DOCUMENT RESUME

ED 284 904 TM 870 457

AUTHOR Solomon, Alan

TITLE Norm-Referenced Standardized Mathematics Achievement Tests at the Secondary School Level and Their Relationship to the National Assessment of Educational Progress Content Objectives and Subobjectives.

PUB DATE Apr 87

NOTE 50p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).

PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Content Validity; *Educational Objectives; High Schools; *Mathematical Concepts; Mathematics Achievement; *Mathematics Skills; *Mathematics Tests; *Norm Referenced Tests; Standardized Tests; Test Construction; *Test Items; Test Validity

IDENTIFIERS California Achievement Tests; Comprehensive Tests of Basic Skills; Metropolitan Achievement Tests; *National Assessment of Educational Progress; SRA Survey of Basic Skills; Stanford Achievement Tests

ABSTRACT

A panel of expert referees from the Philadelphia school district categorized items from secondary-level standardized mathematics tests according to National Assessment of Educational Progress (NAEP) subobjectives for mathematics. The following tests were covered by the study: (1) California Achievement Tests (Levels 19 and 20); (2) Comprehensive Tests of Basic Skills (Levels J and K); (3) Metropolitan Achievement Tests (Advanced 1 and 2); (4) SRA Survey of Basic Skills (Levels 36 and 37); and (5) Stanford Achievement Tests (Advanced, Task 1, and Task 2). Of these 11 tests, 7 appropriate tests were considered for each grade grouping (9, 10, 11, 12). Chi-square one-sample tests showed that these tests did not differ significantly in terms of the number of NAEP objectives or subobjectives addressed on a grade by grade basis. However, the proportion of NAEP objectives covered by the tests showed that the NAEP system is not being followed by publishers of standardized norm-referenced mathematics tests in the secondary school grades. All tests together addressed 82% of NAEP objectives and 48% of subobjectives, but no individual test covered more than 65% of NAEP objectives or 20% of subobjectives. All objectives and subobjectives addressed by these tests are reported in tables. (LPG)

NORM-REFERENCED STANDARDIZED MATHEMATICS ACHIEVEMENT TESTS AT THE SECONDARY SCHOOL LEVEL AND THEIR RELATIONSHIP TO THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS CONTENT OBJECTIVES AND SUBOBJECTIVES

Prepared by:

Alan Solomon

American Educational Research Association

Washington, D. C.

April, 1987

NORM-REFERENCED STANDARDIZED MATHEMATICS ACHIEVEMENT TESTS AT THE SECONDARY SCHOOL LEVEL AND THEIR RELATIONSHIP TO THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS CONTENT OBJECTIVES AND SUBOBJECTIVES

Introduction

Standardized test publishers claim they use different sources for the contents of their examinations. Textbooks, curriculum guides and the opinions of leading educators are usually mentioned in the manuals and handbooks which accompany the tests. One potential source is absent from the references accompanying the current versions of five leading standardized test series, the National Assessment of Educational Progress findings.

The reasons underlying the publishers' choice of sources is not an issue in this research. Clearly, the scope and the intensity of the NAEP's efforts posed against the publishers' choices form a dilemma which other researchers may elect to examine. This research will study the extent to which current norm-referenced standardized achievement tests reflect the curriculum espoused by the NAEP in one subject, secondary mathematics.

Research designed to show if standardized tests varied over time in terms of their content revealed that few changes had taken place.[1] In this research, which was limited to elementary mathematics, the list of objectives prepared by the NAEP for its initial mathematics assessment in

1972-73 served as the criterion. Further analyses showed that a relatively small percent of the NAEP objectives and their components or subobjectives were assessed by the standardized tests.

This finding must be qualified because the list of objectives and their components may not be suitable for all students in America despite educators' assertions that all of the topics may be part of an elementary school mathematics program. Secondary school mathematics programs have not been studied under the NAEP criterion and the present study was conducted in order to determine the degree to which current norm-referenced standardized achievement tests attend to the NAEP objectives and subobjectives. Reliability and validity for the procedures had been established earlier and will be described in detail later in this paper.

Five standardized test series were used in the study, the Stanford Achievement Test, the Metropolitan Achievement Tests, the California Achievement Tests, the Comprehensive Tests of Basic Skills and the SRA Achievement Series. The researcher examined each item in each test and assigned it to one of the NAEP subobjectives. Chi-square was used to analyze the data

The National Assessment of
Educational Progress

The National Assessment of Educational Progress (NAEP) is a cyclical testing program in which tests in ten instructional areas are alternated annually. Examinations in music, reading, writing, art, citizenship, mathematics, science, social studies, literature and occupational development are administered to carefully selected samples at four age levels, nine, thirteen, seventeen, and twenty-six to thirty-five. Results are presented by age, sex, geographic region, race, community type and parental education status.

> Many benefits have been derived from NAEP, not the least of which is the refinement of methodologies for implementing large-scale exercise development and data collection activities. It is hoped that the data have influenced school administrators and federal and state legislators to make rational decisions about the allocations of money for educational programs.[2]

The NAEP is a project of the Education Commission of the States. Designed to determine the nation's progress in education, this project is funded by the National Center for Education Statistics. This assignment was given to the Education Commission of the States by the U. S. Office of Education at its inception in 1867.[3]

Prior to the NAEP's work in assessment, measures of educational quality were based on categorical or demographic information. This information included teacher-student ratios, class size, number of classrooms, and per-pupil expenditures among other data. Meaningful outcome measure were not available. Formal testing programs produced information which could be used to categorize students but this data did not yield information on individual student learning.[4]

Dr. Francis Keppel, the United States Commissioner of Education from 1962 to 1965, became concerned about this dearth of information and initiated a series of conferences designed to find ways to collect data on the nation's progress in education. Keppel's approach bore fruit in 1964 when a group of educators formed the Exploratory Committee on Assessing the Progress of Education (ECAPE). Ralph W. Tyler chaired ECAPE and directed the committee toward determining the feasibility of conducting a national assessment. ECAPE reported that this project was feasible and the responsibility for conducting the assessment was given to the Education Commission of the States and named the National Assessment of Educational Progress in 1968. Funding was supplied by the National Center for Education Statistics.[5]

More than one hundred years passed between the government's assignment to collect data on the status of education in the United States in 1867 and the initiation of the program designed to carry out this task in 1968. Since its start, the NAEP has assessed more Americans in more areas than any other federally sponsored testing program in the nation's history. The 1972-73 mathematics assessment, for instance, included data from 25,000 nine-year olds, 30,000 thirteen year-olds, 33,000 seventeen-year olds and 4,500 young adults.[6]

The NAEP has been directed by leading American educators[7] and has conducted three assessments in mathematics, 1972-73, 1977-78 and 1982-83. While other systems for categorizing mathematics content have been prepared, none have included as many categories as the system constructed for the first assessment.[8] Because of this comprehensive structure, the domain defined and used by the NAEP for its first mathematics assessment was examined and used in this study.

A carefully planned series of activities underscored the development of the content categories used in the NAEP's first assessment of mathematics.[9] The first task faced by the NAEP was to define the universe for each skill assessed while ascertaining that each task was properly defined. According to Wilson, defining the universe included the knowledges, skills and attitudes related to the subject while excluding those which were not. On the other hand, a list of this type would be far too long for meaningful assessment purposes and should be reduced in length by using only relevant items in the process.

Wilson was not satisfied with the construct of universe as applied to the NAEP's purposes for the first assessment of mathematics.

> It is clearly beyond the current state of the art to define the universe of behaviors for a complex area in the strict sense discussed above. Yet, it is equally clear that a set of exercises (test items) which form a coherent assessment of a subject area cannot be constructed without some definition of the domain to be tested.[10]

The NAEP took a judgmental approach to this dilemma by relying on individuals' opinions as opposed to logic or statistics. Thus, NAEP's universe was defined by

> ... a set of objectives that represents a consensus of opinion covering many segments of our society regarding the important goals and outcomes of our educational processes in respect to a given subject area.[11]

The NAEP divided mathematics content into seventeen instructional content objectives. Fifteen content objectives included subdivisions which were called content subobjectives for this study. Overall, 126 content subobjectives were stated. Exercises were prepared to measure the objectives at each of the appropriate age levels set by the NAEP.[12]

At first, the NAEP objectives and subobjectives were prepared by subcontractors. These objectives, according to Wilson, differed in quality and tended to assess only those areas most amenable to measurement. Topics which were not amenable to measurement, but no less important to educators, were not examined.[13] Later, the development of the objectives and subobjectives was assigned to the NAEP's Exercise Development Department.

Comments on the NAEP's work have varied. The NAEP has been criticized by Womer and Mastie who questioned the use and application of the results of the assessments[14] and Katzman and Rosen who inferred that the research design followed by the NAEP was influenced by political considerations.[15] On the other hand, Greenbaum et al. pointed out that the NAEP has taken steps to answer the questions posed by its critics.[16] Payne claimed that many benefits resulted from the NAEP's work including strategies for constructing exercises and collecting data.[17]

Additionally, a number of educators reported their use of NAEP data and the assistance provided by this information. McKillip used NAEP data to recommend improvements in teaching division.[18] Carpenter claimed that NAEP data offered insights to areas of performance differences in calculations with decimals for thirteen-year old students who had been given instruction in the skill as opposed to nine-year old students who had not been given instruction.[19] Post was able to suggest techniques designed to help students improve their skills in adding fractions,[20] while Kahle used NAEP data to suggest approaches for increasing minority student enrollment in science.[21] Lapointe and Koffler stated that NAEP data can help educators develop national educational standards.[22]

The evidence shows that the NAEP has constructed a comprehensive, orderly system for categorizing mathematics content in order to assess individual performance. This system is composed of seventeen content objectives which apply to four age groupings. Fifteen objectives include subobjectives.

Exercises have been prepared for each content objective and subobjective. These exercises vary as a function of the age grouping for which they are designed. Thus, the NAEP has set up a system which categorizes mathematics content through a series of objectives and assesses performance through the achievement of these objectives.

The NAEP reviews the objectives continually and revises them as a result of this review. These modifications are implemented and each mathematics assessment has differed from the previous one because of the NAEP's concerns about its evaluation procedures.

The NAEP attends to the mathematics domain. Standardized mathematics achievement tests should do do as well. Therefore, the content objectives and subobjectives constructed by the NAEP could be used to categorize standardized test content. Other systems are availabnle but we found none which were as comprehensive as that prepared by the NAEP.

Analyzing Standardized Test Content

Investigators have commented on standardized achievement tests since the first edition of the Stanford Achievement Test was published in 1922.[23] These comments have ranged from simple descriptions[24] to detailed statistical analyses of the instruments.[25] At first, researchers restricted themselves to presenting categorical information about the tests they examined. Later, the researchers gave their opinions of the quality of the tests they studied as well.

Educational, Psychological and Personality Tests of 1933 and 1934 (EPT) was the first reference work encountered in the course of this review which provided information on the quality of the tests examined as well as categorical information on them.[26] For EPT, professors of education and testing specialists commented on the characteristics of the tests they reviewed. This policy has continued through the Mental Measurements Yearbook series. The Ninth Mental Measurements Yearbook, for instance, contained references to 1,409 tests[27] with 660 educators contributing reviews.[28]

Test reviews of the type found in the Mental Measurements Yearbook series also appear in the literature at large. The Journal of Educational Measurement, for example, publishes test reviews continuously: Other journals do so as well. Education Index, a reference work which catalogs articles in periodicals dealing with education, listed 1,174 citations over 23 pages for its Tests and Scales category in the thirty-fifth volume of this series. This volume covered the twelve month period which began in July, 1984.[29] Clearly, educational measurement and the analysis of instruments used in assessment make up a meaningful portion of the educational literature.

While a considerable amount of information on assessment instruments has appeared in the educational literature, little attention has been given to comprehensive analyses of current tests in a single subject. Researchers who have worked in testing have devoted their attention to other concerns.

Robert Floden et al. found that the content covered by four standardized mathematics tests designed for use with fourth grade students varied with the differences among them leading to possible consequences in terms of instruction.[30] Floden et al. used the Stanford Achievement Test (SAT), the Iowa Tests of Basic Skills (ITBS), the Metropolitan Achievement Tests (MAT) and the Comprehensive Tests of Basic Skills (CTBS) in their study. The contents of these tests were compared to the topics a teacher might cover in his classroom. For operations, one component of the system used by the investigators, test content was similar in terms of the percentage of items devoted to certain specific tasks in addition, subtraction and division. However, differences were observed for the percentage of test items assigned to the topic of addition, overall. In the MAT, twenty-one percent of the items were assigned to addition while the other tests devoted between twelve and fourteen percent of their items to this topic.

More similarities than differences were found among the tests examined, but the differences were important according to the researchers. Six percent of the CTBS items dealt with percentage while the other tests contained no items to test this skill. Alternative number systems were examined in the MAT and SAT but not in the other tests.

> A school district that emphasizes work with percentages in fourth grade would get a distorted picture of progress from the Iowa, which contains no percentage problems. On the other hand, a district which does not introduce percents until the sixth grade would be unnecessarily discouraged by the results of the CTBS which contains six percent problems (sic) involving percentages.[31]

Floden et al. did not identify the levels of the achievement tests they used in their study. Some publishers use grade overlaps at terminal grades for their achievement tests. Thus norms for a fourth grade student appear in Level 2 and Level 3 of the 1970 CAT. The writers should have specified the levels of the achievement tests they used in their study. Similarly, the level of the test used differs with the time the test is administered. A fourth grade student who is tested in the fall of the school year would take the Primary III Battery of the 1973 edition of the SAT, but if he is tested in the spring, he would take the Intermediate I version. (The SAT does not overlap terminal test grades.) Had the investigators provided this information, readers would be able to make more appropriate judgments. This shortcoming limits the study's findings.

Bonnie Armbruster, Robert Stevens and Barak Rosenshine looked at the coverage of three curricula by two tests.[32] The researchers wanted to identify similarities and differences among the instruments as well as subject emphases. All of the materials were designed to assess the performance of third grade students. The reading series used in the study differed in accordance with their emphasis on reading comprehension, generally, and certain categories subsumed by this skill.[33]

The standardized tests were similar with regard to their emphasis on reading comprehension, but all of the tests differed from the reading series used in the study in this respect. The researchers claimed that this finding showed that tests and texts differed in their content coverage.[34] Moreover, a large percentage of the comprehension items taught were not assessed by the standardized tests. Of sixteen categories constructed for the study, no more than seven were taken up by the tests. Most of the test items focused on detail and paraphrasing while inference was emphasized in the texts.

Although there was a strong discrepancy between teaching and testing, the standardized tests used in the study were similar in terms of the topics examined.

> Despite unanswered questions, the present study is important in its demonstration of a feasible methodology for addressing a long-neglected research problem -- determining content coverage and content emphasis of both curricula and tests. More such studies comparing curricula and tests in different content areas and grade levels are needed.[35]

Worthy wanted to determine if there were significant differences in the reading skills measured on two standardized tests, the CTBS and the SAT.[36] The researcher did find the hypothesized differences.

> Each of the six null hypotheses stating that there was no significant difference between the reading skills emphasized at the third grade level in the three basal reading series and the skills measured on the two standardized achievement tests was rejected.[37]

Freeman et al. questioned the use of standardized test scores for instructional purposes.

> Specifically, teachers are encouraged to use standardized test scores to evaluate student achievement on both a group and individual level, to identify students with learning problems, and to assess the effectiveness of instructional strategies that have been used.[38]

The use of standardized test scores for any of these functions, however, must be tempered by the teacher's knowledge of the extent to which the content of the test parallels the content of instruction. Differences in textbooks, school objectives and teacher behaviors as well as the contents of standardized tests may contribute to a discontinuity between content and instruction. This lack of consistency will generate scores which will underestimate student achievement.

The investigators used four standardized elementary mathematics achievement tests in their study, the 1973 SAT, the 1970 MAT, the 1976 CTBS and the 1971 ITBS. Then, Freeman et al. constructed a taxonomy which was made up of three components, (1) presentation mode, (2) material and (3) operations. Through this strategy, the investigators uncovered some differences in test content. For students enrolled in fourth grade, for instance, sixty-three percent of the SAT test items involved whole numbers while the MAT assigned fifty-three percent of its items to this area, and the ITBS, forty-five percent. Other differences were cited but they were not as meaningful as the content assignments.

Freeman et al. concluded that standardized mathematics tests differ in their content. Therefore, the match between the subjects taught and assessed will differ if an inappropriate standardized test is selected. The curriculum may be changed in accordance with the test but this procedure may not be in the students' best interests.

> The match between content taught and content tested is a crucial context for using tests to diagnose student strengths and weaknesses as well as for assessing the strengths and weaknesses of instruction provided.[39]

Educators who decide to use standardized mathematics achievement tests for student assessment may be interested in determining the extent to which their objectives have been achieved. Some tests may be more sensitive to certain objectives than others. Moreover, a test, to some extent, must be sensitive to the objective of assessment or it will not serve as a sound indicator for that purpose. Researchers have made attempts to determine how well a test completes its purpose, but these attempts have included small numbers of tests. This study will analyze the contents of five current standardized secondary mathematics achievement tests in an attempt to provide a comprehensive data base for educators interested in selecting appropriate norm-referenced standardized achievement tests in order to assess their students' performance.

Procedures

At the time of the study's inception, The Ninth Mental Measurements Yearbook[40] was the most comprehensive listing of commercially-prepared standardized achievement tests in the United

States. Therefore, this reference work is an important source for educators working in any area which calls for assessment.

Reviews of The Ninth Mental Measurements Yearbook had not reached the literature when this study began because of its recent publication. However, reviews of past editions were positive, attesting to the Mental Measurements series' value to educators. Raths recommended The Seventh Mental Measurements Yearbook[41] to, "anyone interested in research and/or evaluation."[42] Raths also commented on O. K. Buros' outstanding contribution to the profession.[43] Proger called attention to the enormous effort made by the "father of test reviews,"[44] while Englehard praised Buros by reporting:

> It is difficult to find unused superlatives to characterize The Seventh Mental Measurements Yearbook and its predecessors. They are indeed weighty - 36 pounds on my bathroom scale. I can't believe I read the whole thing, but I have read enough to conclude by saying "Oscar, you are incredible !"[45]

Wilson called the Eighth Mental Measurements Yearbook "a work of immense proportion"[46] and a comprehensive source of information because the editor cited the strengths and weaknesses of each test listed in the test profiles. Through this approach, decisions regarding test selection and use are left to the reader.

Adams described Buros as a critic who looked for honest, objective appraisals from his reviewers who were asked to criticize poor work, call attention to good work and make suggestions for improvements. "Test entries for the 1,184 tests in the Eighth MMY are complete, accurate and helpful."[47] Thompson asked that someone continue Buros' work so that those who deal with tests will not have to rely on a trial and error approach to evaluate them in the future.[48]

Validity: Since no studies were encountered during the course of the literature review which established the validity and reliability of the NAEP classifications for analyzing content, this determination became the first step in the study at hand. "Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. The inferences regarding specific uses are validated, not the test itself."[49] In this context, questions dealing with validity are directed toward inferences about the subject of the assessment and inferences about other behaviors. To answer the first question, the researcher must take steps to determine how well the instrument samples the domain of the topic measured. For the second question, the researcher must assess the value of the instrument as a predictor of other behaviors.

Three types of evidence may be used to describe assessment instruments with regard to validity; criterion-related evidence, content-related evidence and construct-related evidence.[50] The same information may be used for each form of validity. The approach employed, however, differs according to the type.

Criterion-related evidence "demonstrates that test scores are systematically related to one or more outcome criteria."[51] Here the criterion is the variable of primary interest to the researcher. Naturally, the choice of the criterion and the means used to examine it are crucial matters. Researchers may use two strategies to collect evidence for establishing criterion-related validity. For predictive work, the researcher seeks information designed to estimate criterion scores which will emerge at some time in the future. For concurrent work, both sets of information are collected simultaneously. The choice of strategies depends upon the researcher's concerns. In a general sense, the difference between the two forms of evidence is based on time.

Content-related evidence is used to determine if a sample of behaviors represent those of the domain under study. Therefore, the researcher must ascertain if the items included in the instrument are similar to those making up the domain. Since content-related evidence is a major concern during instrument development procedures, professional judgment takes on a key role in terms of deciding what will be measured by the instrument.

Construct-related evidence looks at the test score as a measure of the psychological characteristic under study and may be implied as the researcher examines a criterion for the construct. In turn, a construct may be defined as something which cannot be observed but is stated by the investigator to summarize regularities in a person's behavior.[52] Or, a construct is an idea prepared by the investigator to explain and organize an aspect of existing knowledge, "... because construct validation refers to a broader and more abstract kind of behavioral description, and because there is no single acceptable criterion measure against which to validate a measure of a construct, construct validity typically requires the gradual accumulation of evidence from a number of sources."[53] The accumulation of evidence from various sources was cited by Sax who listed six steps in the construct validation process, (1) justifying the construct in terms of its educational and psychological properties, (2) distinguishing the construct examined from similar constructs, (3) measurability, (4) aquiring evidence from different sources, (5) demonstrating that the construct does not correlate highly with irrelevant variables and (6) modifying the construct in accordance with the evidence gathered.[54]

Face validity was not discussed in the 1985 edition of Standards but was attended to in the 1974 version. It is mentioned here because of its classical value and for completeness. Face validity is the appearance of validity, has no value for developing inferences from scores, and is the "reasonableness and acceptability of a test for use with a particular group."[55] While face validity may be useful in some instances, it cannot be used as a substitute for the other forms of evidence because judgments of face validity cannot be used to develop conclusions as to how faithfully a score represents the topic in question or to predict behavior.

Reliability: Reliability is, "the degree to which test scores are free from errors of measurement."[56] A respondent's effort, ease, and fatigue among other characteristics may vary from one test administration to another. Consequently, scores will differ from one test to another.

At least two sets of measurements are necessary for estimating reliability. These measures may be obtained by administering the same instrument to a subject twice or giving either alternate or parallel forms of the instrument to different subjects believed to have no biases which would affect their responses. The first approach does not control for memory and accepts this factor as a potential systematic source of variance. The second approach does not control for item inequivalence and accepts this factor as a potential source of systematic variance.

The reliability and validity of the NAEP objectives and subobjectives for use in this study were determined by forming a panel of referees who were asked to comment on the use of the objectives and subobjectives for this purpose. The referees selected for this component of the study were employees of the School District of Philadelphia. Each referee had at least five years of classroom teaching experience in the Philadelphia public schools, held an advanced degree in mathematics education, had taken at least one course in tests and measurements or statistics and was serving as a principal, mathematics supervisor or mathematics coordinator when the study was conducted.

Each referee was asked to categorize the items from three sample tests. The researcher joined this component of the study by categorizing the items twice. The researcher's categorizations were separated by ten days in order to help control for memory.[57] This procedure was designed to establish interrater and intrarater reliability. In both instances, acceptable reliability coefficients emerged from the analysis.

Item Classification: The researcher acquired the tests used in the study from their publishers and prepared a tally sheet which included spaces for the name of the test examined, the date of its publication, the test's authors, the form examined, the level examined, and the grades for which the test was designed. The tally sheet accommodated sixty items, with multiple sheets used when a test's item count exceeded this figure.

Each item in each test was examined and assigned to an NAEP subobjective. The grade or grade cluster covered by each test was used to group the tests for analysis. Tests which were designed for use with students enrolled in more than one grade were assigned to each eligible group. Thus, a test designed to measure student performance in grades nine and ten was analyzed in both groups. For study purposes, tests were treated as if they were administered at the end of the school year. Each group of tests was analyzed separately for the seventeen content objectives and the 126 content subobjectives. Thus, eight analyses were planned, one for each grade on the objectives and one,

for the subobjectives.

Statistical Procedures: A chi-square one-sample test was used to determine if the number of content objectives examined in each test used in the study differed significantly.[58] The same strategy was used to analyze the subobjectives. The *Statistical Package for the Social Sciences* ($SPSS^x$) contains a program designed to perform a single-sample chi-square and it was used to analyze the data.[59]

Results

Table 1 shows the tests used in the study and the grades they covered. Eleven tests were used and seven were assigned to each grade grouping. Eleven tests appear, representing five series. Three of the eleven tests included norms for students in grades nine through twelve, addressing all of the grades used in the study. While three tests could be used to assess ninth grade student performance alone, there were no tests which contained norms for eleventh grade or twelfth grade students individually.

Since the tests designed to measure student performance in eleventh and twelfth grades were the same, only one analysis was necessary to cover both. Thus, the number of analyses used in the study was reduced from eight to six. Tables 2, 3, and 4 show the relevant information for the study. Each table presents a list of the tests involved and the number of objectives and subobjectives addressed by each. Six one-sample chi-square analyses were conducted and significance was not reached in any analysis. Therefore, the standardized tests examined in this study did not differ in terms of the numbers of NAEP objectives and subobjectives examined. Consequently, it would be appealing to say that any test used in the study for student assessment would yield the same information. This type of statement would be simplistic because the tests differ in the proportion and numbers of items directed toward individual objectives and subobjectives. Tables 5 through 15 show the objectives and subobjectives addressed by each test examined in the study.

Comments

The tests examined in the study did not differ significantly in terms of the number of NAEP objectives or subobjectives addressed on a grade by grade basis. Consequently, it seems as if similar information pertaining to student assessment would be provided by each test if number alone is used as the criterion

Although fourteen of seventeen (82 %) NAEP objectives were studied by all of the tests combined, no test addressed more than eleven (65 %). Properties of Numbers, Mathematical Proof and Attitude and Interest Items did not appear in any test. Similarly, sixty of the 126 subobjectives (48 %) were examined with no test dealing with more than twenty-five (20 %). This information shows that the NAEP system is not being followed by the publishers of standardized norm-referenced mathematics tests in the secondary school grades.

Some subobjectives may be appropriate to the elementary school grades only and their absence in the secondary grade tests may be legitimate. Given this point, it still remains clear that test publishers are not using the NAEP system in preparing their tests. With recent research demonstrating the relatively low status of American students among their peers in other countries, it may be time for test publishers to consider the NAEP system when they construct their tests.

Table 1

Standardized Mathematics Tests Examined in the Study:
Test Level and Grade Coverage

Test	Level	Grade Coverage 9	10	11	12
California Achievement Tests	19	x	x		
	20		x	x	x
Comprehensive Tests of Basic Skills	J	x	x	x	x
	K			x	x
Metropolitan Achievement Tests	Advanced 1	x			
	Advanced 2		x	x	x
SRA Survey of Basic Skills	36	x			
	37		x	x	x
Stanford Achievement Test	Task 1	x	x	x	x
	Task 2	x	x	x	x
		7	7	7	7

Table 2

Number of NAEP Objectives and Subobjectives Addressed by Standardized Mathematics Achievement Tests - Grade 9

Test	Number of Objectives	Number of Subobjectives
CAT (19)	9	22
CTBS (J)	10	21
MAT (Advanced 1)	7	21
SAT (Advanced)	9	24
SAT (TASK 1)	6	12
SAT (TASK 2)	10	20
SRA (36)	10	18

Chi-Square Objectives 1.77, Significance .94
Subobjectives 4.90, Significance .56

Table 3

Number of NAEP Objectives and Subobjectives Addressed by Standardized Mathematics Achievement Tests – Grade 10

Test	Number of Objectives	Number of Subobjectives
CAT (19)	9	22
CAT (20)	11	25
CTBS (J)	10	21
MAT (Advanced 2)	9	21
SAT (TASK 1)	6	12
SAT (TASK 2)	10	20
SRA (37)	10	19

Chi-Square Objectives 1.41, Significance .96
Subobjectives 3.50, Significance .74

Table 4

Number of NAEP Objectives and Subobjectives Addressed by Standardized Mathematics Achievement Tests - Grades 10 & 11

Test	Number of Objectives	Number of Subobjectives
CAT (20)	11	25
CTBS (J)	10	21
CTBS (K)	10	20
MAT (Advanced 2)	9	24
SAT (TASK 1)	6	12
SAT (TASK 2)	10	20
SRA (37)	10	19

Chi-Square Objectives 1.77, Significance .94
Subobjectives 4.90, Significance .56

Table 5

NAEP Objectives and Subobjectives Addressed by CAT - 19

Objective	Subobjective
Number and Numeration Concepts	Numeration Systems
	Prime & Composite Numbers
	Divisibility, Greatest Common Factor, Least Common Multiple
	The Real Number Line
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion & Percent
	Rounding Off
Estimation and Measurement	Time
	Money
Exponents & Logarithms	Exponential & Logarithmic Equations
Algebraic Expressions	Combining Like Terms
	Evaluating Expressions
Functions	Quadratic Equations & Their Graphs
	Maxima and Minima of Functions
Probability and Statistics	Permutations & Combinations
	Outcomes, Samples, Spaces and Events
	Probability of an Event
	Measures of Central Tendency
Geometry	Circles and Spheres
	Cartesian Coordinates
Business & Consumer Mathematics	Personal and Bank Records

Table 6

NAEP Objectives and Subobjectives Addressed by CAT - 20

Objective	Subobjective
Number and Numeration Concepts	Numeration Systems
	Odd and Even Numbers
	Prime and Composite Numbers
	Real Numbers
	Rational Numbers
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion & Percent
	Rounding Off
Sets	Properties
Estimation & Measurement	Time
	Weight
	Area-Volume
	Money
Exponents & Logarithms	Exponential Equations
Algebraic Expressions	Combining Like Terms
	Removing Parentheses
Equations and Logic	Solving Equations and Inequalities with Absolute Values
Functions	Y-Intercept
Probability and Statistics	Permutations & Combinations
	Outcomes, Samples, Spaces and Events
	Measures of Central Tendency
	Measures of Dispersion
Geometry	Circles and Spheres
Business & Consumer Mathematics	Personal and Bank Records

Table 7

NAEP Objectives and Subobjectives Addressed by CTBS - Level J

Objective	Subobjective
Number and Numeration Systems	Numeration Systems
	Odd and Even Numbers
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion, Percent
	Computation with Approximate Data
Sets	Properties
Estimation & Measurement	Time
	Money
	Conversion Relations
Algebraic Expressions	Properties of Expresssions
	Combining Like Terms
	Removing Parentheses
Equations & Logic	Finding Solutions in One Variable
Functions	Y Intercept
Probability and Statistics	Permutations & Combinations
	Probability of an Event
	Descriptive Statistics
	Measures of Central Tendency
Geometry	Circles & Spheres
Trigonometry	Relations among Functions

Table 8

NAEP Objectives and Subobjectives Addressed by
CTBS-Level K

Objective	Subobjective
Number and Numeration Concepts	Numeration Systems
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion, Percent
Sets	Properties
Estimation & Measurement	Time
	Money
Exponents & Logarithms	Exponential & Logarithmic Equations
Algebraic Expressions	Properties of Expressions
	Manipulation of Expressions
	Combining Like Terms
Equations & Logic	Graphs of Equations
	Maxima & Minima of Functions
Probability & Statistics	Basic Probability Concepts
	Permuations & Combinations
	Outcomes, Samples, Spaces and Events
	Measures of Central Tendency
	Measures of Dispersion
Geometry	Circles & Spheres
Business and Consumer Mathematics	Personal and Bank Records

Table 9

NAEP Objectives and Subobjectives Addressed by
MAT - Advanced 1

Objective	Subobjective
Number and Numeration Concepts	Decimal Place Value
	Prime & Composite Numbers
	Greatest Common Factor Least Common Multiple
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion & Percent
	Rounding Off
Estimation and Measurement	Time
	Distance
	Area-Volume
	Weight
Exponents & Logarithms	Exponential Equations
Algebraic Expressions	Operations with Expressions
	Evaluating Expressions
Probability & Statistics	Probability of an Event
	Representing Data
Geometry	Points, Lines and Planes
	Rays, Segments and Angles
	Polygons and Polyhedra
	Angle Measurement
	Cartesian Coordinates

Table 10

NAEP Objectives and Subobjectives Addressed by
MAT - Advanced 2

Objective	Subobjective
Number and Numeration Concepts	Decimal-Place Value
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion and Percent
	Rounding Off
Estimation and Measurement	Time
	Distance
	Area-Volume
	Conversion Relations
Exponents and Logarithms	Exponential Equations
Algebraic Expressions	Factoring
	Evaluating Expressions
Functions	Evaluating Functions
	Y-Intercept
Probability and Statistics	Basic Probability Concepts
	Representing Data
Geometry	Rays, Segments and Angles
	Polygons and Polyhedra
	Circles and Spheres
	Angle Measurement
Trigonometry	Trigonometric Functions

Table 11

NAEP Objectives and Subobjectives Addressed by
SRA Survey of Basic Skills - 36

Objective	Subobjective
Number and Numeration Concepts	Numeration Systems
	Prime and Composite Numbers
	Greatest Common Factor - Least Common Multiple
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion, Percent
	Rounding Off
Estimation and Measurement	Time
Exponents and Logarithms	Exponential Equations
Algebraic Expressions	Combining Like Terms
	Removing Parentheses
Functions	Writing Equations of Quadratic Functions
	Maxima and Minima of Functions
Probability and Statistics	Permutations and Combinations
	Measures of Central Tendency
Geometry	Circles and Spheres
Trigonometry	Trigonomet: ic Functions
Miscellaneous Topics	Sequences and Series

Table 12

NAEP Objectives and Subobjectives Addressed by
SRA Survey of Basic Skills - 37

Objective	Subobjective
Number and Numeration Concepts	Numeration Systems
	Odd and Even Numbers
	Prime and Composite Numbers
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion and Percent
	Rounding Off
Estimation and Measurement	Time
Exponents and Logarithms	Exponential Equations
Algebraic Expressions	Combining Like Terms
	Operations with Expressions
Functions	Writing Equations of Quadratic Functions
	Maxima and Minima of Functions
Probability and Statistics	Permutations and Combinations
	Probability of an Event
	Measures of Central Tendency
etry	Circles and Spheres
scellaneous Topics	Binomial Expansion
siness and Consumer Mathematics	Personal and Bank Records

Table 13

NAEP Objectives and Subobjectives Addressed by
Stanford Achievement Test - Advanced

Objective	Subobjective
Number and Numeration Concepts	Decimal-Place Value
	Prime and Composite Numbers
	Greatest Common Factor - Least Common Multiple
	Factorials
	Real Numbers
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Complex Numbers
	Ratio, Proportion, Percent
	Rounding Off
Sets	Set Operations
Estimation and Measurement	Time
	Weight
	Area-Volume
	Conversion Relations
Exponents and Logarithms	Exponential Equations
Algebraic Expressions	Evaluating Expressions
Probability and Statistics	Basic Probability Concepts
	Measures of Central Tendency
	Representing Data
	Circles and Spheres
	Cartesian Coordinates
Logic	Logic
Business and Consumer Mathematics	Personal and Bank Records

Table 14

NAEP Objectives and Subobjectives Addressed by
Stanford Achievement Test - TASK 1

Objective	Subobjective
Number and Numeration Concepts	Numeration Systems
	Integers
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion, Percent
	Rounding Off
Estimation and Measurement	Time
	Distance
Algebraic Expressions	Combining Like Terms
	Evaluating Expressions
Probability and Statistics	Permutations and Combinations
Business and Consumer Mathematics	Personal and Bank Records

Table 15

NAEP Objectives and Subobjectives Addressed by
Stanford Achievement Test - TASK 2

Objective	Subobjective
Number and Numeration Concepts	Numeration Systems
	Odd and Even Numbers
	Prime & Composite Numbers
Arithmetic Computation	Whole Numbers
	Rational Numbers
	Ratio, Proportion, Percent
	Rounding Off
Estimation & Measurement	Time
	Weight
	Conversion Relations
Exponents and Logarithms	Exponential Equations
Algebraic Expressions	Evaluating Expressions
Equations and Logic	Solving Equations & Inequalities with Absolute Values
Functions	Evaluating Functions
	Analysis of Graphs of Quadratic Functions
	Maxima and Minima of Functions
Probability and Statistics	Probability of an Event
	Measures of Dispersion
Geometry	Circles and Spheres
Business and Consumer Mathematics	Personal and Bank Records

Footnotes

[1]Alan Solomon, "A Content Analysis and Comparison of Standardized Elementary Mathematics Tests Published in the United States of America from 1902 to 1978" (Doctoral Dissertation, Temple University, 1985).

[2]David A. Payne, "Measurement in Education," Encyclopedia of Educational Research, editor in chief, Harold H. Mitzel (5th ed.; [New York]: The Free Press, 1982), p. 1184.

[3]National Center for Education Statistics, Mathematics Objectives: Second Assessment, U. S. Educational Resources Information Center, ERIC Document ED 156 439, 1978.

[4]H. A. Wilson, "A Judgmental Approach to Criterion Referenced Testing," Problems in Criterion Referenced Measurement ed. C. W. Harris, M. C. Alkin and W. J. Popham (Los Angeles: University of California, 1974), p. 27.

[5]Ibid., pp. 29-31.

[6]National Assessment of Educational Progress, Mathematics Technical Report: Exercise Volume; Mathematics Report No. 04-MA-20 (Washington, D. C.: U. S. Government, 1977), p. 1.

[7]William Greenbaum, Michael S. Garet and Ellen R. Solomon, Measuring Educational Progress: A Study of the National Assessment (New York: McGraw-Hill, 1977), p. 16.

[8]Robert E. Floden et al. Don't They All Measure the Same Thing ? Consequences of Selecting Standardized Tests, U. S. Educational Resources Information Center, ERIC Document ED 167 632, 1978.

[9]Wilson, p. 29.

[10]Ibid.

[11]Ibid.

[12]Eleanor N. Norris and John Bowers, National Assessment of Educational Progress, Mathematics Objectives, U. S. Educational Resources Information Center, ERIC Document ED 063 140, 1970.

[13]Wilson, p. 31.

[14]Frank B. Womer and Marjorie M. Mastie, "How Will National Assessment Change American Education ?" Phi Delta Kappan 53:118-20, October, 1971, p 118.

[15]Martin Katzman and Ronald Rosen, "The Science and Politics of National Education Assessment," The Record 71:571-86, May 1970, pp. 579-80.

[16]Greenbaum et al., pp. 169-70.

[17]Payne, 1982, p 1184.

[18]William D. McKillip, "Computational Skills in Division: Results and Implications from National Assessment," The Arithmetic Teacher 28:34-7, March, 1981.

[19]Thomas P. Carpenter, "Decimals: Results and Implications from National Assessment," The Arithmetic Teacher, 28:34-7, April, 1981.

[20]Thomas D. Post, "Fractions: Results and Implications from National Assessment," The Arithmetic Teacher, 28:26-31, May, 1981.

[21]Jane B. Kahle, "Can Positive Minority Attitudes Lead to Achievement Gains in Science ? Analysis of the 1977 National Assessment of Educational Progress, Attitudes toward Science," Science Education, 66:539-46, July, 1982.

[22]Archie E. Lapointe and Stephan L. Koffler, "Your Standards or Mine ? The Case for the National Assessment of Educational Progress," Educational Researcher, 11:4-9, December, 1982.

[23]Truman L. Kelley, Giles M. Ruch and Lewis M. Terman, The Stanford Achievement Test (Yonkers-on-Hudson, NY: World Book Co, 1922).

[24]Maryann Mercer, "The Content of Two Mathematics Achievement Tests," School Science and Mathematics 78:669-74, December, 1978.

[25]Esther Shani and Joseph Petresko, "Structural Components Derived from Evaluating Standardized Tests," Journal of Educational Measurement 13: 283-96, Winter, 1976.

[26]Oscar K. Buros, ed., Educational, Psychological and Personality Tests of 1933 and 1934, Rutgers University Bulletin, Vol. 11, No. 11, Studies in Education, No. 7, (New Brunswick, NJ: Rutgers University, May, 1935).

[27]James V. Mitchell, ed., The Ninth Mental Measurements Yearbook (Lincoln, NB: The Buros Institute of Mental Measurements, The University of Nebraska - Lincoln, 1985), xiv.

[28]Ibid.

[29]Mary Louise Hewitt, ed., Education Index (New York: H. W. Wilson, 1985).

[30]Robert E. Floden et al., Don't They All Measure The Same Thing ? Consequences of Selecting Standardized Tests, U. S. Educational Resources Information Center, ERIC Document ED 167 632, 1978.

[31]Ibid., p. 8.

[32]Bonnie B. Armbruster, Robert J. Stevens and Barak Rosenshine, Analyzing Content Coverage and Emphases: A Study of Three Curricula and Two Tests, Technical Report No. 26, National Institute of Education, Washington, D. C., U. S. Educational Resources Information Center, ERIC Document ED 136 238, 1977.

[33]The reading series were: (1) The Mysterious Wisteria published by Economy Co., (2) Reading 360 published by Ginn and (3) Fiesta published by Houghton-Mifflin.

[34] Armbruster et al., p. 7.

[35] Ibid., p. 9.

[36] Diane H. Worthy, "A Comparative Content Analysis of Three Basal Reading Series and Two Standardized Achievement Tests" (Doctoral Dissertation, University of South Carolina, 1979).

[37] Ibid., p. 50.

[38] Donald J. Freeman et al., "A Closer Look at Standardized Tests," The Arithmetic Teacher, 29:50-54, March, 1983.

[39] Ibid., p. 54.

[40] Mitchell, 1985.

[41] Oscar K. Buros, ed., The Seventh Mental Measurements Yearbook (Highland Park, NJ: Gryphon Press, 1972).

[42] James Raths, Review of The Seventh Mental Measurements Yearbook, Educational Leadership, 33:69-71, October, 1975, p. 71.

[43] Ibid.

[44] Barton B. Proger, Review of The Seventh Mental Measurements Yearbook, Journal of Special Education, 7:91-93, Spring, 1973, p. 93.

[45] Max D. Englehard, Review of The Seventh Mental Measurements Yearbook, Educational and Psychological Measurement, 32:843-46, Autumn, 1972, p. 846

[46] Edward S. Wilson, Review of The Eighth Mental Measurements Yearbook, Educational and Psychological Measurement, 40:263-65, Spring, 1980.

[47] Georgia S. Adams, Review of The Eighth Mental Measurements Yearbook, Educational Horizons, 58:67-70, Fall, 1979.

[48]Jane Thompson, Review of The Eighth Mental Measurements Yearbook, Measurement and Evaluation in Guidance, 12:187-88, October, 1979.

[49]Standards for Educational and Psychological Testing (Washington, D. C.: American Psychological Association, 1985), p. 9.

[50]Ibid.

[51]Ibid., p 11.

[52]Richard M. Wolf, "Validity of Tests," Encyclopedia of Educational Research, editor-in-chief, Harold H. Mitzel (5th ed.; [New York]: The Free Press, 1982), p. 1993.

[53]Ibid., p. 1994.

[54]Gilbert Sax, Principles of Educational and Psychological Measurement and Evaluation (Belmont, CA: Wordsworth Publishing Co., 1980), pp 296-97

[55]Wolf, "Validity of Tests," p. 1991.

[56]Standards, p. 19

[57]Merle W. Tate, Statistics in Education (New York: Houghton-Mifflin, 1960), p. 338.

[58]Sidney Siegel, Nonparametric Statistics (New York: McGraw-Hill, 1956), p. 42.

[59]SPSSx User's Guide, (New York:McGraw-Hill, 1983), pp. 673-74.

Bibliography

Adams, Georgia S. Review of The Eighth Mental Measurements Yearbook. Educational Horizons. 58:67-70. Fall, 1979.

Armbruster, Bonnie B., Robert J. Stevens and Barak Rosenshine. Analyzing Content Coverage and Emphases: A Study of Three Curricula and Two Tests. Technical Repoort No. 26. National Institute of Education, Washington, D. C. U. S. Educational Resources Information Center, ERIC Document ED 136 238, 1978.

Buros, Oscar K. ed. Educational, Personality and Psychological Tests of 1933 and 1934 Rutgers University Bulletin, Vol. 11, No. 11, Studies in Education, No. 7. New Brunswick, NJ: School of Education, Rutgers University. May, 1935.

———. The Eighth Mental Measurements Yearbook. Highland Park, NJ: Gryphon Press, 1978.

———. The Seventh Mental Measurements Yearbook. Highland Park, NJ. Gryphon Press, 1972.

Carpenter, Thomas P. "Decimals: Results and Implications from National Assessment." The Arithmetic Teacher, 28:334-7. April, 1981.

Encyclopedia of Educational Research, editor-in-chief, Harold H. Mitzel. 5th ed. New York: The Free Press, 1982.

Englehard, Max D. Review of The Seventh Mental Measurements Yearbook. Educational and Psychological Measurement. 32:843-46. Autumn, 1972.

Floden, Robert E., Andrew C. Porter, William H. Schmidt and Donald J. Freeman. Don't They All Measure the Same Thing ? Consequences of Selecting Standardized Tests. U. S., Educational Resources Information Center, ERIC Document ED 167 732, 1978.

Freeman, Donald J., Therese M. Kuhs, Lucy B. Knappen and Andrew C. Porter. "A Closer Look at Standardized Tests." The Arithmetic Teacher, 29:50-54. March, 1982.

Greenbaum, William, Michael S., Garet and Ellen R. Solomon. Measuring Educational Progress: A Study of the National Assessment. New York: McGraw-Hill, 1977.

Harris, Chester W., M. C. Alkin and W. James Popham, eds. Problems in Criterion Referenced Measurement. Los Angeles: University of California, 1974.

Hewitt, Mary Louise, ed. Education Index. New York: H. W. Wilson, 1985.

Kahle, Jane B. "Can Positive Minority Attitudes Lead to Achievement Gains in Science ? Analysis of the 1977 National Assessment of Educational Progress, Attitudes toward Science." Science Education, 66:539-46. July, 1982.

Katzman, Martin and Ronald Rosen, "The Science and Politics of National Educational Assessment," The Record, 71:571-86. May, 1970.

Kelley, Truman L., Giles M. Ruch and Lewis M. Terman. The Stanford Achievement Test (Yonkers-on-Hudson, NY: World Book Co., 1922).

Lapointe, Archie E. and Stephan L. Koffler. "Your Standard or Mine ? The Case for the National Assessment of Educational Progress." Educational Researcher, 11:4-9. December, 1982.

McKillip, William. "Computational Skills in Division: Results and Implications from National Assessment." The Arithmetic Teacher, 28:34-37. March, 1981.

Mercer, Maryann. "The Content of Two Mathematics Achievement Tests." School Science and Mathematics," 78:669-74. December, 1978.

Mitchell, James V. ed. The Ninth Mental Measurements Yearbook. Lincoln, NB: The Buros Institute of Mental Measurements, University of Nebraska - Lincoln, 1985.

National Assessment of Educational Progress. Mathematics Technical Report: Exercise Volume: Mathematics Report No. 04-MA-20. Washington, D. C.: U. S. Government, 1977.

National Center for Education Statistics. Mathematics Objectives: Second Assessment. U. S. Educational Resources Information Center, ERIC Document ED 156 439, 1978.

Norris, Eleanor L. and John E. Bowes. National Assessment of Educational Progress, Mathematics Objectives. Denver, CO: Education Commission of the States, 1970. U. S. Educational Resources Information Center, ERIC Document ED 063 140, 1970.

Post, Thomas D. "Fractions: Results and Implications from National Assessment." The Arithmetic Teacher, 28:26-31. May, 1981.

Proger, Barton B. Review of The Seventh Mental Measurements Yearbook. Journal of Special Education, 7:91-93. Spring, 1973.

Raths, James, Review of The Seventh Mental Measurements Yearbook. Educational Leadership, 33:69-71. October, 1975.

Sax, Gilbert. Principles of Educational and Psychological Measurement and Evaluation. Belmont, CA: Wordsworth Publishing Co., 1980.

Shani, Esther and Joseph Petresko. "Structural Components Derived from Evaluating Standardized Tests." Journal of Educational Measurement, 13:283-96. Winter, 1976.

Siegel, Sidney. Nonparametric Statistics. New York: McGraw-Hill, 1956.

Solomon, Alan. "A Content Analysis and Comparison of Standardized Elementary Mathematics Tests Published in the United States of America from 1902 to 1978." EdD Dissertation, Temple University, 1985.

SPSSx User's Guide. New York: McGraw-Hill, 1983.

Standards for Educational and Psychological Testing. Washington, D. C.: American Psychological Association, 1985.

Tate, Merle W. Statistics in Education. New York: Houghton-Mifflin, 1960.

Thompson, Jane. Review of The Eighth Mental Measurements Yearbook. Measurement and Evaluation in Guidance. 12:187-88. October, 1979.

Wilson, Edward S. Review of the Eighth Mental Measurements Yearbook. Educational and Psychological Measurement. 40:263-65, Spring, 1980.

Womer, Frank B. and Marjorie M. Mastie. "How Will National Assessment Change American Education ?" Phi Delta Kappan, 53:118-20. October, 1971.

Worthy, Diane H. "A Comparative Content Analysis of Three Basal Reading Series and Two Standardized Achievement Tests." EdD Dissertation, University of South Carolina, 1979.