DOCUMENT RESUME

ED 200 627 TM 810 212

AUTHOR Blust, Ross S.; Kohr, Richard L.
TITLE An Examination of the Relationship Between School Scores Derived From Commercial Achievement Tests and Those From Statewide Assessment.
INSTITUTION Pennsylvania State Dept. of Education, Harrisburg.
PUB DATE Apr 81
NOTE 18p.: Paper presented at the Annual Meeting of the National Council on Measurement in Education (Los Angeles, CA, April 11-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Basic Skills; *Comparative Testing; *Correlation; Elementary Education; Elementary School Mathematics; *Grade 5; *Grade Equivalent Scores; Reading Skills; Standardized Tests; State Programs; Test Interpretation; Writing Skills
IDENTIFIERS *Pennsylvania Educational Quality Assessment; *Percentile Ranks

ABSTRACT

An apparent discrepancy between building level scores in basic skills produced by Pennsylvania's state assessment program (EQA) and building summary scores, generally a grade equivalent, provided by commercial standardized achievement tests is investigated. The impetus for the study came from occasional reports by school administrators that their school level grade equivalent suggested above average performance while the state assessment percentile rank seemed low by comparison. Achievement test scores at the building level, obtained from approximately 302 school districts, were merged with state assessment scores in reading, writing skills and mathematics for grades 5, 8, and 11. Correlations in the .70 to .85 range revealed a similar rank ordering of schools by commercial and state assessment tests. Rather than average building performance being equivalent to the grade-month of testing, as believed by many administrators, the distribution of building mean scores for several commercial tests indicated that "average" corresponded to a grade equivalent of six months to a year beyond that point. Commercial achievement test scores examined included those from the California Achievement Tests, Comprehensive Tests of Basic Skills, Iowa Tests of Basic Skills, Metropolitan Achievement Tests, Science Research Associate Assessment Survey and the Stanford Achievement Test. (Author/RL)

ED200627

AN EXAMINATION OF THE RELATIONSHIP BETWEEN SCHOOL SCORES DERIVED FROM COMMERCIAL ACHIEVEMENT TESTS AND THOSE FROM STATEWIDE ASSESSMENT

Ross S. Blust
and
Richard L. Kohr

Pennsylvania Department of Education

Presented at the Annual Meeting of the
National Council on Measurement in Education
Los Angeles, California
April 1981

TM 810 212

Title: An Examination of the Relationship Between School Scores Derived from Commercial Achievement Tests and Those From Statewide Assessment

Authors: Ross S. Blust and Richard L. Kohr

Institution: Pennsylvania Department of Education

ABSTRACT

This study investigated an apparent discrepancy between building level scores in basic skills produced by Pennsylvania's state assessment program and building summary scores, generally a grade equivalent, provided by commercial standardized achievement tests. The impetus for the study came from occasional reports by school administrators that their school level grade equivalent suggested above average performance while the state assessment percentile rank seemed low by comparison. Achievement test scores at the building level, obtained from approximately 302 school districts, were merged with state assessment scores in reading, writing skills and mathematics. Correlations in the .70 to .85 range revealed a similar rank ordering of schools by commercial and state assessment tests. Rather than average building performance being equivalent to the grade-month of testing, as believed by many administrators, the distribution of building mean scores for several commercial tests indicated that "average" corresponded to a grade equivalent of six months to a year beyond that point.

An Examination of the Relationship Between School Scores Derived From Commercial Achievement Tests and Those From Statewide Assessment

INTRODUCTION

This study was conducted to investigate the perceived discrepancy in building level basic skill scores provided by the Pennsylvania Educational Quality Assessment Program (EQA) and the building summary scores, generally a grade equivalent, obtained from commercial standardized achievement tests. When comparing the building level results from these two data sources, school administrators sometimes noticed that their standardized test results appeared to be higher than the state assessment results. This discrepancy presented public relations problems when attempting to report seemingly conflicting test results. An informal review of the school districts making this complaint suggested that the problem might be confined to just two or three of the major standardized tests. In such instances, the general senario was one in which the school building grade equivalent suggested above average performance whereas the EQA state percentile seemed low by comparison. Expecting higher EQA performance seemed to originate from the assumption by school administrators that a grade equivalent commensurate with the grade-month at which testing took place was indicative of "average" performance (with respect to a national norm). Furthermore, a grade equivalent higher than the grade-month of testing was interpreted as above average performance.

The Pennsylvania Educational Quality Assessment (EQA) program utilizes state norms developed each year of the assessment on Pennsylvania building mean scores. National norms produced by commercial achievement test companies on a base norm year are calculated on student level scores. Thus, part of the difference in performance is due to differences in aggregation level and norm samples. Also, commercial achievement tests utilize a base year to norm while EQA has renormed each year.

During the construction of the EQA tests, student level correlations in the .60 to .80 range were found between EQA subject tests and similar tests from several prominent standardized tests. From these results it was anticipated that EQA building level scores in reading, writing and mathematics would show reasonably close agreement with reading, language, and mathematics scores obtained from the commercial achievement tests. This expectation is based on the fact that when aggregated data such as a building mean score is the unit of analysis, a larger magnitude in the correlation coefficient is generally observed (Robinson, 1950).

Instruments

The present EQA test package, developed by Pennsylvania Department of Education staff, with the assistance of advisory committees, has been in use since 1978. The assessment package provides school districts with information at the school building level in both cognitive and affective areas at grades five, eight and eleven. Reading, writing, and mathematics were the areas of primary concern in the present study. All EQA tests are multiple choice format. The grade five reading test, consisting of 48 items, is comprised of predominantly inferential comprehension items. The writing skills test includes 45 items measuring mechanics and usage, sentence sense, paragraph sense and style, tone and flavor. The 60 item mathematics test, measuring conceptual, computational, and problem solving levels, contains items dealing with number systems, numeration, notation, geometry, measurement, number pattern, relationships and other topics. The reliability and validity of these instruments is documented in the manual, Getting Inside The EQA Inventory (Kohr, Hertzog and Seiverling, 1980).

Commercial achievement tests included were the following: California Achievement Tests, Comprehensive Tests of Basic Skills (CTBS), Iowa Tests of Basic Skills, Metropolitan Achievement Tests, SRA (Science Research Associate) Assessment Survey, and the Stanford Achievement Test. Reading, language, and mathematics scores were obtained for the achievement tests previously listed.

It should be noted that there were differences in the subscales combined to produce total reading, total mathematics, or language scores. For example, the California Achievement Test includes in the language score both mechanics and language expression. Language mechanics contains capitalization and punctuation items while language expression contains usage, sentence structure and paragraph organization (CTB/McGraw-Hill, 1978). The Metropolitan Achievement Test includes items measuring listening comprehension, grammar and syntax, spelling, study skills, punctuation and capitalization, and usage as a part of the language score (Psychological Corporation, 1979). Based on the content description offered by the publishers, there are differences in the content included for areas such as language. Hence, there are not only differences in the items from test to test but differences in the content covered.

METHOD

Over 380 school districts who participated in the state assessment during the three year period of 1978 to 1980 were requested to submit photocopies of the building level standardized achievement test results for grades five, eight, and eleven. The request was for test results only for the year that the school district participated in the EQA program. Replies were received from 302 school districts by the end of July 1980. Some returns were unusable because achievement tests were not administered in the required grades. Data representing six test companies were included among the usable returns.

Each school's summary statistics, including the mean raw score, grade equivalent, percentile rank, stanine, and standard score were merged with EQA test scores and condition variables. The EQA program gathers data on thirty-five condition variables which reflect the operational conditions of the school including socioeconomic indicators. These data enabled analyses to be conducted, comparing the samples of schools using different commercial achievement tests, with respect to socioeconomic and other condition variables.

Achievement test companies were assigned a code letter so that individual companies could not be identified in the study. For grades eight and eleven, the sample gathered for each achievement test was rather small. This was due mainly to a decrease in the use of achievement tests at higher grade levels. Therefore, the only data analyzed and reported were for the fifth grade.

RESULTS

Correlations with Grade Equivalents

Correlations were calculated to evaluate the nature and magnitude of the relationship between EQA reading, writing, and mathematics scores and the appropriate scores derived from each of the standardized tests. Initially, the intent was to use mean number of items correct or a standard score from the standardized achievement test results as a metric to correlate with the EQA building mean raw scores. Upon examination of the data, it was found that, because school districts can opt to receive the scores they wish, an insufficient number of schools requested the desired scores. All schools did receive building level grade equivalent scores. Thus, the grade equivalent was chosen as the metric for analysis. The correlations presented are between EQA building mean raw scores and achievement test grade equivalents. Correlations were calculated for two groups, those schools that administered commercial achievement tests in the fall and those testing in the spring.

Table 1 presents the correlations for grade five reading scores. These correlations reveal a strong positive relationship between EQA mean scores and achievement test grade equivalents for reading. The highest correlations were found in the case of schools administering the achievement test in the spring. This is to be expected since the EQA tests are also given in the spring of each year. As noted in Table 1, the sample size is extremely low for several tests under the fall testing column. In these instances, the correlations were omitted (indicated by NA). In general, the correlations indicate that schools scoring high on the EQA measure of reading also tended to have high grade equivalent scores in reading on commercially produced achievement tests. Conversely, schools with low EQA scores tend to have low grade equivalent scores.

Table 1.

Building Level Correlations Between Grade Five EQA Reading Mean Raw Scores and Achievement Test Mean Reading Grade Equivalent Scores

Standardized Test	Fall Testing	Spring Testing
A	.86 n = 16	.67 n = 32
B	NA n = 5	.83 n = 35
C	.10 n = 29	.81 n = 86
D	NA n = 7	.59 n = 61
E	.75 n = 33	.86 n = 69
F	.73 n = 51	.85 n = 211

As revealed by the correlations found in Table 2, there was a strong positive statistical relationship between EQA writing scores and grade equivalent scores in writing on commercial achievement tests. Achievement test companies A and B were used by only a few schools thus the correlations should be viewed with caution. For both reading and writing correlations the relationship between EQA scores and achievement tests was rather strong as expected due to the similarity of content of the EQA and commercial achievement tests. It was noted that correlations for the larger samples of over 50 schools ranged from .71 for achievement test D for spring testing to .87 for achievement test E for spring testing.

Correlations between the EQA mean scores and achievement test grade equivalents in mathematics were placed in Table 3. For mathematics there was a strong positive relationship between all of the achievement tests and EQA. Correlations for the larger samples of over 50 schools ranged from .75 on achievement test F for fall testing to .85 for achievement test F spring testing.

In summary, the statistical relationship found between EQA mean scores and achievement test grade equivalents was rather strong and always positive for reading, writing and mathematics. Correlations between EQA scores and mean school scores administering achievement tests in the spring were almost always the highest. Schools administering achievement tests in the fall were anticipated to have lower mean grade equivalent scores than those schools administering achievement tests in the spring. Achievement tests B, C and D at times had slightly lower correlations with EQA than other achievement tests considered. In general the correlations indicated high agreement between achievement test scores and EQA scores.

Table 2

Building Level Correlations Between Grade Five EQA Mean Writing Raw Scores and Achievement Test Mean Writing Grade Equivalent Scores

Standardized Test	Fall Testing	Spring Testing
A	.70 n = 16	.83 n = 28
B	NA n = 5	.64 n = 25
C	.71 n = 29	.81 n = 86
D	NA n = 7	.71 n = 58
E	.79 n = 33	.87 n = 69
F	.75 n = 51	.85 n = 211

8

Table 3

Building Level Correlations Between Grade Five
EQA Mean Mathematics Raw Scores and Achievement Test
Mean Mathematics Grade Equivalent Scores

Standardized Test	Fall Testing	Spring Testing
A	.84 n = 16	.87 n = 28
B	NA n = 5	.89 n = 35
C	.51 n = 29	.76 n = 86
D	NA n = 7	.79 n = 65
E	.80 n = 33	.84 n = 69
F	.75 n = 131	.85 n = 211

Mean Grade Equivalents

Since the correlational analyses suggested a similar rank ordering of schools by EQA and commercial achievement tests, an examination of mean grade equivalent scores was undertaken. It was suspected that the interpretative dilemma experienced by school administrators lay in the assumption that "average performance" was "expected" to be a building mean grade equivalent that was equal to the grade-month of testing. One could anticipate variation among schools using different testing programs on the basis of Williams' (1980) report, documenting the inconsistency among tests in the grade equivalent corresponding to different percentiles on student norms.

The reading grade equivalent means for achievement tests, found in Table 4, revealed that achievement tests A, B, D, and F had means that were higher than "expected." An analysis of the means for schools testing in the spring found that test F had the highest mean which was over one year higher than anticipated. Means were not included in Table 4 for tests B and D where sample size was extremely small. Also, the means presented for the rather small samples should be viewed with caution.

Grade five writing means were placed in Table 5 for each achievement test. Writing achievement tests A, B, D, and F had grade equivalent means higher than expected. None of the means for spring testing was below 6.2 for writing. The means for three achievement tests were above 7.0 for spring testing which is over one year higher than expected.

Table 4

Building Level Mean Grade Equivalent Scores for Reading at Grade Five

Standardized Test	Fall Testing	Spring Testing
A	5.78 n = 16	6.58 n = 32
B	NA n = 5	6.64 n = 35
C	5.29 n = 29	5.97 n = 86
D	NA n = 7	6.66 n = 61
E	5.17 n = 33	6.18 n = 69
F	5.93 n = 51	7.09 n = 21

Table 5

Building Level Mean Grade Equivalent
Scores for Writing at Grade Five

Standardized Test	Fall Testing	Spring Testing
A	6.09 n = 16	7.13 n = 28
B	NA n = 5	7.75 n = 25
C	5.65 n = 29	6.21 n = 86
D	NA n = 7	6.94 n = 58
E	5.20 n = 33	6.53 n = 67
F	5.66 n = 44	7.03 n = 209

Summarized in Table 6 are the mathematics mean grade equivalents for each achievement test. Mathematics had lower mean grade equivalents than reading and writing; however, the means for achievement tests A, B, D, and F were higher than expected.

Overall, the grade equivalent means indicated schools were scoring much higher than anticipated for all three areas: reading, writing, and mathematics. The grade equivalent means revealed large differences between achievement test means with tests A, B, D, and F providing higher scores than tests C and E. Mean grade equivalents for tests C and E on reading and mathematics came closest to the grade equivalents anticipated.

Table 6

Building Level Mean Grade Equivalent Scores for Mathematics at Grade Five

Standardized Test	Fall Testing	Spring Testing
A	5.48 n = 16	6.60 n = 28
B	NA n = 5	6.62 n = 35
C	5.28 n = 29	5.99 n = 86
D	NA n = 7	6.44 n = 65
E	5.04 n = 33	6.33 n = 69
F	5.61 n = 51	6.71 n = 211

The results helped to explain why several school district employees indicated there was a discrepancy between EQA and achievement test scores. In order to be above the state mean a grade five mean school score on an achievement test for spring testing often had to be over a grade equivalent of 6.6 and in several cases even higher. From an analysis of 211 elementary schools using test F (the most extreme case) an average grade equivalent for spring testing of 7.09 was found for total reading and 6.71 for total math. It is instructive to examine the distribution of grade equivalent scores for the 262 schools using test F. Thirty-seven percent of the reading scores were 7.0 or higher. Only seventeen percent of the schools scored lower than 6.0 Table 7 summarizes the building level grade equivalent scores associated with percentile ranks based on the sample of schools providing scores. Thus, a school with a mean grade equivalent of 6.4 in reading was found to have a percentile rank in the sample of schools using test F that was equivalent to approximately the 30th percentile. Many school district employees perceived mean achievement test scores as being high when the grade equivalents were above the year and month of testing. This view appears to be incorrect since much higher grade equivalent scores are required to be above the state mean on several achievement tests.

Table 7

Percentile Rank in Reading and Mathematics
Grade Equivalents for Schools Using Achievement Test F

Percentile Rank	Reading Grade Equivalent	Mathematics Grade Equivalent
90	8.0	7.6
70	7.4	6.9
50	6.9	6.5
30	6.3	6.0
10	5.7	5.4

n = 262

With the results of this study EQA staff members and school district employees should be better able to interpret both EQA scores and achievement test scores especially grade equivalents from the achievement tests. Also, it should be noted that grade equivalents are not endorsed by the authors as being the most useful statistic that could be employed to present achievement test results. Grade equivalents are not consistent from achievement test to achievement test and from subject to subject. For example, a student scoring 2.0 years above grade level in reading may not receive the same percentile rank in reading and mathematics although scoring 2.0 years above grade level in math. Differences in achievement test grade equivalents were noted in Tables 4, 5, and 6 exemplifying the inconsistencies from test to test. Also, employees of state assessment programs outside of Pennsylvania may find this study helpful when interpreting state assessment results.

Comparability of the Sample to the State

Since data were collected from a total of 302 school districts representing several achievement test companies, an examination of the sample utilized in the study was performed. The sample might be questioned as to whether it is representative of the population of schools undergoing assessment during the 1978 to 1980 period. To facilitate this comparison, mean scores on a set of variables depicting EQA basic skills achievement, socioeconomic background, demographic characteristics and indices of teaching staff perceptions are presented in Table 8. To determine whether the mean scores for the 660 schools providing standardized test data differed significantly from the 1590 schools assessed from 1978 to 1980, a z test was employed (Hays, 1963, p. 250). All tests were nonsignificant with the exception of parental education ($z = 4.58$, $p < .01$), parental occupation ($z = 2.99$, $p < .01$), population density ($z = 2.82$, $p < .01$), family size ($z = 2.60$, $p < .01$) and amount of reading material in the home ($z = 2.91$, $p < .01$). To assist in evaluating these significant differences, given the large sample, a measure of "effect size" (Cohen, 1969) was calculated by forming a ratio of the observed difference to the population standard deviation. In each case, the effect size was quite small, only .07 except for parental education which reached .11. All can be safely disregarded.

The examination of the comparability of the schools comprising each sample was continued by calculating percentiles and EQA variables. The characteristics of the schools providing data on the different standardized tests were analyzed by converting mean scores and EQA variables to percentile ranks. Percentiles for the EQA variables were placed on Table 9 by achievement tests published. The samples gathered for the study when analyzed by test had most EQA scores at or close to the state median (50th percentile). This indicated the samples were based on schools that, when averaged together, were representative of the state. Achievement test B had the lowest percentiles in the EQA basic skill areas of reading, writing, and mathematics and was the test that differed most from the state median. A review of achievement test B mean grade equivalents found in Tables 4, 5, and 6 revealed grade equivalents higher than anticipated. Thus, the schools using achievement test B have the lowest EQA mean scores, but the mean grade equivalents were among the highest. Possibly in part, the unexpected findings for test B could be explained by the rather small sample of only forty schools creating inconsistent results. Achievement test A had the highest percentiles in the EQA basic skill areas and on the other EQA variables. For test A the mean grade equivalents found in Tables 4, 5, and 6 were rather high which may be explained by the median achievement levels in EQA scores. But, the mean grade equivalents for test A were rather high even for schools that were above the state median at the 55th percentile or 60th percentile. In general, the schools sampled were on EQA mean scores typical of the state. This supported the assumption that grade equivalents were for the most achievement tests producing higher than anticipated scores.

14

Table 8

Comparability of Schools Supplying Standardized Test Data With All Schools Undergoing Assessment During 1978-1980

Variable	Schools Supplying Test Data (n=660)		All Schools Assessed (n=1590)	
	Mean	SD	Mean	SD
Reading	27.62	3.94	27.58	3.75
Writing	29.08	3.24	28.92	3.11
Mathematics	37.45	4.25	37.20	4.03
Grade enrollment	62.66	40.22	63.61	42.46
Percentage of low income students	20.39	14.81	19.95	13.38
Parental education	3.69	0.65	3.62	0.61
Parental occupation	56.41	12.85	55.49	12.28
Population density	2.35	1.83	2.22	1.84
Percentage of white students	92.44	13.95	92.88	12.61
Family size	2.36	0.44	2.39	0.46
Amount of reading materials in home	10.96	1.00	10.89	0.96
Factors disruptive to classroom management	20.04	1.49	19.99	1.47
Discipline problems	10.91	1.95	10.99	1.82
Teacher influence on instructional decisions	20.13	2.58	20.05	2.61

Table 9

Percentiles on EQA Variables by Achievement Test Samples

EQA Variable	Test A	Test B	Test C	Test D	Test E	Test F
Reading	60	35	55	45	50	45
Writing	55	35	55	42	50	50
Mathematics	60	40	55	50	60	45
Percentage of low income students	55	75	50	55	60	55
Factors disruptive to classroom management	55	45	55	45	45	45
Discipline problems	50	45	50	45	40	45
Parental education	67	55	65	60	65	65
Parental occupation	60	47	60	55	55	60
Population density	55	60	65	55	60	60
Student perception of parental interest in school	55	45	45	45	45	50

Note: For Test A n = 64, Test B n = 40, Test C n = 116, Test D n = 72, Test E n = 106, Test F n = 262

16

ANCILLARY FINDINGS

Data were available on the testing dates for each achievement test, hence; a summary was produced for grade five test F the achievement test with the most schools responding. The number of schools testing was summarized in Table 10 for each month from September to June. It was interesting to find testing conducted months before and after the fall and spring norming dates. This is a questionable practice on the part of school district employees. It was noted that over 60 percent of the schools test in the spring.

Table 10

Achievement Test F Testing Dates
Grade Five

Month	Frequency	Percentage	Cumulative Percentage
September	20	7.6	7.6
October	15	5.7	13.3
November	16	6.1	19.4
December	—	—	19.4
January	21	8.0	27.4
February	—	—	27.4
March	15	5.7	33.1
April	128	49.0	82.1
May	43	16.4	98.5
June	4	1.5	100.0
Total	262		

REFERENCES

Cohen, J. Statistical Power Analysis for the Behavioral Sciences, New York: Academic Press, 1969.

CTB/McGraw-Hill. California Achievement Test: Class Management Guide, Forms C and D. Monterey, CA: CTB/McGraw-Hill, 1978.

Hays, W. L. Statistics for Psychologists, New York: Holt, Rinehart and Winston, 1963.

Kohr, R. L., Hertzog, J. F. and Seiverling, R. Getting Inside The EQA Inventory. Harrisburg, Pennsylvania Department of Education, 1980.

Psychological Corporation. Resources For Decision Making. New York: The Psychological Corporation, 1979.

Robinson, W. S. Ecological correlations and the behavior of individuals. American Sociological Review, 1950, 15, 351-357.

Williams, T. B. The distributions of NCE, percentile and grade equivalent scores among twelve nationally standardized tests. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, April 1980.