DOCUMENT RESUME

ED 064 286 TM 001 402

TITLE	Technical Report of Selected Aspects of the 1969-70 Michigan Educational Assessment Program.
INSTITUTION	Educational Testing Service, Princeton, N.J.; Michigan State Dept. of Education, Lansing.
PUB DATE	Aug 71
NOTE	49p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Achievement Tests; Attitude Tests; *Educational Testing; *Measurement Instruments; *Questionnaires; Statistical Analysis; *Test Construction; *Test Interpretation; Test Reliability; Test Validity
IDENTIFIERS	*Michigan

ABSTRACT

This report describes the development of the 1969-70 Michigan Educational Assessment measures used in assessing the levels and distribution of educational performance for Michigan's districts, schools, and pupils. The report has four sections. The first section contains a brief description of the 1969-70 assessment program, including a statement of purposes and a summary of procedures. The second section explains in detail the construction of both the achievement portion of the battery and the pupil questionnaire. Section three describes the development of the scaled scores arising from the statistical treatment of responses to the items in the various parts of the battery. Nine such scores were developed. Section four deals with the characteristics of the battery components. (Author/DB)

ED 064286

U.S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TECHNICAL REPORT

of Selected Aspects of the

1969-70 MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM

TM 001 402

Prepared by Educational Testing Service
Princeton, New Jersey

and

Michigan Department of Education
Lansing, Michigan

Published by Michigan Department of Education

August, 1971

FOREWORD

This report describes the development of the 1969-70 Michigan Educational Assessment measures which were used in assessing the levels and distribution of educational performance for Michigan's districts, schools, and pupils. It was prepared jointly by Educational Testing Service and the Michigan Department of Education, and is published by the Michigan Department of Education.

The contents of this document are intended to fulfill the requirements of the Standards for Educational and Psychological Tests and Manuals. Therefore, it is technical in nature and is intended for persons interested in the specifications and psychometric properties of the tests.

Thanks are due to a large number of individuals and groups for making the Michigan Educational Assessment Program a reality: The State Board of Education for proposing it, the Governor and Legislature for actively supporting it, and Michigan educators for assisting with it. The program was designed and administered by Research, Evaluation, and Assessment Services, Michigan Department of Education, with the assistance of Educational Testing Service, and counsel of several ad hoc advisory groups.

John W. Porter
Superintendent of
Public Instruction

TECHNICAL REPORT OF SELECTED ASPECTS OF THE 1969-70 MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM

Introduction

An assessment program is of value only when those who are to develop a course of action as a result of its implementation fully understand the construction and scoring of the instruments used in the program and the extent to which the battery has achieved its purposes. Consequently, this report attempts to answer three questions: How was the battery constructed? How were scores derived? How well did the battery succeed in achieving the intended goals?

This report has four sections. The first section contains a brief description of the 1969-70 assessment program, including a statement of purposes and a summary of procedures.

The second section explains in detail the construction of both the achievement portion of the battery and the pupil questionnaire. Construction of the four achievement tests is described in terms of content, item-writing, test assembly methods, and review procedures. In addition, sources of questionnaire items and methods of selecting those questions to be included are described.

Section three describes the development of the scaled scores arising from the statistical treatment of responses to the items in the various parts of the battery. Nine such scores were developed. Five are concerned with achievement (vocabulary, reading, English expression, mathematics,

and composite); four are concerned with student perceptions of educational correlates (socio-economic status, the importance of school achievement, self-perceptions, and attitude toward school).

Section four deals with the characteristics of the battery components. Describing the characteristics of the achievement tests and the questionnaire, this section enables the reader to evaluate the aspects of the program that are dependent upon student responses.

Description of the Assessment Program

The plans and procedures of the 1969-70 Michigan Assessment Program were designed to determine the levels of educational performance in basic skills not only for the state as a whole but also for various subdivisions within the state. The program also called for the development and institution of means of determining the status and distribution of several educational correlates within the state and within those same subdivisions. Implementation of the plans called for the development of achievement and attitude measures to be administered to students and for the creation of a data-file containing information regarding school district financial and human resources. The achievement and attitude measures were developed by Educational Testing Service in consultation with the State Department of Education. The information concerning school district resources was furnished by the State Department from existing records.

The assessment battery was administered to fourth and seventh grade pupils in January of 1970. Table 1 shows the number of school districts,

schools, and students in the fourth and seventh grades that took part in the 1969-70 assessment program.

TABLE 1

Participants in the 1969-70 Assessment Program

Grade	School Districts* N	Schools* N	Pupils N
4	585	2,492	158,713
7	566	909	159,407

*With enrollment of at least 5 students.

Construction of the 1969-70 Assessment Battery

Achievement Battery

Before the process of constructing the 1969-70 assessment battery is described, regrettable fact must be brought to the reader's attention. Because of the lateness of the funding authorization, the period of time allowed for test construction was altogether too brief. Although every effort was made to involve as many Michigan educators as possible in the process, the time restriction did not allow for as widespread participation of this group as would have been desirable.

Several processes in the development of tests in the achievement areas covered by the Michigan Assessment Program were the same regardless of the subject under consideration. These processes are described below and are followed by a detailed description of the tests in each area: vocabulary, reading, English expression, and mathematics.

The first step in the construction of any test is the development of specifications. Test specifications have been likened to a skeleton. They are the bones or the framework that determine the shape and structure of the finished product. Test specifications set forth in detail the content, coverage, and emphasis of the test. For example, test specifications for a reading test would include the type of material with which the student must work (words, sentences, paragraphs), the type of knowledge called for (synonyms, associations, factual, interpretation), and the number of questions involving each type of material and each type of knowledge to be included in the test. So that the achievement tests used in the Michigan Assessment Program might accurately reflect the objectives of education in the state, the professional staff of each of the several departments of the Test Development Division of Educational Testing Service wrote a set of preliminary test specifications based on texts in use in Michigan schools.

Text-based specifications require a careful examination of the textbooks that are in use and consultation with educators who are responsible for instruction. The purposes of such procedures are to bring into sharper focus the specific objectives in each of the subject-matter areas under consideration, to determine the relative emphasis given to each topic within any given curriculum, and to point up the extent to which any innovative programs (e.g., modern mathematics) are being adopted throughout the state.

The preliminary test specifications so developed were reviewed by the ad hoc Assessment Battery Specifications Committee appointed by the

Michigan Department of Education. Members of this committee suggested several revisions for the improvement of the content and coverage of the proposed tests. Educational Testing Service revised the test specifications in accordance with these suggestions. Content specifications for the reading, English expression, and mathematics tests are presented in Appendix A.

The second standard step in test development is the writing and assembly of items or questions of the type and in the quantity called for by the test specifications. The ETS procedure for writing items is as follows: An individual is assigned the task of producing a certain number of items along stringent lines carefully laid out in the specifications. After writing the items, he submits them to at least two other staff members for review.

The review process for both the test questions and the tests themselves involves more than a mere reading of the material. Some of the considerations borne in mind in both instances are:

1. Does each item have one and only one correct answer?
2. Is the language simple, direct, and free of ambiguity?
3. Is each item independent of other items? (It is not fair for the student to get two items right (or two wrong) because the same information is needed to answer both items.)
4. An item should be difficult because it requires sophisticated reasoning, not because it tests obscure or esoteric subject matter.

5. Do the items cover a range of difficulty? Are they ordered from easy to hard?
6. Does the test meet the content specifications?
7. Is the content of any item "unfair" or "emotionally offensive" to any segment of the population?

Continuing the question-writing phase of test development, the items are returned to the individual who wrote them. He makes whatever changes are necessary along the lines suggested by reviewers.

A different staff member is assigned the task of assembling the test, which is the third step in test development. He receives and reviews each item, making changes, rejecting items considered unsuitable, and putting finishing touches on the items to be used. After the test is assembled, he submits the test for a Test Specialist review. Here the test as a whole is evaluated and each individual item again subjected to scrutiny.

The test is then reviewed in turn by the department chairman, the coordinator of the program for which the test is being used, and the editorial staff. After each review, changes are made in line with the suggestions of the reviewer. Upon completion of the process within the Test Development department, the test is submitted for review by a committee of those responsible for using the test; i.e., the Assessment Battery Specifications Committee in this case. The committee evaluates the test and makes suggestions for changes that are carried out by the test developers.

Test Content: Vocabulary

The vocabulary section of the Michigan Assessment of Basic Skills Test for 1969-70 consists of 50 analogy questions. Each question begins with two words that go together in a certain way. Under them, there are four other pairs of words lettered A, B, C, and D. The student is asked to find the lettered pair of words that go together in the same way as the first pair of words. For example: CALF:COW::

(A) puppy:dog
(B) nest:bird
(C) horse:bull
(D) shell:turtle

The first pair of words go together in this way: A calf is a young cow. The only lettered pair of words that go together in this way is puppy:dog; that is, a puppy is a young dog.

Items of this type use words that are recognizable by the population being tested. The student is asked, however, to be able to think in somewhat greater depth than simple recognition; he must be able to tell in what way a word is similar in meaning to, or different in meaning from, other words.

To apply learning gained in one situation to other similar situations, one must be able to see in what way similarities exist. One may learn that two apples and two apples make four apples, but to use this knowledge he must be able to see that two pears (or two of anything else) have certain similarities to two apples although apples are not pears. It is this kind of learning that most educators strive for rather than rote memorization of specific bits of information.

Test Content: Reading

The reading section of the Michigan Basic Skills Test contains ten word-association items, ten sentence items, and twenty questions based on reading passages. The word-association items indicate the students' ability to read and understand the meaning of individual words. The sentence items measure both the ability to read and understand the meaning of a sentence and to derive the meaning of a word in the context of a sentence. The reading passages measure the students' ability to read a short passage and understand and interpret facts.

In the word-association test, the student is asked to read a word and then choose another word or phrase that goes best with it. The word that goes "best" with the first word is one of three types. It may be a word that means almost the same thing as the first word (cheerful:happy); it may be a phrase that illustrates an emotion or action appropriate to the first word (cry:boo-hoo); or it may be one or more words that describe something commonly associated with the first word (ask:May I?).

The sentence items are of two types. In one type, inference, the student is asked to pick a word which describes what has happened in the sentence. For example: "The car would not move because the wheels sank down into the mud." "The car was _____?" "Stuck" is the answer which the student must pick from three other options.

A second type of sentence, comprehension, asks that a student understand a word contained in a sentence, the meaning of which is largely explained by the rest of the sentence. The following is an example.

Sally was <u>miserable</u> because she had the measles and could not go to the party.

<u>Miserable</u> means

(A) happy
(B) sad
(C) calm
(D) smiling

The section with reading passages contains six short passages, each followed by questions. Approximately two-thirds of the questions have to do with understanding the factual contents of the passage; one-third asks for something that was either implied or could be inferred from the passage. Both types of questions are illustrated in the following example:

Many bodies of water are called seas, but the term "the Seven Seas" refers to the largest of all, the oceans: the Arctic and Antarctic, the North and South Atlantic, the North and South Pacific, and the Indian.

The Indian Ocean takes its name from the country along its northern shore, India, whose name in turn came from "Sindhu," meaning river. India was called, in olden days, "the land beyond the river."

In the Middle Ages, the Atlantic Ocean was called "the sea of darkness." Men knew its eastern shore, but beyond there lay darkness, the darkness which came from ignorance and fear.

The Pacific, the largest ocean of all, was first seen by a Spanish explorer who named it "Mar del Sud" -- the Sea of the South. A later explorer named it "Pacific," meaning peaceful, because of the serenity of its surface.

(1) Which of these is NOT one of the Seven Seas?

(A) Atlantic (B) Indian (C) Baltic (D) Arctic

(2) Why is the Pacific Ocean called "Pacific?"

(A) It is in the North.
(B) It divides the continents.
(C) It is large and dark.
(D) It looks serene.

Test Content: English Expression

There are four parts to the English Expression Test: Spelling, Effectiveness of Expression, Grammar and Usage, and Capitalization. As is indicated below, the number of questions varies with the grade level for which the test is intended.

A. Spelling (Grade 4, 15 items; Grade 7, 20 items)

The questions for this part present four words, one of which may be misspelled. Sources of spelling error include such possibilities as misunderstanding of the rules for word formation, reversing of letters, and spelling by sound alone.

Example: (A) monney
(B) funny
(C) sunny
(D) no error

B. Effectiveness of Expression (Grades 4 and 7, 10 items)

Questions testing effectiveness of expression ask the student to demonstrate his sensitivity to language by selecting the wording which, in addition to being grammatically correct, is best in sentence structure and word order and is most precise in idiom and diction.

Example: (A) Her car is red like this here one.
(B) Her car is red like this one.
(C) This car is red like the one of hers.
(D) Like her car, this here one is red.

C. Grammar and Usage (Grade 4, 10 items; Grade 7, 13 items)

The questions for this part assess the students' ability to work with such matters of appropriate sentence construction as subject-verb agreement, adjective forms (including a and an), and, for grade seven only, agreement between a pronoun and its antecedent.

Example: Before he ran (A) for office, Mr. Lloyd had spoke (B) out against (C) high taxes. No error. (D)

D. Punctuation and Capitalization (Grade 4, 10 items; Grade 7, 12 items)

These questions assess the students' mastery of such common matters as the use of capital letters to begin proper names and to signify the first-person "I", the use of end-of-sentence marks, and the use of the apostrophe to form possessives and contractions.

Example: (A) My teacher's name is Mrs. Smith.
(B) My teacher's name is Mrs Smith.
(C) My teachers name is Mrs Smith.
(D) My teachers name is mrs. Smith.

The English Expression Test is not a direct measure of a student's ability to write. It is not, and it does not purport to be, a test of a student's ability to think imaginatively, to write "creatively," or to function in his own world of communication. Rather, it is a valid and

reliable predictor of a student's mastery of certain aspects of the language arts curriculum and of his ability to produce the kind of language that is often called "standard written English," the language of most of the books that students read for school.

Test Content: Mathematics

The mathematics section of the Michigan Basic Skills Test contains 30 questions arranged in order of difficulty. Although most of the questions are of middle-difficulty, they range from very easy ones that require little more than recall or recognition to some challenging ones that test reasoning ability. The majority of the items are designed to test comprehension of concepts that have been taught up to the appropriate grade level.

The test does not emphasize computation; however, care was exercised to include as many different number facts as possible in the course of solving problems. The problems are phrased in simple and direct language that would not present a reading problem to most students. Only the more generally recognizable symbols of mathematical notation are used.

For each question, the three incorrect options were determined by considering the common misconceptions students have, the most prevalent computation errors they make, or the answers they would get if they chose the wrong method of solving the problem. The four options are arranged in order from least to greatest value or vice-versa, except in those cases where determining the order is precisely what is being tested. There follows two example items of the type in the mathematics section.

Example: △ - 14 = 61

What number does △ represent?

(A) 47
(B) 75
(C) 79
(D) 84

Example: If N is a whole number greater than 10, which of the following number is least?

(A) N - 1
(B) N - 4
(C) N - 9
(D) N + 3

Construction of Pupil Questionnaire

Plans for the 1969-70 Michigan Assessment Program called for a preliminary investigation of the status of several educational correlates, their distribution, and their possible relationships to basic skills achievement. Consequently, the literature on the correlates of school performance was reviewed. This review pointed up the fact that the questions posed in the study conducted by Coleman and others[1] have proved to be the most useful in those situations where the information desired is to be gotten from students. The responses of the national sample of sixth-grade students to questions from this study were examined, and those that seemed appropriate and that differentiated well among various segments of the population were adopted for use in this program. To the

---

[1]James S. Coleman and others, Equality of Educational Opportunity, Washington, D. C., U. S. Government Printing Office, 1966

questions thus gleaned, several others, written by ETS professional staff members and approved by the Michigan Department of Education, were added.

The resulting questionnaire, labeled "General Information," consists of 26 questions designed to provide measures of such educational correlates as family socio-economic status, attitude toward school, educational aspirations, and self-concepts. Examples of the questions included are:

Does your family have a dictionary?
Do you like school?
Are you planning to go to college?
Do most of your classmates like you?

Development of Scaled Scores

Basic Skill Scores

The score that is obtained by counting the number of questions a student answered correctly is a raw score. Raw scores always refer to a particular test and can never be compared to raw scores on any other test. Identical raw scores obtained on different tests may represent different levels of performance, depending on the length of the tests and their relative difficulties. Therefore, the raw scores on each of the basic skills tests in the Michigan Assessment Battery were transformed to standard or scaled scores with an average of 50 and a standard deviation of 10. The equations for transforming the raw scores on each test at each grade level are tabled in Appendix B.

The transformation of raw scores to standard scores makes them comparable; that is, the same standard scores from different tests represent the same relative degree of achievement. Furthermore, standard scores from different tests may be added together or averaged.

After the standard score scale for each test had been established, a composite achievement score was computed for each student. This score is the average of an individual's scaled scores on reading, English expression, and mathematics. Vocabulary scores were not included in the computation of composite achievement scores.

Scores from the Questionnaire

For purposes of developing scales from the pupil questionnaire, a spaced sample of approximately 7,000 seventh-grade students was drawn from those who participated in the Assessment Program. The statistical technique known as principal components factor analysis was used to create scales from the responses of these students to 25* of the 26 questions included in the General Information section of the Assessment Battery. This means that all of the responses to these questions were analyzed to determine which groups of items tended to measure the same pupil perceptions. For example, questions such as "If you had your choice, would you rather go to a school other than this one?" and "Do you like school?" were indicated as measuring the same student perception: attitude toward school. These were only two of a number of items which contributed to this attitude scale. The statistical analysis then weighted each of the items in terms of its contribution to each scale. The end result of these analyses was the construction of four scales: a measure of socio-economic

---

*The responses to the first item, "Are you a girl or a boy?", were not included in the analysis.

status and three measures of student attitude (importance of school achievement, self-perception, and attitude toward school).

Inasmuch as the Michigan Assessment Battery was designed to measure the achievement and perceptions of groups of students rather than individuals, the scores of all students within a school were averaged to provide an estimate of the levels of educational performance and related factors for each school that participated in the program. These school average scores were then combined for all schools within each of the several districts in the state to indicate the district-wide status of the variables under consideration. These combinations were of two types: percentile distributions and averages.

The percentile distributions contained in Assessment Reports 4, 5, and 6, are based on school average scores combined in various groupings; e.g., districts and SES standings. School averages based on less than five students were not used in the preparation of these percentile distributions.

Similarly, the unweighted district averages or means upon which the percentile distributions in Assessment Reports 4 and 5 are based did not include school averages for those schools where less than five students responded to the questions in the Assessment Battery.

Analysis of Selected Facets of the Assessment Battery

Achievement

Tests may be evaluated by determining several of their characteristics. The test characteristics that are important in evaluating group measures, such as those in this battery, include: validity, reliability, difficulty, and speededness. Each of these terms will be defined before the characteristics of either the fourth or seventh grade battery are discussed.

The validity of a test is an indication of the extent to which it does the job it is intended to do. The most important type of validity for achievement tests is content validity. Content validity means that a test which claims to measure 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 represent a good sample of the topics that are involved in the subject tested.

Content validity is ensured by entrusting the development of the tests to specialists in test construction and in the subject to be covered. The development of the achievement tests in the Michigan Assessment Battery has been described earlier and was designed to produce tests of high content validity. The success of these endeavors may be judged by examining the test copy.

The reliability of a test is an estimation of its consistency. A test is reliable if it consistently does the job it is intended to do; that is, if all of the questions are on the same general topic and if a student would receive roughly the same score if he could take the test 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 widely accepted rule of thumb states that a test possesses an acceptable degree of reliability if the coefficient is .80 or higher. The Kuder-Richardson Formula 20 was used in estimating the reliability of the basic skills tests in the Michigan Assessment Battery.

The difficulty of a test is an indication of how well it is suited for the group being tested. A test that is appropriate for a given group should be of middle difficulty, and the scores obtained by that group should extend throughout the possible range of scores. Middle difficulty of a test consisting of four-choice items implies that the students would know the correct answer to half of the items and would get another quarter of them correct by guessing.

One way of estimating the difficulty of a test is to convert the raw score average to a percentage of the number of items in the test. In these terms, middle difficulty for a four-choice test would be 62.5%. A percentage significantly higher than 62.5% would indicate an easy test; a percentage significantly lower than 62.5% would indicate a hard one.

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.

GRADE 4

A spaced sample of 1,005 student records was drawn from the records of all the fourth grade students who participated in the program. The responses of this sample were used to analyze the characteristics of the fourth grade Basic Skills Battery. The results are given in Table 2.

The tabled reliability coefficients for the fourth grade Basic Skills Battery range from .83 for mathematics to .90 for reading, and the estimated reliability for the composite achievement measure is .95. According to the previously stated rule of thumb, all of these tests are reliable; the reading and composite achievement measures are highly so.

The indications of difficulty presented in Table 2 show that the vocabulary test is difficult for the fourth grade students. The remaining tests--reading, English expression, and mathematics--fall within the acceptable range of middle difficulty. None are easy.

The vocabulary test is speeded; only two-thirds of the students reached the last item and only four-fifths of them got three-quarters of the way through the test. The three remaining tests are clearly unspeeded.

A detailed description of the results of the analysis of the tests in the fourth grade Basic Skills Battery will be found in Appendix B1.

TABLE 2

Characteristics of the Michigan Assessment
Basic Skills Battery, 1969-70

Grade 4, N = 1005

	Vocabulary	Reading	English Expression	Mathematics
Number of items	50	40	45	30
Average*	21.8	24.1	26.3	16.3
Standard deviation*	8.5	8.2	8.1	5.7
Reliability	.87	.90	.88	.83
Difficulty	43.5%	60.3%	58.5%	55.3%
Speededness:				
Proportion reaching last item	66%	88%	93%**	88%
Proportion reaching 3/4 point	86%	97%	98%**	97%

* In raw score units

**Average for separately timed parts

GRADE 7

A spaced sample of the records of 1005 students of all the seventh grade students who participated in the program was drawn for a detailed analysis of the characteristics of the tests in the seventh grade Basic Skills Battery. Table 3 shows the results of these analyses.

The reliability coefficients for the seventh grade tests range from .82 to .88. The estimated reliability of the composite achievement scores is .94. All of these measures show acceptable to high reliability.

All of the tests in the seventh grade battery proved to be of slightly greater than middle difficulty, but all of them fall within the acceptable range. None are easy.

There is no evidence of undue speededness in any of the tests. In each instance, more than 80% of the students finished the tests and more than 90% of them reached the three-quarter point.

A detailed description of the results of the analysis of the tests in the seventh grade Basic Skills Battery will be found in Appendix B2.

TABLE 3

Characteristics of the Michigan Assessment
Basic Skills Battery, 1969-70

Grade 7, N = 1005

	Vocabulary	Reading	English Expression	Mathematics
Number of items	50	40	55	30
Average*	27.3	22.4	31.5	16.0
Standard deviation*	8.1	6.3	8.9	5.9
Reliability	.86	.82	.88	.84
Difficulty	54.6%	55.9%	57.2%	53.3%
Speededness:				
Proportion reaching last item	84%	90%	96%**	90%
Proportion reaching 3/4 point	97%	98%	99%**	98%

* In raw score units

**Average for separately timed parts

APPENDIX A

TEST SPECIFICATIONS

VERBAL ABILITY SPECIFICATIONS: MICHIGAN TESTING PROGRAM 1969-70

Verbal Ability -- (Analogies)

20 minutes

50 items 4 choice questions

Bases for analogies: synonyms, antonyms, part-whole, part-part, cause-effect, object-purpose or action, class-sub-class, place-inhabitant measurement, expression of, associated with, character of, related to.

Subject matter: animals, household, human relationships, transports, tools, weather, aesthetics, sports, measurement, abstract, general, geography.

Vocabulary level in accordance with medium difficulty level for respective grades.

Reading Test Content Specifications
(1969-70)

	Grade 4	Grade 7
I. WORDS	10 items	10 items
A. Synonymous (BEGIN: start)	3	3
B. Associative (DRY: desert)	4	4
C. Illustrative (RELUCTANT: "I'd really rather not.")	3	3
II. SENTENCES	10 items	10 items
A. Inference (External--tests word not in sentence but inferred by sentence)	5	5
B. Comprehension (Internal--tests word in sentence, underlined or boldface--straightforward comprehension)	5	5
III. READING COMPREHENSION		
A. Items	20 items	20 items
1. Factual (explicit)	13	11
2. Interpretive (inference)	7	9
B. Passage Content	6 passages	6 passages
1. Narrative	3	2
2. Science	2	2
3. Social Studies	1	1
4. Humanities		1

English Expression Test Content Specifications (1969-70)

		Grade 4	Grade 7
I.	SPELLING	15 items	20 items
	A. Misunderstanding of rules for word formation	4	9
	B. Misunderstanding of rules for word transformation	2	5
	C. Reversing of letters	2	1
	D. Common mispronunciation	1	1
	E. Spelling by sound alone	4	4
	F. Other	2	
II.	EFFECTIVENESS OF EXPRESSION	10 items	10 items

Questions testing effectiveness of expression ask the student to demonstrate his sensitivity to language by selecting the wording which, in addition to being grammatically correct, is best in sentence structure and word order and is most precise and appropriate in idiom and diction.

III.	GRAMMAR AND USAGE	10 items	13 items
	A. Subject-verb agreement	2	4
	B. Verb forms	2	1
	C. Double negative	1	1
	D. Adjective/adverb confusion	1	1
	E. Adjective forms (including a-an)	1	1
	F. Pronoun forms	1	2
	G. Pronoun-antecedent agreement		1
	No error	2	2
IV.	PUNCTUATION AND CAPITALIZATION	10 sentences	12 sentences
	A. Capital letter		
	1. First word of sentence	4	
	2. Proper names	4	10
	3. First person "I"	1	1
	4. In titles	1	
	5. In quotations		2

English Expression Test Content Specifications (Continued)
(1969-70)

IV. PUNCTUATION AND CAPITALIZATION (Cont'd)	Grade 4	Grade 7
B. Period		
1. End of Sentence	1	1
2. Abbreviation	1	2
C. Question Mark	2	1
D. Comma		
1. Address, date	2	3
2. Direct address	1	1
3. Series	1	
4. In quotations	1	2
5. Apposition		1
6. With exclamation		1
E. Quotation Marks		
1. Direct address, direct quote	1	8
F. Apostrophe		
1. Possessive	2	4
2. Contraction	1	2
H. Semicolon		1

Mathematics Test Content Specifications (1969-70)

		Grade 4	Grade 7
I.	Number and Operations	15 items	10 items
	A. Operations with integers	4	1
	B. Place value	2	1
	C. Properties of integers, divisibility	2	1
	D. Proper fractions	2	1
	E. Decimals and per cents	0	1
	F. Properties of operations (commutative, associative, distributive, closure)	2	2
	G. Estimation	1	0
	H. Special properties of zero and one	2	2
	I. Average	0	1
II.	Geometry and Measurement	3 items	7 items
	A. Units of measure: length, weight, time, temperature, money	1	2
	B. Perimteres and areas of simple polygons	1	1
	C. Scale drawings and maps	0	1
	D. Properties of polygons and the circle	1	2
	E. Angles and intuitive ideas of geometry	0	1
III.	Relations, Functions, Graphs	2 items	4 items
	A. Use of mathematical formula	1	1
	B. Reading and interpreting graphs	1	3
IV.	Logical Thinking	1 item	1 item
	A. Intuitive ideas: counterexample, reasoning	1	1
V.	Mathematical Sentences	2 items	4 items
	A. Equations	1	2
	B. Inequalities	1	2

Mathematics Test Content Specifications (Continued)
(1969-70)

	Grade 4	Grade 7
	7 items	4 items
VI. Applications		
A. Word problems (other than those already listed in one of the categories above)	7	4
Note: At least one-third of the problems could be classified as applications.		
TOTAL	30 items	30 items

APPENDIX B-1

GRADE 4 TEST ANALYSIS

Grade 4

The Michigan Educational Assessment Program became operational during the 1969-70 academic year. Appropriate levels of the test battery, "Michigan Assessment of Basic Skills," Form SMT, were administered to all Michigan pupils in Grades 4 and 7 during January, 1970. The present report concerns the Grade 4 pupils. An outline of the test battery follows:

Vocabulary (20 minutes)

Reading (30 minutes)

English Expression

A. Spelling (5 minutes)
B. Effectiveness of Expression (7 minutes)
C. Grammar and Usage (6 minutes)
D. Punctuation and Capitalization (7 minutes)

Mathematics (25 minutes)

Five scores are obtained for each pupil, namely, one score based on each of the four tests and a composite achievement score, which is the sum of the Reading, English Expression, and Mathematics scaled scores.

NOTES ON PRINCIPAL FINDINGS

Total Group

158,713 Grade 4 pupils.

Sample

Spaced sample of 1,005 cases represents total group.

Appropriateness of Tests to Group

Scores extend throughout possible range on each of the four tests. Each mean is lower than middle difficulty for the group.

Score Conversion

Scores on each test converted to scaled with a mean of 50 and a standard deviation of 10. Scaled scores used in all reporting.

Reliability

Estimates of .869, Vocabulary; .899, Reading; .879, English Expression; .826, Mathematics; .947, Composite Achievement.

Speededness

Strong evidence of speededness in Vocabulary score; no important speed element in remaining scores.

Mean Item Difficulty

Mean deltas[1] of 13.7, Vocabulary; 11.8, Reading; 12.0, English Expression; 12.5, Mathematics. Middle-difficulty reference value is 11.7.

Mean Biserial Correlations

Means range from .42 for Vocabulary items to .58 for Reading items.

TOTAL-GROUP STATISTICS

Frequency distributions of the four test scores are presented on pages B1-5 and B1-6. The records of all pupils who responded to fewer than four items on a test are not included in the data for that test. Most of the missing scores represent absentees, for the tests were not all given at the same sitting.

In each distribution the scores cover the entire range of 0 to the highest possible value. With a low mean that suggests considerable difficulty, the Vocabulary distribution exhibits high positive skewness. There is less skewness apparent in the remaining distributions.

The raw scores were transmuted to scores with means of 50 and standard deviations of 10 for the total groups of the present distributions. The conversion parameters appear at the bottom of pages B1-5 and B1-6.

[1] The difficulty of the items in the Michigan Educational Assessment Battery have been expressed in terms of "delta" (Δ). The following explanation of this statistic is taken from Chung-Teh Fan's Item Analysis Table (Princeton: Educational Testing Service, 1952), p. 3.

"Delta is related in the following manner to the normal deviate, x, corresponding to p (proportion of correct responses in the total sample);

$$\Delta = 13 + 4x$$

where x is taken as positive for p's less than .50 and negative for p's greater than .50. Thus delta increases with increasing item difficulty. The parameters, 13 and 4 are arbitrary numbers which serve to eliminate negative values and to provide a sufficiently broad range that only integral values could be used if desired. The principal advantage of delta over p lies in the fact that equal increments in delta, unlike p, may reasonably be assumed to represent equal increments in difficulty. This characteristic of linearity of delta permits comparisons to be made between groups taking different test forms."

SAMPLE STATISTICS

A spaced sample of 1,005 records was drawn for detailed analysis. Score statistics for the sample can be compared with the total-group figures from the following tabulation:

	Sample		Total Group	
	Mean	S.D.	Mean	S.D.
Vocabulary	21.77	8.46	21.61	8.48
Reading	24.11	8.15	23.72	8.31
English Expression	26.33	8.10	26.11	8.24
Mathematics	16.30	5.68	16.06	5.55

There are no statistically significant differences between corresponding measures. It will be assumed that other statistics based on the sample will also describe the total group.

Reliability estimates are presented at the top of page B1-7. Those for the separately timed parts were computed by the Kuder-Richardson formula (20), and those for the English Expression total score and the Composite Achievement score were computed by the formula,

$$\text{reliability} = 1 - \frac{\sum w^2 SE_{Meas}^2}{\sigma_t^2}$$

In the case of the English Expression total score, the standard errors of measurement are those for the four subscores, the weights, w, are each unity, and σ_t^2 is the total-score variance. For the Composite Achievement score the weights are the appropriate "A" values in the conversion equations, Y = AX + B, on pages B1-5 and B1-6.

The estimated reliability of the brief subscores ranges from .65 to .72; for the four test scores, from .83 to .90; and for the Composite Achievement score it is .95. Except for Composite Achievement, the standard errors of measurement are listed in raw-score units. In scaled-score units they are 3.6 for Vocabulary, 3.1 for Reading, 3.4 for English Expression, and 4.3 for Mathematics.

Intercorrelations among the various scores are presented in the middle of page B1-7. Those among the four English subscores, which range from .49 to .57, are all well below the corresponding reliability estimates, an indication that each subtest takes a unique contribution to the total score. The correlations among the four test scores, although substantial, are also smaller than the corresponding reliability estimates. The smallest of these correlations, .672, is the Vocabulary-Mathematics figure, and the largest, .824, is Reading--English.

Data concerning speededness appear at the bottom of page B1-7. A test may be regarded as essentially unspeeded for a group if at least 80 per cent of the group reach the last item and if virtually every one reaches at least three-quarters of the items. By these criteria, Vocabulary is clearly speeded. The remaining tests are probably not unduly speeded, although in no instance did every one reach three-quarters of the items.

Item statistics are summarized on page B1-8. At the top of the page are frequency distributions of the difficulty index, delta. A 4-choice item that is known by one-half of the group and answered at random by the remainder would be expected to yield a delta of about 11.7. The observed mean for Vocabulary exceeds this middle-difficulty reference value by two delta points; Reading and English Expression are within 0.3 of middle difficulty, on the average; and Mathematics is more difficult by 0.8 delta point. More important, perhaps, than the average difficulty is the spread of difficulty among the items. When the group to be tested represents a broad range of ability, and when discrimination at both extremes of the ability range is required, then it is necessary that the items cover a broad difficulty range, for items of middle difficulty for the group as a whole may be altogether too easy for one extreme group and too hard for the other. Here, the upper portion of the item difficulty range appears to be well represented, but there is a notable paucity of items with deltas in the range of 6.0 to 9.0. The inclusion of even a few items within this range in each test would materially reduce the proportions of scores that fall in the chance area.

At the bottom of page B1-8 are distributions of the biserial correlations between item scores and criterion scores. The criterion for the items in the English Expression part scores is the 45-item total score. Otherwise, the criterion for each set is the score on the test noted at the top of the same column. Mean values range from .42 for Vocabulary to .58 for Reading. Only 8 of the 165 correlations are under .20.

TEST ANALYSIS REPORT FORM

Test Michigan Assessment of Basic Skills Subject _____ Form Form SMT

Taken by Grade 4 Date January 1970

Vocabulary				Reading			
Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval	Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval
48-50	81-83	47	99.97	40	70	143	99.9
45-47	78-80	438	99.7	38-39	67-68	2195	98.5
42-44	74-76	1608	98.7	36-37	65-66	6562	94.4
39-41	71-73	3589	96.4	34-35	62-64	10595	87.6
36-38	67-69	5861	92.7	32-33	60-61	13055	79.4
33-35	63-66	8303	87.4	30-31	58-59	14198	70.4
30-32	60-62	10550	80.7	28-29	55-56	13980	61.5
27-29	56-59	13134	72.4	26-27	53-54	13387	53.0
24-26	53-55	15764	62.4	24-25	50-52	12377	45.2
21-23	49-52	18858	50.5	22-23	48-49	11008	38.2
18-20	46-48	22211	36.4	20-21	46-47	9904	31.9
15-17	42-45	23497	21.5	18-19	43-44	9107	26.1
12-14	39-41	18575	9.7	16-17	41-42	8721	20.6
9-11	35-37	10160	3.3	14-15	38-40	8802	15.0
6- 8	32-34	3940	0.8	12-13	36-37	8598	9.6
3- 5	28-30	1098	0.1	10-11	33-35	7432	4.9
0- 2	25-27	177	0.0	8- 9	31-32	4680	1.9
		157810		6- 7	29-30	2087	0.6
				4- 5	26-27	688	0.2
				2- 3	24-25	197	0.03
				0- 1	21-23	41	0.00
						157757	

M_x = 21.61

σ_x = 8.48

M_y = 50

σ_y = 10

Md_x = 20.39

(50 items)

Conversion Data

Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.

Y = 1.1795 X + 24.5067

M_x = 23.72

σ_x = 8.31

M_y = 50

σ_y = 10

Md_x = 24.75

(40 items)

Conversion Data

Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.

Y = 1.2036 X + 21.4497

TEST ANALYSIS REPORT FORM

Test Michigan Assessment of Basic Skills Subject _____ Form SMT

Taken by Grade 4 Date January 1970

English Expression				Mathematics			
Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval	Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval
45	73	95	99.9	30	75	95	99.9
42-44	69-72	2085	98.6	28-29	72-73	1446	99.0
39-41	66-68	6722	94.4	26-27	68-70	5320	95.6
36-38	62-64	12646	86.3	24-25	64-66	9698	89.5
33-35	58-61	17569	75.2	22-23	61-63	13394	81.0
30-32	55-57	19866	62.6	20-21	57-59	16045	70.8
27-29	51-54	20342	49.7	18-19	54-55	17663	59.6
24-26	47-50	19546	37.3	16-17	50-52	18905	47.6
21-23	44-46	17186	26.4	14-15	46-48	19625	35.2
18-20	40-43	14724	17.1	12-13	43-44	18639	23.3
15-17	37-39	11679	9.7	10-11	39-41	16359	13.0
12-14	33-35	8528	4.3	8- 9	35-37	11658	5.6
9-11	29-32	4749	1.3	6- 7	32-34	6111	1.7
6- 8	26-28	1619	0.2	4- 5	28-30	2119	0.3
3- 5	22-24	327	0.03	2- 3	25-26	474	0.03
0- 2	18-21	45	0.00	0- 1	21-23	47	0.00
		157728				157598	

M_x = 26.11	Conversion Data	M_x = 16.06	Conversion Data
σ_x = 8.24	Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.	σ_x = 5.55	Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.
M_y = 50		M_y = 50	
σ_y = 10		σ_y = 10	
Md_x = 26.57		Md_x = 15.89	
(45 items)	Y = 1.2140 X + 18.3069	(30 items)	Y = 1.8021 X + 21.0627

Test Michigan Assessment of Basic Skills Subject Grade 4 Form SMT

N = 1005

Description of Sample:

Spaced sample.

Scoring Formulae and Reliability Coefficients for Sections

Section of Test	Scoring Formula	Reliability*	SE meas.	Section of Test	Scoring Formula	Reliability*	SE meas.
Vocabulary	R	.869	3.06	D Punc., Cap.	R	.675	1.34
Reading	R	.899	2.59	Total	R	.879**	2.82
English Express.:				Mathematics	R	.826	2.37
A Spelling	R	.694	1.62				
B Effect. Ex.	R	.719	1.28	Composite Achieve.		.947**	6.30
C Grammar	R	.652	1.37				

*Kuder-Richardson formula (20). **See text.

Intercorrelations of Sections

			English Expression								
Section	Voc.	Read.	A	B	C	D	Total	Math.		Mean	S.D.
Vocabulary		.755	.598	.604	.612	.576	.740	.672		21.77	8.46
Reading	.755		.661	.707	.666	.628	.824	.717		24.11	8.15
Engl. Exp.: A	.598	.661		.553	.536	.493	.824	.551		8.78	2.93
B	.604	.707	.553		.567	.546	.820	.599		6.46	2.41
C	.612	.666	.536	.567		.521	.802	.558		5.08	2.32
D	.576	.628	.493	.546	.521		.782	.608		6.01	2.36
Total	.740	.824	.824	.820	.802	.782		.716		26.33	8.10
Mathematics	.672	.717	.551	.599	.558	.608	.716			16.30	5.68
Comp. Ach.	.791	.925					.925	.891		151.17	27.28

Speededness of Sections

			English Expression								
Section	Voc.		Read.		A	B	C	D		Math.	
Per cent completing test..	65.9		87.6		90.0	92.3	94.3	94.3		87.7	
Per cent completing 75 per cent of test .	85.8		97.3		97.7	97.7	97.5	98.2		97.4	
Number of items reached by 80 per cent of the candidates	41		40		15	10	10	10		30	
Total number of items	50		40		15	10	10	10		30	

Test Michigan Assessment of Basic Skills Subject Grade 4 Form SMT

Frequency Distributions of Original Deltas
and Biserial Correlations, by Sections

Delta	English Expression									
	Voc.	Read.		A	B	C	D	Total		Math.
19.0 up ..										
18.0-18.9										
17.0-17.9	3									
16.0-16.9	3	2		1				1		1
15.0-15.9	8	2		1		1		2		2
14.0-14.9	11	2		1		1	1	3		5
13.0-13.9	9	5		2	3	4	2	11		5
12.0-12.9	4	8		2	1	1	3	7		7
11.0-11.9	8	5		3	1	-	-	4		4
10.0-10.9	3	7		1	3	3	3	10		1
9.0- 9.9	1	7		3	1		1	5		4
8.0- 8.9		-		1	1			2		-
7.0- 7.9		2								1
6.0- 6.9										
Total	50	40		15	10	10	10	45		30
Mean	13.7	11.8		12.0	11.3	12.9	11.9	12.0		12.5
σ	1.9	2.1		2.3	1.7	1.7	1.6	2.0		2.0
a										
b										
r_{bis}										
.90-.99										
.80-.89		1								
.70-.79	1	8								1
.60-.69	3	15		6	5	2	2	15		3
.50-.59	12	7		1	2	3	3	9		19
.40-.49	21	3		3	3	3	4	13		1
.30-.39	4	3		3		2	1	6		5
.20-.29	4	2		1				1		-
.10-.19	4	1		1				1		1
.00-.09	-									
Negative .	1									
Total	50	40		15	10	10	10	45		30
Not Comp.										
Mean	.42	.58		.47	.56	.50	.50	.51		.51
σ	.15	.16		.15	.08	.11	.08	.12		.12

Standard Δ = a(original Δ) + b

APPENDIX B-2

GRADE 7 TEST ANALYSIS

Grade 7

The Michigan Educational Assessment Program became operational during the 1969-70 academic year. Appropriate levels of the test battery, "Michigan Assessment of Basic Skills," Form SMT, were administered to all Michigan pupils in Grades 4 and 7 during January, 1970. The present report concerns the Grade 7 pupils. An outline of the test battery follows:

Vocabulary (20 minutes)

Reading (30 minutes)

English Expression

A. Spelling (6 minutes)
B. Effectiveness of Expression (7 minutes)
C. Grammar and Usage (6 minutes)
D. Punctuation and Capitalization (6 minutes)

Mathematics (25 minutes)

Five scores are obtained for each pupil, namely, one score based on each of the four tests and a composite achievement score, which is the sum of the Reading, English Expression, and Mathematics scaled scores.

NOTES ON PRINCIPAL FINDINGS

Total Group

159,407 Grade 7 pupils.

Sample

Spaced sample of 1,005 cases represents total group.

Appropriateness of Tests to Group

Scores extend throughout possible range on each of the four tests. Each mean is lower than middle difficulty for the group.

Score Conversion

Scores on each test converted to scale with a mean of 50 and a standard deviation of 10. Scaled scores used in all reporting.

Reliability

Estimates of .862, Vocabulary; .818, Reading; .883, English Expression; .844, Mathematics; .93[illegible], Composite Achievement.

Speededness

No evidence of undue speededness in any of the tests.

Mean Item Difficulty

Mean deltas[1] of 12.4, Vocabulary; 12.2, Reading; 12.2, English Expression; 12.6, Mathematics. Middle-difficulty reference value is 11.7.

Mean Biserial Correlations

Means range from .46 for Vocabulary items and for Reading items to .55 for Mathematics items.

TOTAL-GROUP STATISTICS

Frequency distributions of the four test scores are presented on pages B2-5 and B2-6. The records of all pupils who responded to fewer than four items on a test are not included in the data for that test. Most of the missing scores represent absentees, for the tests were not all given at the same sitting.

In each distribution the scores cover the entire range of 0 to the highest possible value. There is little skewness apparent in any of the distributions.

The raw scores were transmuted to scores with means of 50 and standard deviations of 10 for the total groups of the present distributions. The conversion parameters appear at the bottom of pages B2-5 and B2-6.

[1] The difficulty of the items in the Michigan Educational Assessment Battery have been expressed in terms of "delta" (Δ). The following explanation of this statistic is taken from Chung-Teh Fan's Item Analysis Table (Princeton: Educational Testing Service, 1952), p. 3.

"Delta is related in the following manner to the normal deviate, x, corresponding to p (proportion of correct responses in the total sample);

$$\Delta = 13 + 4x$$

where x is taken as positive for p's less than .50 and negative for p's greater than .50. Thus delta increases with increasing item difficulty. The parameters, 13 and 4 are arbitrary numbers which serve to eliminate negative values and to provide a sufficiently broad range that only integral values could be used if desired. The principal advantage of delta over p lies in the fact that equal increments in delta, unlike p, may reasonably be assumed to represent equal increments in difficulty. This characteristic of linearity of delta permits comparisons to be made between groups taking different test forms."

SAMPLE STATISTICS

A spaced sample of 1,005 records was drawn for detailed analysis. Score statistics for the sample can be compared with the total-group figures from the following tabulation:

	Sample		Total Group	
	Mean	S.D.	Mean	S.D.
Vocabulary	27.31	8.12	27.46	8.32
Reading	22.37	6.27	22.27	6.52
English Expression	31.47	8.94	31.68	9.03
Mathematics	15.98	5.88	16.14	6.03

There are no statistically significant differences between corresponding measures. It will be assumed that other statistics based on the sample will also describe the total group.

Reliability estimates are presented at the top of page B2-7. Those for the separately timed parts were computed by the Kuder-Richardson formula (20), and those for the English Expression total score and the Composite Achievement socre were computed by the formula,

$$\text{reliability} = 1 - \frac{\sum w^2 SE_{Meas}{}^2}{\sigma_t{}^2}$$

In the case of the English Expression total score, the standard errors of measurement are those for the four subscores, the weights, w, are each unity, and $\sigma_t{}^2$ is the total-score variance. For the Composite Achievement score the weights are the appropriate "A" values in the conversion equations, $Y = AX + B$, on pages B2-5 and B2-6.

The estimated reliability of the brief subscores ranges from .61 to .75; for the four test scores, from .82 to .88; and for the Composite Achievement score it is .94. Except for Composite Achievement, the standard errors of measurement are listed in raw-score units. In scaled-score units they are 3.6 for Vocabulary 4.1 for Reading, 3.4 for English Expression, and 3.9 for Mathematics.

Intercorrelations among the various scores are presented in the middle of page B2-7. Those among the four English subscores, which range from .47 to .59, are all below the corresponding reliability estimates, as indication that each subtest makes a unique contribution to the total score. The correlations among the four test scores, although substantial, are also smaller than the corresponding reliability estimates. The smallest of these correlations, .665, is the Reading-Mathematics figure, and the largest; .755, is Reading-English.

Data concerning speededness appear at the bottom of page B2-7. A test may be regarded as essentially unspeeded for a group if at least 80 per cent of the group reach the last item and if virtually every one reaches at least three-quarters of the items. Each separately timed part was completed by more than 80 per cent of the group, but in no instance did every one reach three-quarters of the items.

Item statistics are summarized on page B2-8. At the top of the page are frequency distributions of the difficulty index, delta. A 4-choice item that is known by one-half of the group and answered at random by the remainder would be expected to yield a delta of about 11.7. The observed means for the four tests all lie within 12.2 to 12.6, somewhat higher than the middle-difficulty reference value. More important, perhaps, than the average difficulty is the spread of difficulty among the items. When the group to be tested represents a broad range of ability, and when discrimination at both extremes of the ability range is required, then it is necessary that the items cover a broad difficulty range, for items of middle difficulty for the group as a whole may be altogether too easy for one extreme group and too hard for the other. Here, the upper portion of the item difficulty range appears to be well represented in each of the four tests, but a few additional items with deltas in the range of 6.0 to 9.0 in the appropriate parts or tests would materially reduce the proportions of scores that fall in the chance area.

At the bottom of page B2-8 are distributions of the biserial correlations between item scores and criterion scores. The criterion for the items in the English Expression part scores is the 55-item total score. Otherwise, the criterion for each set is the score on the test noted at the top of the same column. Mean values range from .46 for Vocabulary and for Reading to .55 for Mathematics. Only 3 of the 175 correlations are under .20.

TEST ANALYSIS REPORT FORM

Test Michigan Assessment of Basic Skills Subject _____ Form SMT

Taken by Grade 7 Date January 1970

Vocabulary				Reading			
Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval	Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval
48-50	75-77	177	99.9	40	77	46	99.97
45-47	71-73	1617	98.9	38-39	74-76	666	99.6
42-44	67-70	4972	95.7	36-37	71-73	2070	98.2
39-41	64-66	9610	89.7	34-35	68-70	3939	95.8
36-38	60-63	13498	81.1	32-33	65-66	6499	91.6
33-35	57-59	17098	70.3	30-31	62-63	9534	85.6
30-32	53-55	18866	58.4	28-29	59-60	12561	77.7
27-29	49-52	19139	46.3	26-27	56-57	15549	67.9
24-26	46-48	19078	34.3	24-25	53-54	17390	56.9
21-23	42-45	18157	22.8	22-23	50-51	18186	45.4
18-20	39-41	16007	12.7	20-21	47-48	17213	34.5
15-17	35-37	11525	5.4	18-19	43-45	15626	24.6
12-14	31-34	5731	1.8	16-17	40-42	13416	16.2
9-11	28-30	2111	0.5	14-15	37-39	10531	9.5
6- 8	24-27	600	0.1	12-13	34-36	7366	4.9
3- 5	21-23	138	0.02	10-11	31-33	4382	2.1
0- 2	17-19	29	0.00	8- 9	28-30	2194	0.7
				6- 7	25-27	837	0.2
		158353		4- 5	22-24	222	0.04
				2- 3	19-20	62	0.01
				0- 1	16-17	9	0.00
						158298	

M_x = 27.46

σ_x = 8.32

M_y = 50

σ_y = 10

Md_x = 27.41

(50 items)

Conversion Data

Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.

Y = 1.2024 X + 16.9806

M_x = 22.27

σ_x = 6.52

M_y = 50

σ_y = 10

Md_x = 22.30

(40 items)

Conversion Data

Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.

Y = 1.5337 X + 15.8386

TEST ANALYSIS REPORT FORM

Test Michigan Assessment of Basic Skills Subject _____ Form SMT

Taken by Grade 7 Date January 1970

English Expression				Mathematics			
Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval	Raw Score X	Standard Score Y	f	Percentile Rank of Lower Limit of Interval
54-55	75-76	101	99.9	30	73	495	99.7
51-53	71-74	1045	99.3	28-29	70-71	3398	97.5
48-50	68-70	3697	96.9	26-27	66-68	7143	93.0
45-47	65-67	7325	92.3	24-25	63-65	10436	86.4
42-44	61-64	11325	85.1	22-23	60-61	12894	78.2
39-41	58-60	15241	75.5	20-21	56-58	14563	69.0
36-38	55-57	18002	64.1	18-19	53-55	16073	58.8
33-35	51-54	19170	52.0	16-17	50-51	17200	48.0
30-32	48-50	18781	40.1	14-15	46-48	17787	36.7
27-29	45-47	14775	29.0	12-13	43-45	17934	25.3
24-26	41-44	14775	19.7	10-11	40-41	16003	15.2
21-23	38-40	11929	12.2	8- 9	36-38	12790	7.1
18-20	35-37	8638	6.7	6- 7	33-35	7617	2.3
15-17	32-34	5767	3.0	4- 5	30-32	2992	0.4
12-14	28-30	3187	1.0	2- 3	27-28	581	0.03
9-11	25-27	1215	0.3	0- 1	23-25	50	0.00
6- 8	22-24	307	0.1			157956	
3- 5	18-20	85	0.01				
0- 2	15-17	23	0.00				
		158183					

M_x = 31.68

σ_x = 9.03

M_y = 50

σ_y = 10

Md_x = 32.00

(55 items)

Conversion Data

Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.

Y = 1.1075 X + 14.9745

M_x = 16.14

σ_x = 6.03

M_y = 50

σ_y = 10

Md_x = 15.87

(30 items)

Conversion Data

Raw scores converted to scale with mean of 50 and standard deviation of 10 for this group.

Y= 1.6585 X + 23.2280

Test Michigan Assessment of Basic Skills Subject Grade 7 Form SMT

N = 1005

Description of Sample:

Spaced sample.

Scoring Formulae and Reliability Coefficients for Sections

Section of Test	Scoring Formula	Reliability*	$SE_{meas.}$	Section of Test	Scoring Formula	Reliability*	$SE_{meas.}$
Vocabulary	R	.862	3.02	D Punc., Cap.	R	.689	1.50
Reading	R	.818	2.67	Total	R	.883**	3.06
English Express.:				Mathematics	R	.844	2.32
A Spelling	R	.755	1.82				
B Effect. Exp.	R	.607	1.36	Composite Achieve.		.936**	6.57
C Grammar	R	.651	1.46				

*Kuder-Richardson formula (20). **See text.

Intercorrelations of Sections

English Expression

Section	Voc.	Read.	A	B	C	D	Total	Math.		Mean	S.D.
Vocabulary		.748	.543	.549	.634	.564	.701	.709		27.31	8.12
Reading	.748		.610	.596	.655	.594	.755	.645		22.37	6.27
Engl. Exp.: A	.543	.610		.472	.553	.547	.843	.486		11.90	3.68
B	.549	.596	.472		.563	.525	.750	.514		5.82	2.16
C	.634	.655	.553	.563		.591	.818	.583		7.38	2.48
D	.564	.594	.547	.525	.591		.816	.594		6.37	2.68
Total	.701	.755	.843	.750	.818	.816		.664		31.47	8.94
Mathematics	.709	.645	.486	.514	.583	.594	.664			15.98	5.88
Comp. Ach.	.808	.897					.907	.865		149.64	26.05

Speededness of Sections

English Expression

Section	Voc.		Read.		A	B	C	D		Math.	
Per cent completing test.	83.8		90.0		94.6	96.8	96.5	94.1		90.0	
Per cent completing 75 per cent of test .	96.7		97.9		97.7	99.7	98.7	98.2		98.2	
Number of items reached by 80 per cent of the candidates	50		40		20	10	13	12		30	
Total number of ems	50		40		20	10	13	12		30	

Test Michigan Assessment of Basic Skills Subject Grade 7 Form SMT

Frequency Distributions of Original Deltas and Biserial Correlations, by Sections

Delta	English Expression									
	Voc.	Read.		A	B	C	D	Total		Math.
19.0 up ..										
18.0-18.9										
17.0-17.9	1	1		1		1		2		1
16.0-16.9	1	1		2	1	-	1	4		1
15.0-15.9	7	3		-	-	2	1	3		1
14.0-14.9	4	5		-	-	1	-	1		6
13.0-13.9	8	7		2	3	2	2	9		5
12.0-12.9	8	8		5	1	-	5	11		5
11.0-11.9	8	4		3	2	2	2	9		2
10.0-10.9	6	4		1	2	2	-	5		6
9.0- 9.9	1	2		2	1	-	-	3		2
8.0- 8.9	3	1		3		2	1	6		-
7.0- 7.9	3	3		1		1		2		1
6.0- 6.9		1								
Total	50	40		20	10	13	12	55		30
Mean	12.4	12.2		11.9	12.1	12.2	12.6	12.2		12.6
σ	2.4	2.6		2.6	1.9	2.9	1.9	2.5		2.1
a										
b										
r_{bis}										
.90-.99										
.80-.89										
.70-.79	1	2		1				1		1
.60-.69	5	6		4	1	1	2	8		10
.50-.59	15	8		3	2	6	3	14		9
.40-.49	17	13		8	5	3	4	20		6
.30-.39	7	5		3	2	2	3	10		4
.20-.29	5	4		1		-		1		
.10-.19		1				1		1		
.00-.09		1								
Negative .										
Total	50	40		20	10	13	12	55		30
Not Comp.										
Mean	.46	.46		.48	.46	.48	.49	.48		.55
σ	.12	.15		.12	.09	.11	.09	.11		.12

Standard Δ = a(original Δ) + b