myself.	( )	( )	( )	( )	( )
17. I like to decide for myself what I study in school.	( )	( )	( )	( )	( )
18. I have a lot of questions I never get a chance to ask.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
19. When I hear the word arithmetic, I have an unpleasant feeling.	( )	( )	( )	( )	( )
20. Arithmetic is as important as any other subject.	( )	( )	( )	( )	( )
21. I won't need arithmetic when I grow up.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
22. I wish I had more science in school.	( )	( )	( )	( )	( )
23. I am interested in learning more about science.	( )	( )	( )	( )	( )
24. Science has little value for children.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
25. I don't like to discuss current events.	( )	( )	( )	( )	( )
26. It is important to learn about the people of other countries.	( )	( )	( )	( )	( )
27. Everyone should know how government works.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
28. I prefer to choose the people I want to work with in class.	( )	( )	( )	( )	( )
29. I like to talk in a small group about my ideas.	( )	( )	( )	( )	( )
30. I like to listen to the teacher talk to the whole class.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
31. I enjoy talking with other people in large groups.	( )	( )	( )	( )	( )
32. I like to learn by working with other students.	( )	( )	( )	( )	( )
33. I like to go on to new topics, even if I haven't learned much about the topics I studied before.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
34. I get confused when I don't know why I'm studying some things.	( )	( )	( )	( )	( )
35. I like to work at my own speed no matter what others are doing.	( )	( )	( )	( )	( )
36. I hate to make a mistake in class.	( )	( )	( )	( )	( )

STOP

Wait for more directions.

PART 2

DIRECTIONS:

Let's read this story.

Some engineers want to run a heavy television cable through a pipe. The pipe is 500 feet long and 6 inches in diameter. The pipe is about 10 feet below the ground. It is open at both ends so the engineers can work on it. The pipe is not straight, but is made up of sections that twist and curve.

The engineers have already tried to push the television cable through the pipe from either end. Each time they try, the cable twists and gets stuck after only a few feet.

The problem is to think of ways to run this television cable through the pipe, without ripping up the pipe or digging down to it.

Some other children have also read this story about the engineers. Then they wrote sentences about the story. Now we want to know whether you agree with their ideas.

The directions are the same as for the first part.

1. Read each sentence.
2. Decide which phrase tells how <u>you</u> feel.
3. Check the phrase that tells how <u>you</u> feel.

Now turn this page. Begin.

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
37. There is probably only one answer to a problem like this one.	( )	( )	( )	( )	( )
38. If someone gets an idea that no one else has thought of, he should keep it to himself.	( )	( )	( )	( )	( )
39. In a problem like this one, the best answer will be the one that most of the class decides is right.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
40. An idea for solving a problem could lead to a wrong answer but still be a good idea.	( )	( )	( )	( )	( )
41. The best answer is the one that the teacher thinks is right.	( )	( )	( )	( )	( )
42. It is best to make sure that an idea is a good one before sharing it with the class.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
43. There is probably only one best way to solve a problem like this one.	( )	( )	( )	( )	( )
44. I would like to work on a problem like this one.	( )	( )	( )	( )	( )
45. If I worked on this problem, I would get a lot of good ideas.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
46. I don't think I should ask too many questions about problems in class.	( )	( )	( )	( )	( )
47. Other students know more about problems like this than I do.	( )	( )	( )	( )	( )
48. If I already have one good idea, I would rather stick with it than look for more ideas.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
49. I would like to work on a problem like this, even though I might not be able to solve it.	( )	( )	( )	( )	( )
50. I think that my ideas for solving this problem would be better than ideas given by other students.	( )	( )	( )	( )	( )
51. I am very good at thinking and solving problems.	( )	( )	( )	( )	( )

APPENDIX K

Attitude Scale - Pre-Test

Form 1

1									
1	2	3	4	5	6	7	8	9	10

Teacher's Name: ______________________

Grade: ______________________

DIRECTIONS:

Please write your teacher's name and your grade on the lines above.

The sentences on the next pages tell how some children feel about things you do in school.

1. Read each sentence.
2. Decide which phrase tells how you feel. Here are the phrases:

Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
( )	( )	( )	( )	( )

3. Check the phrase that tells how you feel about each sentence.

Here are some examples we'll try together:

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
a. Listening to music on the radio is fun.	( )	( )	( )	( )	( )
b. Recess is boring.	( )	( )	( )	( )	( )

If you have any questions, ask them now.

Start on the next page. When you get to page 3, you will see the word STOP. Put you pencil down and wait for the instructions for Part 2.

Now turn the page. You will see the number 2 in the upper right hand corner. Begin.

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
1. I avoid arithmetic because I am not very good with numbers.	( )	( )	( )	( )	( )
2. I enjoy arithmetic.	( )	( )	( )	( )	( )
3. Arithmetic is not useful to children.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
4. Studying science bores me.	( )	( )	( )	( )	( )
5. Science is necessary to improve the world.	( )	( )	( )	( )	( )
6. What you learn in science is often the basis of a good hobby.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
7. I like to read history books.	( )	( )	( )	( )	( )
8. Social studies is boring.	( )	( )	( )	( )	( )
9. I like to have meetings in school to discuss school problems.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
10. I like to chosse what I want to learn.	( )	( )	( )	( )	( )
11. I don't like to talk to the whole class about my ideas.	( )	( )	( )	( )	( )
12. Talking with other students in small groups is fun.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
13. I like to help other students learn.	( )	( )	( )	( )	( )
14. I don't like to listen to other students talk to the class.	( )	( )	( )	( )	( )
15. I like my teachers to tell me what I'm supposed to learn.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
16. I enjoy doing school work by myself.	( )	( )	( )	( )	( )
17. I like to decide for myself what I study in school.	( )	( )	( )	( )	( )
18. I have a lot of questions I never get a chance to ask.	( )	( )	( )	( )	( )

STOP

Wait for more directions.

PART 2

DIRECTIONS:

Let's read this story.

Some engineers want to run a heavy television cable through a pipe. The pipe is 500 feet long and 6 inches in diameter. The pipe is about 10 feet below the ground. It is open at both ends so the engineers can work on it. The pipe is not straight, but is made up of sections that twist and curve.

The engineers have already tried to push the television cable through the pipe from either end. Each time they try, the cable twists and gets stuck after only a few feet.

The problem is to think of ways to run this television cable through the pipe, without ripping up the pipe or digging down to it.

Some other children have also read this story about the engineers. Then they wrote sentences about the story. Now we want to know whether you agree with their ideas.

The directions are the same as for the first part.

1. Read each sentence.
2. Decide which phrase tells how <u>you</u> feel.
3. Check the phrase that tells how <u>you</u> feel.

Now turn this page. Begin.

	Agree A L[ot]		Not Sure	Disagree A Little	Disagree A Lot
19. If someone gets an idea that no one else has thought of, he should keep it to himself.	( )	( )	( )	( )	( )
20. In a problem like this one, the best answer will be the one that most of the class decides is right.	( )	( )	( )	( )	( )
21. An idea for solving a problem could lead to a wrong answer but still be a good idea.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
22. There is probably only one best way to solve a problem like this one.	( )	( )	( )	( )	( )
23. Other students know more about problems like this than I do.	( )	( )	( )	( )	( )
24. I am very good at thinking and solving problems.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
25. If I already have one good idea, I would rather stick with it than look for more ideas.	( )	( )	( )	( )	( )
26. If I worked on this problem, I would get a lot of good ideas.	( )	( )	( )	( )	( )

APPENDIX L

Attitude Scale - Pre-Test

Form 2

2									
1	2	3	4	5	6	7	8	9	10

Teacher's Name: ______________________________

Grade: ______________________________

DIRECTIONS:

Please write your teacher's name and your grade on the lines above.

The sentences on the next pages tell how some children feel about things you do in school.

1. Read each sentence.

2. Decide which phrase tells how you feel. Here are the phrases:

Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
( )	( )	( )	( )	( )

3. Check the phrase that tells how you feel about each sentence.

Here are some examples we'll try together:

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
a. Listening to music on the radio is fun.	( )	( )	( )	( )	( )
b. Recess is boring.	( )	( )	( )	( )	( )

If you have any questions, ask them now.

Start on the next page. When you get to page 3, you will see the word STOP. Put you pencil down and wait for the instructions for Part 2.

Now turn the page. You will see the number 2 in the upper right hand corner. Begin.

		Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
1.	When I hear the word arithmetic, I have an unpleasant feeling.	( )	( )	( )	( )	( )
2.	Arithmetic is as important as any other subject.	( )	( )	( )	( )	( )
3.	I won't need arithmetic when I grow up.	( )	( )	( )	( )	( )

		Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
4.	I wish I had more science in school.	( )	( )	( )	( )	( )
5.	I am interested in learning more [illegible] science.	( )	( )	( )	( )	( )
6.	Science has little va[illegible] child[illegible].	( )	( )	( )	( )	( )

		Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
7.	I don't like to discuss current events.	( )	( )	( )	( )	( )
8.	It is important to learn about the people of other countries.	( )	( )	( )	( )	( )
9.	Everyone should know how government works.	( )	( )	( )	( )	( )

		Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
10.	I prefer to choose the people I want to work with in class.	( )	( )	( )	( )	( )
11.	I like to talk in a small group about my ideas.	( )	( )	( )	( )	( )
12.	I like to listen to the teacher talk to the whole class.	( )	( )	( )	( )	( )

		Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
13.	I enjoy talking with other people in large groups.	( )	( )	( )	( )	( )
14.	I like to learn by working with other students.	( )	( )	( )	( )	( )
15.	I like to go on to new topics, even if I haven't learned much about the topics I studied before.	( )	( )	( )	( )	( )

		Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
16.	I get confused when I don't know why I'm studying some things.	( )	( )	( )	( )	( )
17.	I like to work at my own speed no matter what others are doing.	( )	( )	( )	( )	( )
18.	I hate to make a mistake in class.	( )	( )	( )	( )	( )

STOP

Wait for more directions.

PART 2

DIRECTIONS:

Let's read this story.

Some engineers want to run a heavy television cable through a pipe. The pipe is 500 feet long and 6 inches in diameter. The pipe is about 10 feet below the ground. It is open at both ends so the engineers can work on it. The pipe is not straight, but is made up of sections that twist and curve.

The engineers have already tried to push the television cable through the pipe from either end. Each time they try, the cable twists and gets stuck after only a few feet.

The problem is to think of ways to run this television cable through the pipe, without ripping up the pipe or digging down to it.

Some other children have also read this story about the engineers. Then they wrote sentences about the story. Now we want to know whether you agree with their ideas.

The directions are the same as for the first part.

1. Read each sentence.
2. Decide which phrase tells how <u>you</u> feel.
3. Check the phrase that tells how <u>you</u> feel.

Now turn this page. Begin.

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
19. It is best to make sure that an idea is a good one before sharing it with the class.	( )	( )	( )	( )	( )
20. The best answer is the one that the teacher thinks is right.	( )	( )	( )	( )	( )
21. There is probably only one answer to a problem like this one.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
22. I would like to work on a problem like this, even though I might not be able to solve it.	( )	( )	( )	( )	( )
23. I think that my ideas for solving this problem would be better than ideas given by other students.	( )	( )	( )	( )	( )
24. I don't think I should ask too many questions about problems in class.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
25. I would like to work on a problem like this one.	( )	( )	( )	( )	( )

APPENDIX M

Attitude Scale - Post Test

Form 3

3									
1	2	3	4	5	6	7	8	9	10

Teacher's Name: ______________________________

Grade: ______________________________

DIRECTIONS:

Please write your teacher's name and your grade on the lines above.

The sentences on the next pages tell how some children feel about things you do in school.

1. Read each sentence.

2. Decide which phrase tells how you feel. Here are the phrases:

Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
( )	( )	( )	( )	( )

3. Check the phrase that tells how you feel about each sentence.

Here are some examples we'll try together:

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
a. Listening to music on the radio is fun.	( )	( )	( )	( )	( )
b. Recess is boring.	( )	( )	( )	( )	( )

If you have any questions, ask them now.

Start on the next page. When you get to page 4, you will see the word STOP. Put you pencil down and wait for the instructions for Part 2.

Now turn the page. You will see the number 2 in the upper right hand corner. Begin.

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
1. I try not to do much arithmetic because I am not good with numbers.	( )	( )	( )	( )	( )
2. I enjoy arithmetic.	( )	( )	( )	( )	( )
3. Arithmetic is not useful to children.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
4. When I hear the word arithmetic, I have an unpleasant feeling.	( )	( )	( )	( )	( )
5. Arithmetic is as important as any other subject.	( )	( )	( )	( )	( )
6. I won't need arithmetic when I grow up.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
7. I wish I had more science in school.	( )	( )	( )	( )	( )
8. I am interested in learning more about science.	( )	( )	( )	( )	( )
9. Science is not useful for children.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
10. Studying science bores me.	( )	( )	( )	( )	( )
11. Science helps to improve the world.	( )	( )	( )	( )	( )
12. What you learn in science is often the basis of a good hobby.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
13. I like to choose what I want to learn.	( )	( )	( )	( )	( )
14. I don't like to talk to the whole class about my ideas.	( )	( )	( )	( )	( )
15. Talking with other students in small groups is fun.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
16. I like to help other students learn.	( )	( )	( )	( )	( )
17. I don't like to listen to other students talk to the class.	( )	( )	( )	( )	( )
18. I like my teachers to tell me what I'm supposed to learn.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
19. I like to study lots of things, even if I don't learn them well.	( )	( )	( )	( )	( )
20. I enjoy doing school work by myself.	( )	( )	( )	( )	( )
21. I like to decide for myself what I study in school.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
22. I have a lot of questions I never get a chance to ask.	( )	( )	( )	( )	( )
23. I prefer to choose the people I want to work with in class.	( )	( )	( )	( )	( )
24. I like to talk in a small group about my ideas.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
25. I like to listen to the teacher talk to the whole class.	( )	( )	( )	( )	( )
26. I enjoy talking with other people in large groups.	( )	( )	( )	( )	( )
27. I like to learn by working with other students.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
28. I like to go on to new topics, even if I haven't learned much about the topics I studied before.	( )	( )	( )	( )	( )
29. I get confused when I don't know why I'm studying some things.	( )	( )	( )	( )	( )
30. I like to work at my own speed no matter what others are doing.	( )	( )	( )	( )	( )
31. I hate to make a mistake in class.	( )	( )	( )	( )	( )

STOP

Wait for more directions.

PART 2

DIRECTIONS:

Let's read this story.

Some engineers want to run a heavy television cable through a pipe. The pipe is 500 feet-long and 6 inches in diameter. The pipe is about 10 feet below the ground. It is open at both ends so the engineers can work on it. The pipe is not straight, but is made up of sections that twist and curve.

The engineers have already tried to push the television cable through the pipe from either end. Each time they try, the cable twists and gets stuck after only a few feet.

The problem is to think of ways to run this television cable through the pipe, without ripping up the pipe or digging down to it.

Some other children have also read this story about the engineers. Then they wrote sentences about the story. Now we want to know whether you agree with their ideas.

The directions are the same as for the first part.

1. Read each sentence.
2. Decide which phrase tells how <u>you</u> feel.
3. Check the phrase that tells how <u>you</u> feel.

Now turn this page. Begin.

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
1. There is probably only one answer to a problem like this one.	( )	( )	( )	( )	( )
2. If someone gets an idea that no one else has thought of, he should keep it to himself.	( )	( )	( )	( )	( )
3. In a problem like this one, the best answer will be the one that most of the class decides is right.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
4. An idea for solving a problem could lead to a wrong answer but still be a good idea.	( )	( )	( )	( )	( )
5. The best answer is the one that the teacher thinks is right.	( )	( )	( )	( )	( )
6. It is best to make sure that an idea is a good one before sharing it with the class.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
7. There is probably only one best way to solve a problem like this one.	( )	( )	( )	( )	( )
8. I would like to work on a problem like this one.	( )	( )	( )	( )	( )
9. If I worked on this problem, I would get a lot of good ideas.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
10. I don't think I should ask too many questions about problems in class.	( )	( )	( )	( )	( )
11. Other students know more about problems like this than I do.	( )	( )	( )	( )	( )
12. If I already have one good idea, I would rather stick with it than look for more ideas.	( )	( )	( )	( )	( )

	Agree A Lot	Agree A Little	Not Sure	Disagree A Little	Disagree A Lot
13. I would like to work on a problem like this, even though I might not be able to solve it.	( )	( )	( )	( )	( )
14. I think that my ideas for solving this problem would be better than ideas given by other students.	( )	( )	( )	( )	( )
15. I am very good at thinking and solving problems.	( )	( )	( )	( )	( )