DOCUMENT RESUME

ED 104 897 TM 004 339

TITLE	Individual Pupil Report: Explanatory Materials. The Second Report of the 1972-73 Michigan Educational Assessment Program.
INSTITUTION	Michigan State Dept. of Education, Lansing. Research, Evaluation, and Assessment Services.
PUB DATE	Mar 73
NOTE	37p.; For related documents, see TM 004 340-342 and 555-556
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Achievement Tests; *Basic Skills; Cognitive Development; Comparative Statistics; *Educational Assessment; Elementary Education; Elementary School Students; Raw Scores; Scores; *State Programs; Statistical Analysis; *Testing; *Test Interpretation; Test Reliability; Test Validity
IDENTIFIERS	*Michigan Educational Assessment Program

ABSTRACT

This explanatory booklet and the materials that accompany it have two major purposes. The first purpose is to provide local school officials information regarding the performance on basic skills achievement of each student who took the 1972-73 Michigan Educational Assessment Battery. The second purpose is to provide local officials with information that will assist them in understanding and interpreting their students' scores. This booklet has four major sections. The first section describes the content of each test--word relationships, reading, mechanics of written English, and mathematics--in the educational assessment battery. The second section describes cautions that must be exercised in the interpretation of individual pupil scores from the program. The third section describes the materials that accompany this booklet and explains the pupil scores that they contain. The fourth section defines statistical terms used in the educational assessment program and provides technical information about the educational assessment battery. (Author)

INDIVIDUAL PUPIL REPORT: EXPLANATORY MATERIALS

The Second Report of the 1972-73
Michigan Educational Assessment Program

Prepared by Research, Evaluation and Assessment Services

Michigan Department of Education

March, 1973

FOREWORD

The Michigan Educational Assessment Program (MEAP) was initiated by the State Board of Education, supported by the Governor, and funded by the legislature initially through enactment of Act 307 of the Public Acts of 1969, and subsequently under Act 38 of the Public Acts of 1970. This report, the second in the 1972-73 series, provides information which will assist school district staff in the interpretation of the educational assessment results for individual pupils.

The State Board of Education has adopted a six-step process as a guide or model for improving Michigan education. The six steps are: the identification of common goals, the development of performance objectives, the assessment of educational needs, the analysis of delivery systems, the evaluation and testing of these systems or programs, and recommendations for educational improvement. This report presents information for the third step--the assessment of educational needs. Educational assessment provides general information on student needs which, along with other information gathered by local educators, will assist in identifying areas of need on the part of local schools and pupils. Analysis of the systems for delivering educational services and the specific evaluations of the areas so identified may then be initiated by local school officials in order to determine the extent to which changes in curricula and resource allocations are justified. Thus, the educational assessment program can contribute to the improvement of educational programs for Michigan children and youth.

Thanks are due to a large number of individuals and groups for making the Michigan Educational Assessment Program a reality and for continuing to work with it in its fourth year, 1972-73: to the State Board of Education for initially proposing it and continuing to support it, to the Governor and legislature for actively supporting it, and to Michigan educators for assisting with it. The program was designed and administered by the Research, Evaluation, and Assessment Services Unit, Michigan Department of Education, with the assistance of Educational Testing Service of Princeton, New Jersey, and the advice of the MEAP Council.

This report was prepared by Mr. Robert Huyser with the assistance of Dr. Thomas Fisher, Mrs. June Olsen, Dr. Daniel E. Schooley, and Dr. David Donovan. Questions or requests for additional information relative to this report should be directed to the educational assessment staff.

John W. Porter
Superintendent of
Public Instruction

TABLE OF CONTENTS

INTRODUCTION

This explanatory booklet and the materials that accompany it have two major purposes. This first purpose is to provide local school officials with information regarding the performance on basic skills achievement of each student who took the 1972-73 Michigan Educational Assessment Battery. The second purpose is to provide local officials with information that will assist them in understanding and interpreting their students' scores.

This booklet has four major sections. The first section describes the content of each test--word relationships, reading, mechanics of written English, and mathematics--in the educational assessment battery. The second section describes cautions that must be exercised in the interpretation of individual pupil scores from the program. The third section describes the materials that accompany this booklet and explains the pupil scores that they contain. The fourth section defines statistical terms used in the educational assessment program and provides technical information about the educational assessment battery.

SECTION I

DESCRIPTION OF THE EDUCATIONAL ASSESSMENT MEASURES

The educational assessment battery given in January, 1973, was developed according to specifications developed jointly by Educational Testing Service and the Michigan Department of Education. The tests from 1972 were reviewed and revised by panels of Michigan teachers, subject matter specialists, and others.* The resulting 1973 tests incorporated a number of the questions that were developed by the panels and pretested in 1972. However, enough questions from the previous test were retained to permit comparison of test results between years.

The 1972-73 educational assessment battery consisted of four different achievement measures--reading, word relationships, mechanics of written English, and mathematics administered in that order. Each measure was a separate, timed test at the fourth and seventh grade. This section will describe the tests in the order in which they appear on the reports.

Word Relationships

The fourth grade word relationships test contained 45 verbal analogy problems which were designed to measure students' knowledge of the meaning of words and the relationships between words and concepts. Twenty minutes were allowed to work on the test. The seventh grade test contained 38 questions of the same type. The time allowed to work on this test was 15 minutes.

---

*Department members and an Educational Testing Service representative were also present at the panel meetings. The names and affiliations of these panel members are available upon request from the Office of Research, Evaluation and Assessment Services.

Reading

The fourth grade reading test contained 50 questions which assessed paragraph comprehension, ability to understand words from the context in which they are encountered, and ability to identify the correct synonym for a word. Students at the fourth grade level were allowed 35 minutes to work on this test. The seventh grade test was similar in content but contained 60 questions to be answered in 40 minutes.

Mechanics of Written English

The mechanics of written English test consisted of three parts for fourth graders, and three for seventh graders, each separately timed. In part A, spelling, students were to identify misspelled words. The fourth grade test presented 15 items to be completed in 5 minutes; the seventh grade test presented 20 items to be completed in 6 minutes. Part B, effectiveness of written expression required students to select the best way of expressing a thought or the best word or phrase to complete a sentence. The fourth grade test contained 28 items while the seventh grade test contained 30 items; all pupils were allowed 17 minutes to complete part B. Recognizing errors of punctuation and capitalization was the object of part C. The fourth grade booklet had 12 items and allowed 8 minutes, and the seventh grade booklet had 14 items and allowed 7 minutes.

Mathematics

The mathematics test at both grade levels involved mathematical reasoning, problem solving and computation. In addition, problems in the seventh grade test involved algebraic and geometric concepts. Pupils at both grade levels had 30 minutes in which to answer 40 questions.

Composite Achievement

A composite achievement score was computed for each student. The composite score was obtained by averaging the individual's standard scores on the reading, the mechanics of written English, and the mathematics tests. The test scores were combined in this way so that each score would contribute equally to the average--despite the fact that the number of items was different on the three tests.

It should be noted that the word relationships test score was not included in the calculation of the composite achievement score. Analogies such as those contained in the word relationships test are not a common subject of direct instruction. Furthermore the word relationships score is believed to respond more slowly than the other scores to the influence of schooling and may be considered to be a measure of developed verbal ability. Therefore, it was excluded to focus the composite achievement score upon those aspects of basic skills achievement that respond most readily to instruction.

SECTION II

PRECAUTIONS IN THE INTERPRETATION OF PUPIL SCORES

A number of precautions should be observed in interpreting the individual pupil scores from the Michigan Educational Assessment Program. The following precautions are particularly important.

Measurement Error

All measurement is subject to error. Scores resulting from educational assessment tests are no exception. Therefore, in any use of the results an allowance for error should be made. The discussion on reliability in Section IV of this report contains estimates of the amount of error associated with individual pupil scores on each of the assessment instruments.

The score of a pupil at any time will be the result of a number of influences. These influences include, but are not limited to: previous educational experience, nature of the curriculum, teaching effectiveness, home environment, and peer culture. Furthermore, a pupil's score is likely to reflect the combined impact of these forces over several years--even at the fourth grade level.

Content is General

The educational assessment results provide a general idea of the basic skills achievement levels of each pupil compared to the basic skill levels of pupils throughout the state. In other words, the tests were purposely made general in their content so that they would be useful in the varied school situations that are to be found throughout the state. Thus an effort

has been made to focus upon the broader outcomes in reading, mechanics of writing, and mathematics sought by all schools. Because the tests are general, they will not reflect the skills and achievements that are taught less widely nor those unmeasurable in machine-scorable tests.

Results Not Diagnostic

General achievement batteries are not designed to provide diagnostic information about individuals upon which specific ir tructional plans may be based. The Michigan educational assessment tests, being general achievement batteries, do not provide diagnostic information. Rather, they provide a general indication of a pupil's skill levels. If a particular student's results on the achievement tests suggest a problem, an appropriate diagnostic instrument may provide information useful for planning instruction to correct the problem. Information about diagnostic tests* in each skill area may be found in the series of Mental Measurements Yearbooks prepared by O.K. Buros. However, the administration of full diagnostic batteries to all pupils is generally unnecessary and prohibitively expensive.

Many areas of pupil achievement and development are not included in the educational assessment battery. It must be remembered that the goals of local schools and of the Michigan Department of Education are much broader than those assessed by this battery. Therefore, persons wishing to judge the overall achievement of pupils must look to additional measures for judgment in other areas of pupil development.

---

*A test is considered to be diagnostic if it provides a detailed analysis of strengths and weaknesses in an area. Such an analysis will often suggest possible causes for the deficiencies revealed and indicate possible remedial steps.

SECTION III

INTERPRETATION OF PUPIL SCORES

The purpose of this section is to provide the reader with information which will help him to interpret and use pupil scores. This section contains three parts: a description of the norming group, definitions of the terms necessary for understanding the pupil rosters and labels, and descriptions of the pupil rosters and labels.

Description of the Norming Group

The 1972-73 Michigan Educational Assessment Battery was administered to public school students in the fourth and seventh grades. Pupils in non-graded programs were included if they were identified as fourth or seventh graders or, if not identifiable by grade, were in their fourth or seventh year beyond kindergarten. Also included were shared-time pupils who received instruction in the basic skills in the public school.

Excluded were type A mentally handicapped students and pupils receiving instruction in the content areas of reading, English, and mathematics in special classes for the handicapped. But remedial reading pupils were not excluded, nor were pupils receiving special education services (e.g., hard of hearing, physically handicapped, educable mentally retarded, emotionally disturbed) on an itinerant basis, provided that they received instruction in the regular class program in the areas of reading, English, and mathematics. However, students whose disabilities would prevent them from taking the tests under standard conditions (e.g., blind or deaf pupils) were not included.

Definition of Terms

This part defines the terms which are necessary for understanding the scores recorded on pupil rosters and labels. The defined terms are: mean, standard deviation, standard scores, equated standard scores and percent below.

Mean

A mean score is the average of a set of scores. It is obtained by adding all of the scores and dividing the sum by the number of scores.

Standard Deviation

In addition to establishing the mean of a distribution of scores, it is often useful to know the "spread" of the scores. Two groups of scores could have the same mean but still be quite different. For example, one district might have children whose scores on composite achievement are very similar and have a mean score of fifty. In this district, the "spread" of scores would be small. Another district might have a number of children with high scores and a number of children with low scores and still have a mean score of fifty. In this district, however, the "spread" of scores would be large.

One common way of indicating the "spread" of a set of scores is to calculate the standard deviation. In the familiar, bell-shaped "normal" distribution two-thirds of the scores will fall between one standard deviation above and one standard deviation below the mean. The larger the standard deviation, the larger will be the "spread" or variability among the scores of a distribution. In the example at the top of the page,

the district with similar scores would have a smaller standard deviation than would the district with the mixture of high and low scores.

Standard Scores

Standard scores are derived from the number correct, called raw scores, using the mean and standard deviation. In the Michigan Educational Assessment Program, standard scores were developed each year prior to 1972-73 so that the mean of pupil scores on any assessment test was 50 and the standard deviation was 10, when computed for all pupils at the same grade level. As a result, a pupil with a standard score of 40 on reading would be one standard deviation below the state mean; a pupil with a score of 60 would be one standard deviation above the mean; a pupil with a score of 65 would be one and one-half standard deviations above the mean; and so forth.

Equated Standard Scores

Equated standard scores are obtained by transforming the raw scores on a test into the standard score units previously established for another test--generally into the units of a test which is parallel in difficulty and content. The methods used for transforming raw scores on the 1972-73 assessment tests into the standard score units of the 1969-70 tests will be described in the 1972-73 Technical Report.

Pupil scores on the 1972-73 tests are reported in units equated to the standard scores established for the corresponding 1969-70 assessment tests. As a result these 1972-73 pupil scores are comparable to 1969-70 scores on the corresponding tests. For example, 1972-73 equated word relationships may be compared to 1969-70 standard scores on

vocabulary (the reader should note that the name of the verbal analogies test was changed in 1970-71 to word relationships). Similarly, 1972-73 equated standard scores on reading, mechanics of written English, and mathematics may be compared to 1969-70 standard scores on reading, English expression and mathematics, respectively.

The meaning of score comparability is that equal scores indicate the same level of performance in both years. For example, a reading score of 60 in 1972-73 represents the same level of reading skill as a score of 60 in 1969-70. Allowance has been made in the equating process for the effects of year-to-year differences in the number and difficulty of questions that make up the tests.

Equated standard scores may be used in the same way and generally for the same purpose as standard scores. For most uses pupil scores expressed in equated score units are not distinguishable from the standard scores reported in previous years of the assessment program. One difference will be apparent--the state mean of all pupil equated standard scores may not be exactly equal to 50 as before, nor will the standard deviation be exactly equal to 10 as before. Departures from those values will reflect differences between the groups tested in 1969-70 and in 1972-73. The effects of differences in the tests have been removed by the equating process.

The advantage of equated standard scores is that year-to-year gains in the basic skills achievement of all pupils will be reflected in the results, and not be cancelled out by the annual renorming process. Thus, if the fourth graders in Michigan are taught to a higher reading skill level each year, the equated standard score means will rise; conversely, if their skill levels are lower, the equated standard score means will

drop. Unequated standard score means would remain the same for the state as a whole.

School and district averages computed from equated standard scores will share the comparability of the pupil scores from which they are computed. Thus, school and district means for 1972-73 will be comparable to the school and district means for 1969-70. An equating report to be published in the spring of 1973 will contain tables and instructions for converting scores and means reported in 1970-71 and in 1971-72 into score units equated to the 1969-70 scales. By combining 1969-70 base year means with 1970-71 and 1971-72 means, converted by the local district into equated units, and with the 1972-73 means, which will be reported into equated units, a district will have results that are comparable across all four years of the assessment program.

Percent Below

A percent below corresponding to any given standard score or equated standard score is the percentage of pupils in the norm group who received lower scores. Thus, a percent below score of 75 on word relationships indicates that 75 percent of the pupils in the state received a lower score, and that 25 percent of the pupils in the state received the same or a higher score; a percent below score of 16 would mean that 16 percent of the pupils in the state received a lower score and 84 percent of the pupils received the same or a higher score. And so on.

ROSTER OF SCORES 1972-73

MICHIGAN DEPARTMENT OF EDUCATION - MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM

SCHOOL NAME ABLE ELEMENTARY SCHOOL (3487) DISTRICT NAME MICHVILLE PUBLIC SCHOOLS (01-010) GRADE 4

STUDENT NAME			DATE OF BIRTH		SEX	WORD RELATIONSHIPS			READING			MECHANICS OF WRITTEN ENGLISH				MATH				COMPOSITE ACHIEVEMENT	
LAST	FIRST	MI.	MONTH	YEAR		NUMBER RIGHT / NUMBER OF QUESTIONS	STANDARD SCORE	PERCENT BELOW	NUMBER RIGHT / NUMBER OF QUESTIONS	STANDARD SCORE	PERCENT BELOW	NUMBER RIGHT / NUMBER OF QUESTIONS	STANDARD SCORE	PERCENT BELOW		NUMBER RIGHT / NUMBER OF QUESTIONS	STANDARD SCORE	PERCENT BELOW		STANDARD SCORE	PERCENT BELOW
AABERSON	ALEX	M	07	63	M	25/45	48	47	30/50	49	41	35/55	50	47		20/40	45	27		48	38

Figure 1

Description of Pupil Roster and Pupil Labels

The pupil roster and pupil labels yield exactly the same information in different formats. Therefore, only the pupil roster will be fully described.

Pupil Roster

Figure 1, on page 12 represents a pupil roster. An example from it will now be described. Alex M. Aaberson was born in July, 1963, is a male, and is in fourth grade. He received a raw score of 25 out of a possible 45 questions on the word relationships test. This raw score of 25 translates to a standard score (equated) of 49, and 47 percent of the pupils in the state received a lower score. He answered correctly 30 of 50 questions on the reading test for a standard score of 48, a score which exceeded 41 percent of the pupils in the state. On mechanics of written English, Alex got 35 of 55 questions correct for a standard score of 50 and 47 percent of the students in the state received a lower score. On mathematics, he got 20 of 40 questions correct for a standard score of 45 and 27 percent of the students in the state received a lower score. Finally his composite achievement standard score was 48, and 38 percent of the pupils in the state received a lower score. The remaining names and scores on the roster may be described in a similar manner. A pupil roster is provided for each school in the district which has a fourth and/or seventh grade. All 1972-73 standard scores are in units equated to 1969-70 standard scores.

Pupil Labels

The pupil label is displayed below and contains the same information as the pupil roster.

Figure 2

MICHIGAN EDUCATIONAL ASSESSMENT

NAME			GRADE	SEX	DATE OF BIRTH	
LAST	FIRST	M.I.			MONTH	YEAR
AABERSON	ALEX	M	4	M	07	63

1972-73

	WORD RELATIONSHIP	READING	MECHANICS OF WRITTEN ENG	MATHEMATICS	COMPOSITE ACHIEVEMENT
NUMBER RIGHT / QUESTIONS IN TEST	25/45	30/50	35/55	20/40	
STANDARD SCORE	48	49	50	45	48
PERCENT BELOW	47	41	47	27	38

MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM

SECTION IV

STATISTICAL CHARACTERISTICS OF THE 1972-73 MICHIGAN EDUCATIONAL ASSESSMENT BATTERY

The purpose of this section is to provide information concerning the statistical properties of the tests. The section is divided into three major parts. The first part will define the statistical terms which are necessary for the readers' understanding of test characteristics. The next part will discuss the statistical characteristics of the achievement measures at the fourth grade and seventh grade levels. The final part will evaluate the test characteristics and will indicate the areas in which additional information is needed.

Definition of Terms

This part defines the statistical terms: validity, reliability, difficulty, speededness, and standard error of measurement.

Validity

The validity of a test is an indication of the extent to which it measures what it is intended to measure. The most important type of validity for achievement tests is content validity. Content validity means that a test which claims to measure elementary mathematics, for example, should contain questions in mathematics and that those questions should be appropriate for the grade level for which the test is intended. The content validity of a test is dependent upon the extent to which the questions in the test constitute a representative sample of the

topics that comprise the subject tested.

It is virtually impossible to experimentally determine the content validity of a test of school achievement, or to report content validity as a numerical coefficient. The best available evidence of validity is found in the test itself and in a description of the procedures used to construct it.

Content validity is likely to be achieved if the development of the tests is the joint responsibility of specialists in test construction and specialists in the skills to be taught. The development of the achievement tests in the Michigan assessment battery has been described in Section I and could be expected to produce tests of high content validity.

Concurrent validity is another type of validity. It is usually reported as a correlation coefficient which indicates the extent to which two tests measure the same subject or characteristic.

Reliability

The reliability of a test is its consistency or stability. A test is reliable if it measures consistently whatever it measures; the most reliable tests yield relatively precise results for each student, and a student would receive roughly the same score on such a test if he could take it more than once under the same conditions. The reliability of a test is reported as a coefficient--that is, a two-place decimal figure. A reliability coefficient can range between .00 and 1.00. A test increases in reliability as the coefficient increases in value. An internal consistency method (Kuder-Richardson Formula #20) was used in estimating the reliability of the basic skills tests in the Michigan educational assessment

battery. This method employs information about the length of the test and the extent to which the questions in the test contribute mutually confirming or consistent information.

Difficulty

The difficulty of a test is an indication of how well suited it is to the ability of the group being tested. A test of middle difficulty is appropriate when the group being tested is heterogeneous, such as the group being tested in the Michigan Educational Assessment Program. A test is of middle difficulty when the mean score is near the point midway between a perfect score and the expected chance score. If the test includes 50 items a perfect score would, of course, be 50. The expected chance score, when the 50 items each offer four alternative answers, is 1/4 of 50 or 12.5. Hence the ideal mean (midway between chance and perfect) is 31.25. Expressed as a percentage, this would be 62.5%. In the Michigan educational assessment battery, a percentage significantly higher than 62.5% would indicate an easy test; a percentage significantly lower than 62.5% would indicate a difficult one.

Speededness

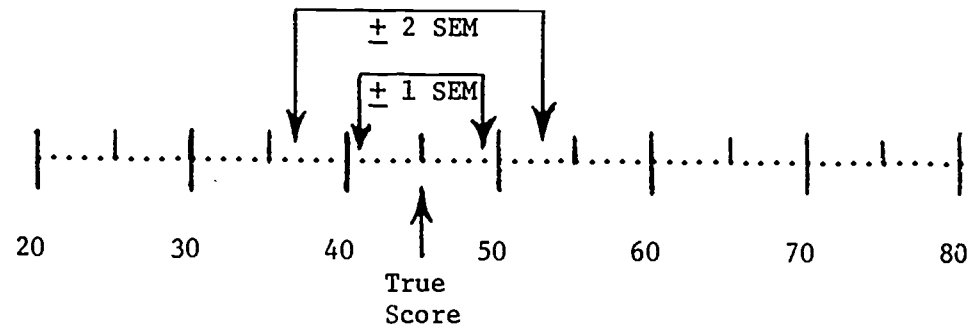
Speededness is a measure of the extent to which test performance is affected by the time limit placed on the test. The criterion used in judging the speededness of a test is two-fold: the proportion of students who answered (1) the last question and (2) the question that is three-quarters of the way through the test. This does not mean that the students answered all the questions up to these points; it means that they reached these particular items. If 80 percent of the students complete the last item and virtually all the students complete 75 percent

of the items, the test is usually judged to be unspeeded. It should be noted that it is desirable for tests in the Michigan Educational Assessment Program to be unspeeded.

Standard Error of Measurement

The standard error of measurement yields an index of the error associated with a test score. It may be used to form an error band around a score extending from a given number of standard errors of measurement above the score to the same number of standard errors of measurement below the score. Figure 3 illustrates the error band around the true score for a standard error of measurement of 4 score units.

Figure 3



If many individuals with the same true score of 45 were tested about 68 percent of them would receive scores not more than one standard error of measurement (SEM) from 45, and about 95 percent would receive scores not more than two standard errors of measurement (SEM) from 45.

Statistical Characteristics of the Fourth Grade Basic Skills Battery

A random sample of 1,020 student tests was drawn from about 155,000 of the fourth grade answer sheets received earliest in the 1972-73 Michigan Educational Assessment Program. (About 160,000 fourth grade answer sheets have been received each year.) This sample's responses were used to analyze the characteristics of the fourth grade educational assessment battery.

The Test Analysis Sample

The means and standard deviations for the sample and for the larger group of 158,000 available for comparison are shown in Table 1 along with the number of questions contained in each test. From the table it can be seen that the characteristics of the sample conform closely to the characteristics of the larger group of fourth graders. Therefore the results presented in the following tables may be considered representative of the fourth grade population. Detailed information concerning the distributions of raw and standard scores is reported in the Appendix.

Table 1

STATISTICS ON THE GRADE FOUR 1972-73 EDUCATIONAL ASSESSMENT TESTS FOR THE TEST ANALYSIS SAMPLE AND FOR THE LARGER GROUP OF FOURTH GRADERS

		Sample		Larger Group	
Score	Number of Questions	Mean	Standard Deviation	Mean	Standard Deviation
Word Relationships	45	24.9	9.2	25.4	9.3
Reading	50	30.8	10.4	31.2	10.3
Mechanics of Written English	55	34.0	10.5	34.4	10.6
Mathematics	40	24.5	7.7	24.8	7.8

Reliability

Table 2 presents the reliability coefficients and standard errors of measurement. The reliability coefficients were: .904 for word relationships, .923 for reading, .913 for mechanics of written English, .888 for mathematics and .964 for composite achievement. The standard errors of measurement expressed in raw score units were: 2.9 on word relationships, 2.9 on reading, 3.1 on mechanics of written English, 2.6 on mathematics, and (not defined) on composite achievement. The standard errors of measurement expressed in standard score units were: 3.5 on word relationships, 2.8 on reading, 2.9 on mechanics of written English, 3.2 on mathematics, and 1.7 on composite achievement. The reliability coefficients are adequately high and the standard errors of measurement are adequately low to justify reporting of individual student scores.

Table 2

RELIABILITY COEFFICIENTS AND STANDARD ERRORS OF MEASUREMENT FOR A SAMPLE OF 1,020 FOURTH GRADERS ON THE 1972-73 EDUCATIONAL ASSESSMENT BATTERY

Score	Reliability Coefficient	Standard Error of Measurement	
		Raw Score Units	Standard Score Units
Word Relationships	.904	2.9	3.5
Reading	.923	2.9	2.8
Mechanics of Written English	.913	3.1	2.9
Mathematics	.888	2.6	3.2
Composite Achievement	.964	(not defined)	1.7

Difficulty and Speededness

Table 3 presents the difficulty and speededness indices. The difficulty indices on word relationships, reading, mechanics of written English, and mathematics are within an acceptable range of middle difficulty. All of the tests are unspeeded.

Table 3

DIFFICULTY AND SPEEDEDNESS FOR A SAMPLE OF 1,020 FOURTH GRADERS ON THE 1972-73 EDUCATIONAL ASSESSMENT BATTERY

Score	Average Item Difficulty (Percent passing)	Speededness: Percent Reaching Last Item	Speededness: Percent Completing 75% of test
Word Relationships	55.3	83.3	93.6
Reading	61.5	82.1	94.1
Mechanics of Written English	61.9	*	*
A. Spelling	60.0	93.3	97.2
B1. Effectiveness of Written Expression	61.1	89.9	96.1
B2. Written Usage	69.2	96.4	98.8
C. Punctuation and Capitalization	56.5	94.6	98.5
Mathematics	61.2	89.6	97.0

* - not applicable

Statistical Characteristics of the Seventh Grade Basic Skills Battery

A representative sample of 1,110 student tests was drawn from the first 159,000 seventh grade answer sheets received in the 1972-73 Michigan Educational Assessment Program. (About 160,000 seventh grade answer sheets have been received each year.) The sample's responses were used to analyze the characteristics of the seventh grade educational assessment battery.

The Test Analysis Sample

The means and standard deviations for the sample and for the larger group of 161,000 available for comparison are shown in Table 4 along with the number of questions in each test. From the table it can be seen that the characteristics of the sample conform closely to the characteristics of the larger group of seventh graders. Although some of the mean differences are statistically significant, the results presented in the following tables for all practical purposes may be considered representative of the seventh grade population. Detailed information concerning the distribution of raw and standard scores is reported in the Appendix.

Table 4

STATISTICS ON THE GRADE SEVEN 1972-73 EDUCATIONAL ASSESSMENT TESTS FOR THE TEST ANALYSIS SAMPLE AND FOR THE LARGER GROUP OF SEVENTH GRADERS

		Sample		Larger Group	
Score	Number of Questions	Mean	Standard Deviation	Mean	Standard Deviation
Word Relationships	38	22.6	6.9	22.0	7.0
Reading	60	41.3	11.3	40.5	11.6
Mechanics of Written English	64	43.2	10.7	42.4	11.0
Mathematics	40	24.5	7.8	24.1	7.8

Reliability

Table 5 presents the reliability coefficients and standard errors of measurement. The reliability coefficients were: .861 for word relationships, .928 for reading, .912 for mechanics of written English, .889 for mathematics and .964 for composite achievement. The standard errors of measurement expressed in raw score units were: 2.6 on word relationships, 3.0 on reading, 3.2 on mechanics of written English, 2.6 on mathematics, and (not defined) on composite achievement. The standard errors of measurement expressed in standard score units were: 3.8 on word relationships, 2.4 on reading, 2.9 on mechanics of written English, 3.1 on mathematics, and 1.7 on composite achievement. The reliability coefficients are adequately high and the standard errors of measurement are adequately low to justify reporting of individual student scores.

Table 5

RELIABILITY COEFFICIENTS AND STANDARD ERRORS OF MEASUREMENT FOR A SAMPLE OF 1,110 SEVENTH GRADERS ON THE 1972-73 EDUCATIONAL ASSESSMENT BATTERY

		Standard Error of Measurement	
Score	Reliability Coefficient	Raw Score Units	Standard Score Units
Word Relationships	.861	2.6	3.8
Reading	.928	3.0	2.4
Mechanics of Written English	.912	3.2	2.9
Mathematics	.889	2.6	3.1
Composite Achievement	.964	(not defined)	1.7

Difficulty and Speededness

Table 6 presents the difficulty and speededness indices. The difficulty indices on word relationships, reading, mechanics of written English, and mathematics are within an acceptable range of middle difficulty. All of the tests are unspeeded.

Table 6

DIFFICULTY AND SPEEDEDNESS FOR A SAMPLE OF 1,110 SEVENTH GRADERS ON THE 1972-73 EDUCATIONAL ASSESSMENT BATTERY

Score	Average Item Difficulty (Percent passing)	Speededness: Percent Reaching Last Item	Speededness: Percent Completing 75% of test
Word Relationships	59.5	92.4	98.1
Reading	68.8	*	*
A. Synonyms	63.9	92.3	97.7
B. Comprehension	70.4	91.8	98.6
Mechanics of Written English	67.5	*	*
A. Spelling	64.5	94.7	98.2
B1. Effectiveness of Written Expression	76.1	96.1	98.7
B2. Written Usage	74.5	98.3	98.8
C. Punctuation and Capitalization	55.1	95.9	98.7
Mathematics	61.2	94.5	98.6

* - not applicable

Summary

A perusal of tables 1 through 6 indicates the statistical strength of the achievement batteries. However, certain types of statistical data are presently not available in the Michigan Educational Assessment Program.

The Michigan Educational Assessment Program has provided reasonable content validity in the achievement battery. Programs are currently under way to make the content of the assessment battery even more relevant to Michigan's curricula. Concurrent validity information is available upon request.

APPENDIX

Table of Word Relationships Raw Scores,
Equated Standard Scores and Percents Below

	Grade 4			Grade 7		
Raw Score	Equated Standard Score	Number of Pupils	Percent Below	Equated Standard Score	Number of Pupils	Percent Below
45	72	254	99.8			
44	71	800	99.3			
43	70	1430	98.4			
42	69	2182	97.1			
41	67	2819	95.3			
40	66	3373	93.1			
39	65	3751	90.8			
38	64	4137	88.2	72	86	99.9
37	63	4345	85.4	71	323	99.7
36	61	4364	82.7	69	900	99.2
35	60	4498	79.8	68	1735	98.1
34	59	4637	76.9	66	2787	96.4
33	58	4718	73.9	65	3768	94.1
32	57	4795	70.9	64	4737	91.1
31	55	4891	67.8	62	5714	87.6
30	54	5077	64.6	61	6429	83.6
29	53	4986	61.5	59	6810	79.4
28	52	5400	58.1	58	7364	74.8
27	51	5500	54.6	56	7475	70.2
26	49	5626	51.0	55	7766	65.4
25	48	5794	47.4	52	7840	60.5
24	47	5793	43.7	52	7760	55.7
23	46	5674	40.2	50	7838	50.9
22	45	5745	36.5	49	7429	46.2
21	43	5735	32.9	47	7344	41.7
20	42	5613	29.4	46	7223	37.2
19	41	5330	26.0	44	6763	33.0
18	40	5175	22.7	43	6787	28.8
17	38	4930	19.6	41	6538	24.8
16	37	4719	16.6	40	6510	20.7
15	36	4459	13.8	38	6353	16.8
14	35	4249	11.2	37	5812	13.2
13	34	3829	8.7	36	5257	10.0
12	32	3443	6.6	34	4504	7.2
11	31	3000	4.7	33	3589	4.9
10	30	2425	3.1	31	2746	3.2
9	29	1787	2.0	30	1991	2.0
8	28	1228	1.2	28	1336	1.2
7	26	816	0.7	27	849	0.6
6	25	531	0.4	25	510	0.3
5	24	275	0.2	24	272	0.2
4	23	163	0.1	22	148	0.1
3	22	95	0.1	21	68	0.0
2	20	51	0.0	19	34	0.0
1	19	27	0.0	18	11	0.0
0	18	8	0.0	16	3	0.0

Table of Reading Raw Scores, Equated Standard Scores and Percents Below

	Grade 4			Grade 7		
Score	Equated Standard Score	Number of Pupils	Percent Below	Equated Standard Score	Number of Pupils	Percent Below
60				65	807	99.5
59				64	1755	98.4
58				64	2877	96.6
57				63	3616	94.4
56				62	4182	91.8
55				61	4501	89.0
54				60	4952	86.0
53				60	5000	82.9
52				59	5054	79.7
51				58	5100	76.6
50	68	387	99.8	57	5251	73.3
49	67	1064	99.1	56	5207	70.1
48	66	1759	98.0	56	5139	66.9
47	65	2528	96.4	55	5118	63.8
46	64	3393	94.2	54	5065	60.6
45	63	4056	91.7	53	4935	57.6
44	62	4706	88.7	52	4873	54.6
43	61	5163	85.5	52	4900	51.5
42	60	5542	82.0	51	4837	48.5
41	59	5590	78.4	50	4456	45.8
40	58	5937	74.7	49	4550	43.0
39	57	5891	71.0	48	4393	40.3
38	56	5924	67.2	48	4276	37.6
37	56	5868	63.5	47	4196	35.0
36	55	5655	60.0	46	4004	32.5
35	54	5416	56.6	45	4016	30.0
34	53	5336	53.2	44	3758	27.7
33	52	5390	49.8	43	3526	25.5
32	51	4991	46.6	43	3595	23.3
31	50	4855	43.6	42	3366	21.2
30	49	4577	40.7	41	3153	19.3
29	48	4323	38.0	40	3005	17.4
28	47	4218	35.3	39	2867	15.7
27	46	4110	32.7	39	2673	14.0
26	45	3905	30.3	38	2487	12.5
25	44	3818	27.8	37	2344	11.0
24	43	3702	25.5	36	2227	9.6
23	42	3693	23.2	35	2008	8.4
22	41	3362	21.2	35	1876	7.2
21	40	3380	18.9	34	1771	6.1
20	39	3336	16.8	33	1543	5.2
19	38	3275	14.8	32	1400	4.3
18	37	2984	12.9	31	1333	3.5
17	36	3081	10.9	31	1150	2.8
16	35	2899	9.1	30	975	2.2
15	34	2807	7.3	29	859	1.6
14	33	2601	5.7	28	730	1.2
13	32	2240	4.3	27	548	0.9
12	31	1928	3.1	27	430	0.6
11	30	1454	2.2	26	309	0.4
10	29	1219	1.4	25	249	0.2
9	28	811	0.9	24	135	0.2
8	27	566	0.5	23	79	0.1
7	26	324	0.3	23	67	0.1
6	25	198	0.2	22	51	0.0
5	24	121	0.1	21	21	0.0
4	23	89	0.1	20	17	0.0
3	23	50	0.0	19	11	0.0
2	22	27	0.0	18	8	0.0
1	21	9	0.0	18	6	0.0
0	20	1	0.0	17	1	0.0

Table of Mechanics of Written English Raw Scores, Equated Standard Scores and Percents Below

	Grade 4			Grade 7		
Raw Score	Equated Standard Score	Number of Pupils	Percent Below	Equated Standard Score	Number of Pupils	Percent Below
64				69	113	99.9
63				68	356	99.7
62				67	725	99.3
61				66	1193	98.5
60				65	1712	97.5
59				64	2280	96.0
58				63	2813	94.3
57				62	3327	92.2
56				61	3977	89.8
55	69	183	99.9	60	4327	87.1
54	68	558	99.5	60	4824	84.1
53	67	1044	98.9	59	5246	80.9
52	66	1698	97.8	58	5514	77.4
51	65	2353	96.3	57	5842	73.8
50	64	2929	94.5	56	5783	70.2
49	63	3558	92.2	55	5948	66.6
48	62	4052	89.7	54	6033	62.8
47	61	4546	86.8	53	5987	59.1
46	60	4816	83.8	52	5942	55.4
45	60	5096	80.5	51	5802	51.8
44	59	5270	77.2	51	5702	48.3
43	58	5439	73.8	50	5426	45.0
42	57	5503	70.3	49	5380	41.6
41	56	5476	66.8	48	4995	38.5
40	55	5558	63.3	47	4974	35.4
39	54	5347	60.0	46	4524	32.6
38	53	5392	56.6	45	4392	29.9
37	52	5326	53.2	44	4034	27.4
36	51	5133	50.0	43	3713	25.1
35	50	4940	46.8	42	3566	22.9
34	49	4831	43.8	42	3296	20.9
33	48	4652	40.8	41	3068	19.0
32	47	4657	37.9	40	2891	17.2
31	46	4523	35.1	39	2723	15.5
30	45	4326	32.3	38	2519	13.9
29	44	4135	29.7	37	2388	12.5
28	43	4084	27.1	36	2136	11.1
27	43	3808	24.7	35	2038	9.9
26	42	3695	22.4	34	1868	8.7
25	41	3442	20.2	33	1731	7.6
24	40	3352	18.1	33	1641	6.6
23	39	3155	16.1	32	1476	5.7
22	38	3038	14.2	31	1436	4.8
21	37	2803	12.4	30	1328	4.0
20	36	2675	10.7	29	1176	3.3
19	35	2524	9.1	28	960	2.7
18	34	2387	7.6	27	853	2.1
17	33	2199	6.2	26	770	1.7
16	32	2002	5.0	25	664	1.3
15	31	1750	3.9	24	521	0.9
14	30	1467	3.0	24	422	0.7
13	29	1308	2.1	23	285	0.5
12	28	1053	1.5	22	242	0.3
11	27	784	1.0	21	152	0.3
10	27	586	0.6	20	115	0.2
9	26	380	0.4	19	76	0.1
8	25	214	0.2	18	60	0.1
7	24	152	0.1	17	33	0.1
6	23	78	0.1	16	34	0.1
5	22	52	0.0	15	28	0.0
4	21	33	0.0	15	25	0.0
3	20	20	0.0	14	15	0.0
2	19	13	0.0	13	13	0.0
1	18	7	0.0	12	9	0.0
0	17	1	0.0	11	2	0.0

Table of Mathematics Raw Scores Equated Standard Scores and Percents Below

Grade 4				Grade 7		
Raw Score	Equated Standard Score	Number of Pupils	Percent Below	Equated Standard Score	Number of Pupils	Percent Below
40	70	514	99.7	69	843	99.5
39	68	1522	98.7	68	1723	98.4
38	67	2547	97.1	67	2690	96.7
37	66	3501	94.9	66	3369	94.6
36	65	4196	92.2	65	3988	92.2
35	64	4932	89.1	63	4548	89.4
34	62	5729	85.5	62	4986	86.3
33	61	6223	81.6	61	5327	83.0
32	60	6543	77.4	60	5676	79.4
31	59	6826	73.1	59	5877	75.8
30	57	6912	68.8	57	6172	72.0
29	56	7050	64.3	56	6360	68.0
28	55	7082	59.8	55	6391	64.0
27	54	7141	55.3	54	6492	60.0
26	52	6954	50.9	53	6682	55.9
25	51	6962	46.5	51	6696	51.7
24	50	6722	42.3	50	6682	47.6
23	49	6541	38.1	49	6771	43.4
22	47	6357	34.1	48	6676	39.2
21	46	6154	30.2	47	6678	35.1
20	45	5602	26.7	45	6562	31.0
19	44	5373	23.3	44	6357	27.1
18	43	5020	20.1	43	6183	23.2
17	41	4733	17.1	42	6069	19.5
16	40	4377	14.4	41	5536	16.0
15	39	4189	11.7	39	5194	12.8
14	38	3729	9.4	38	4645	9.9
13	36	3470	7.2	37	4092	7.4
12	35	3032	5.3	36	3550	5.2
11	34	2479	3.7	35	2750	3.5
10	33	2029	2.4	33	2062	2.2
9	31	1493	1.5	32	1450	1.3
8	30	969	0.9	31	967	0.7
7	29	621	0.5	30	539	0.4
6	28	359	0.2	29	309	0.2
5	27	193	0.1	27	135	0.1
4	25	100	0.1	26	70	0.0
3	24	52	0.0	25	38	0.0
2	23	16	0.0	24	9	0.0
1	22	12	0.0	23	4	0.0
0	20	2	0.0	21	1	0.0

Table of Composite Achievement Equated
Standard Scores and Percents Below

Equated Standard Score	Grade 4 Number of Pupils	Grade 4 Percent Below	Grade 7 Number of Pupils	Grade 7 Percent Below
69	24	100.0		
68	176	99.9	6	100.0
67	659	99.5	228	99.9
66	1319	98.6	782	99.4
65	2240	97.2	1573	98.4
64	3058	95.3	2419	96.9
63	3988	92.7	3383	94.8
62	4627	89.8	4141	92.2
61	5327	86.4	4823	89.2
60	5587	82.9	5334	85.8
59	6018	79.1	5787	82.2
58	6285	75.1	5968	78.5
57	6318	71.1	6246	74.6
56	6383	67.1	6637	70.5
55	6178	63.1	6638	66.3
54	6318	59.1	6430	62.3
53	6111	55.3	6722	58.1
52	5894	51.5	6621	54.0
51	5870	47.8	6598	49.8
50	5672	44.2	6332	45.9
49	5364	40.8	6342	41.9
48	4972	37.7	5954	38.2
47	4935	34.5	5623	34.7
46	4808	31.5	5624	31.2
45	4676	28.5	5139	28.0
44	4399	25.8	4898	24.9
43	4246	23.1	4701	22.0
42	4031	20.5	4305	19.3
41	3794	18.1	4051	16.8
40	3666	15.8	3795	14.4
39	3440	13.6	3447	12.3
38	3332	11.5	3167	10.3
37	3156	9.5	2741	8.6
36	2992	7.6	2483	7.0
35	2715	5.9	2304	5.6
34	2553	4.3	2122	4.3
33	2166	2.9	1692	3.2
32	1739	1.8	1506	2.3
31	1253	1.0	1218	1.5
30	790	0.5	971	0.9
29	377	0.3	631	0.5
28	200	0.1	409	0.3
27	86	0.1	219	0.1
26	50	0.0	96	0.1
25	28	0.0	47	0.0
24	15	0.0	26	0.0
23	10	0.0	9	0.0
22	8	0.0	3	0.0
Below			2	0.0
			1	0.0

Raw and Standard Score Means and Standard Deviations
For the Groups on which Pupil Norms were Based

		Grade 4		Grade 7	
Test	Statistic	Raw Score	Standard Score	Raw Score	Standard Score
Word Relationships	Mean	25.4	48.6	22.0	48.8
	Standard Deviation	9.3	11.2	7.0	10.3
Reading	Mean	31.2	49.9	40.5	49.6
	Standard Deviation	10.3	10.0	11.6	9.4
Mechanics of Written English	Mean	34.4	49.5	42.5	49.2
	Standard Deviation	10.6	10.0	11.0	9.9
Mathematics	Mean	24.8	50.9	24.1	50.3
	Standard Deviation	7.8	9.6	7.8	9.3
Composite Achievement	Mean	*	50.1	*	49.7
	Standard Deviation	*	9.1	*	8.7
	Number of Cases		157,854		160,200

* - not applicable