

DEVELOPING CLASSROOM TESTS FOR HIGH  
SCHOOL MODERN ALGEBRA

by

RICHARD JAY JEWETT

B. S., Kansas State University, 1964

---

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

College of Education

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

1966

Approved by:

  
Major Professor

#### ACKNOWLEDGEMENT

The author wishes to express his sincere appreciation to Professor Russel G. Drumright, major professor, for his continued assistance and guidance in writing this report.

~~378.73~~ LD  
K160r 266's  
1966 R4  
J 59  
C. 2

TABLE OF CONTENTS

	PAGE
INTRODUCTION . . . . .	1
Purpose of the Study . . . . .	1
Method of Procedure . . . . .	2
Definitions of Terms . . . . .	2
Measurement . . . . .	2
Test and examination . . . . .	2
Classroom test . . . . .	2
Quiz . . . . .	2
GENERAL PRINCIPLES OF TEST CONSTRUCTION . . . . .	3
Functions of Tests . . . . .	3
The instructional test . . . . .	4
The mastery test . . . . .	4
The measurement test . . . . .	6
Types of Tests and Their Uses . . . . .	6
The oral examination . . . . .	8
The objective test . . . . .	8
The essay test . . . . .	11
The performance test . . . . .	12
Methods of Construction . . . . .	12
The planning stage . . . . .	12
The construction stage . . . . .	15
The evaluation stage . . . . .	17

## THE APPLICATION OF TEST CONSTRUCTION TO HIGH

SCHOOL MODERN ALGEBRA . . . . .	23
Planning the Test . . . . .	24
Objectives . . . . .	24
The test-specification chart . . . . .	26
Constructing the Test . . . . .	30
The item-writing phase . . . . .	30
The compiling phase . . . . .	41
Evaluating the Test . . . . .	42
SUMMARY . . . . .	43
BIBLIOGRAPHY . . . . .	46

## INTRODUCTION

Ultimately, the primary goal of most teaching is to improve pupil performance in a variety of situations by developing his skills, knowledges, and understandings. In order to determine what improvement has been made, it is necessary for the teacher to observe the pupil in each situation and make an interpretation of his behavior. Teachers have tools--tests--which are not really alternatives to teacher observation of student behavior, but rather a specialized technique for extending, refining, interpreting, and recording those observations.<sup>1</sup>

Textbooks in measurement and evaluation usually treat this subject only in general terms. The characteristics which distinguish each subject from another may differ widely, therefore little information is available which applies directly to high school modern algebra. Thus, since many schools are using the modern approach, it would be worthwhile to investigate how a teacher would develop classroom tests for high school modern algebra.

### Purpose of the Study

It was the purpose of this study to (1) present the general principles of test construction as they are commonly found in texts on measurement and evaluation, and (2) offer methods of applying these principles to high school modern algebra.

---

<sup>1</sup>Robert L. Ebel, Measuring Educational Achievement (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965), p. 4.

### Method of Procedure

The method of carrying out this study of test construction for modern algebra was reading and library research. Only literature pertaining to classroom tests in mathematics and references on general measurement techniques which were available in the Kansas State University Library were investigated. The general information was then applied to high school modern algebra by the author.

### Definitions of Terms

Measurement. Throughout this report the following definition of measurement, by Robert L. Ebel, is used:

Measurement is a process of assigning numbers to the individual members of a set of objects or persons for the purpose of indicating differences among them in the degree to which they possess the characteristic being measured. If any characteristic of persons or things can be defined clearly enough so observed differences between them with respect to this characteristic can be consistently verified, the characteristic is measurable.<sup>1</sup>

Test and examination. The terms test and examination are used interchangeably to designate any device or procedure for measuring a student's knowledge of the subject.

Classroom test. The term classroom test, for the purpose of this paper, is defined as any test constructed by a teacher for use in his classroom.

Quiz. The term quiz is used in this report to refer to any classroom test of short duration, usually ten to twenty minutes.

---

<sup>1</sup>Ibid., pp. 454-455.

## GENERAL PRINCIPLES OF TEST CONSTRUCTION

The general principles of test construction apply to essentially all subjects being taught in high school. The specific application of these principles to high school modern algebra is not a matter of developing new principles, but rather one of using the new approaches of the modern algebra with these general rules.

Functions of Tests

There is no one best test; it is for the test-maker to construct the best test for a given purpose. Testing is used throughout the school year to accomplish different goals. Pupils are evaluated at the beginning of the year in order for the teacher to establish a common background for the entire class, and at the end of chapters, units, grading periods, or semesters to determine the ultimate achievement of each student.

Teaching is aimed at improving pupil performance in a variety of situations by developing knowledge, skill, and understanding, which can be successfully applied in life.<sup>1</sup> Ideally all testing should be directed toward this type of pupil performance, since the pupil who applies knowledge has mastered it well. A pupil who has simply memorized the principles may be able to recall facts without the ability to apply them. Teaching which is aimed at knowledge, skill, and understanding can be evaluated by three types of tests: an instructional

---

<sup>1</sup>John A. Green, Teacher-Made Tests (New York: Harper & Row, 1963), p. 7.

test, which checks skill; a mastery test, which evaluates the knowledge of fundamentals; and a measurement test, which measures the pupil's understanding of concepts.

The instructional test. Testing the student's knowledge or skill with an instructional test is aimed more at showing him some of his strengths and weaknesses than at determining his grade. An instructional test emphasizes and reinforces the important points of a unit which a pupil must learn and remember. To be meaningful, with the maximum learning to result, the test is likely to be in the form of a ten- to twenty-minute quiz which is graded by the pupils and discussed immediately afterwards.<sup>1</sup>

An instructional test is usually neither a speed test, having a ceiling in terms of number of items completed, nor a power test, with the ceiling in terms of the difficulty of items, since the student is supposed to know everything on the test and is given sufficient time that he should be able to finish. This test may be used as a teaching device when the pupils have not sufficiently learned the material from classroom instruction or from the homework. It can be used to supplement instruction by presenting a test on the most important areas which will show each student where he needs to concentrate further study.

The mastery test. If instruction is not to proceed so slowly that pupil interest is lost, or advance so rapidly that many pupils fail to master even the minimum essential skills, the teacher must

---

<sup>1</sup>Ibid., p. 2.



pace the instruction to the learning rate of the majority. The mastery test is the teacher's method of determining this learning rate. In order to proceed to more difficult tasks or to work based on previous knowledge, it is necessary for the teacher to determine when the pupils have acquired sufficient competence or skill to enable them to continue. Since these skills will be the foundation for subsequent work, the teacher must expect the students to answer 90 per cent correctly.<sup>1</sup>

A mastery test could be a speed test, since the pupils might be expected to work rapidly as well as accurately in certain cases. The students may have learned to expand binomials by inspection, for instance, and thus the teacher might want to limit the time in order to prevent students from multiplying the binomials out. There would be a minimum number of problems that all pupils would be expected to complete, with ample time allotted for this part, and additional problems would be added. In this case, these would be more problems than any one pupil could finish in the given time, in order to establish each pupil's mastery beyond the minimum requirements.

A mastery test could be a power test, on the other hand, where the teacher would be checking the ultimate achievement of the pupils beyond the required skills. As in the previous example, the teacher would have a minimum number of expansions which all the pupils were expected to complete and then from that point progressively harder or more complicated problems would be included. No one would be expected to complete the most difficult problem.

---

<sup>1</sup>Ibid., p. 2-3.

The measurement test. The examination with which pupils are most familiar is the measurement test. This is the test which purports to give exact measurement of each pupil's achievement throughout the school year, during the semester, or at the end of a unit. This test is the primary grade determiner for each pupil. If it is of the informal objective type, it has a 50 per cent level of difficulty--i.e., the average pupil will answer 50 per cent correctly--and has a power arrangement with the first few items easy enough for the slowest pupil and the last items too difficult for the brightest. This arrangement permits the teacher to establish a point of maximum achievement for each pupil.<sup>1</sup>

#### Types of Tests and Their Uses

Once the function of a test has been determined the teacher is presented with the task of choosing the right type of test to accomplish that function. One situation might require the pupils to prove a particular statement. Another would call for the recognition of terms, definitions, or problem-solving. The teacher may want to check quickly to see if the class has understood a point just covered, or there may have been some practical application just made apparent which needs a practical testing situation. Each of these situations can be classified into an oral, objective, essay, or performance type test.<sup>2</sup> Table I gives a composite picture of the relative strengths and weaknesses of each of these types.

---

<sup>1</sup>Ibid., p. 3.

<sup>2</sup>Ibid., p. 4.

TABLE I  
STRENGTHS AND WEAKNESSES OF VARIOUS TEST FORMS<sup>1</sup>

Strengths	Weaknesses
ORAL	
<ol style="list-style-type: none"> <li>1. Gives extensive measurement.</li> <li>2. Is useful as an instructional device.</li> <li>3. Permits teacher to give cues to elicit desired responses.</li> <li>4. Improves test rapport for pupils who fear written examinations.</li> </ol>	<ol style="list-style-type: none"> <li>1. Is too time-consuming.</li> <li>2. Results in poor pupil performance due to lack of practice.</li> <li>3. Provides limited sample unless pupils are tested individually.</li> <li>4. Is frequently poorly planned.</li> <li>5. Is not subject to refinement.</li> <li>6. Gives poor comparative evaluation of pupils.</li> </ol>
OBJECTIVE	
<ol style="list-style-type: none"> <li>1. Gives an extensive test sample.</li> <li>2. Can be made highly reliable.</li> <li>3. Can be graded objectively and quickly.</li> <li>4. Eliminates bluffing.</li> <li>5. Can be subjected to item analysis and further refinement.</li> <li>6. Can be adapted to several teaching objectives.</li> <li>7. Can be made highly valid for some teaching objectives.</li> </ol>	<ol style="list-style-type: none"> <li>1. Frequently neglects measurement of higher thought processes.</li> <li>2. Overemphasizes rote learning.</li> <li>3. Promotes poor study habits.</li> <li>4. Encourages guessing.</li> <li>5. Is difficult to prepare.</li> <li>6. Costs more than essay to prepare and reproduce.</li> </ol>
ESSAY	
<ol style="list-style-type: none"> <li>1. Is applicable to measurement of writing, organizational ability, and creativeness.</li> <li>2. Is easy to construct.</li> <li>3. Promotes proper type of study.</li> <li>4. Is adaptable to several subject fields.</li> </ol>	<ol style="list-style-type: none"> <li>1. Gives a limited test sample.</li> <li>2. Is difficult to grade.</li> <li>3. Favors the verbally inclined student.</li> <li>4. Has low reliability.</li> <li>5. Encourages bluffing.</li> </ol>
PERFORMANCE	
<ol style="list-style-type: none"> <li>1. Stresses application of knowledge.</li> <li>2. Can be used as a learning device.</li> <li>3. May give a truer achievement picture for the verbally handicapped pupil.</li> <li>4. Measures some skills and abstract abilities not measured by other conventional test forms.</li> </ol>	<ol style="list-style-type: none"> <li>1. Is not adaptable to many fields of learning.</li> <li>2. Is difficult to construct.</li> <li>3. Is often difficult to grade.</li> <li>4. Is often time-consuming.</li> </ol>

<sup>1</sup>Ibid., p. 5.

The oral examination. Oral questioning can give an extensive picture of a pupil's knowledge, but in order to do so sufficient time must be devoted to the examination, and the questions must be carefully prepared ahead of time in order to give an adequate sample of the areas covered. Each exam would take from one-half to one hour. Another form of oral questioning is the familiar teaching device, which is a means of stimulating pupil participation in class discussion, where grades are not counted heavily. Oral questioning permits the examiner to be flexible in his procedure. He may adapt his questions to fit each pupil, and he can draw out additional meanings behind vague or incomplete statements. This is an excellent means of following the thought processes which a pupil has used in solving a problem, and a good way of diagnosing pupil difficulties.

Many areas of teaching stress skills, knowledges, and understandings, which lay the foundation for further learning but for which there is little immediate application. Classroom tests of the essay or objective type will give the best estimate of learning in these areas.<sup>1</sup>

The objective test. The objective test is strong where the essay test is weak. It permits reliable measurement of an extensive amount of factual material, and is the most objective type in scoring, hence the name. It is difficult to construct objective tests which measure such types of learning as complex understandings, creativity, attitude changes, and problem-solving.<sup>2</sup> The objective test has four

---

<sup>1</sup>Ibid., p. 4.

<sup>2</sup>Ibid., p. 6.

common forms which are extensively used in classroom tests; the short-answer form, the alternate-response form, the multiple-choice form, and the matching form.

The short-answer form of objective test may include questions, incomplete sentences, definitions, or identification items. This form is characterized by having its answers in the form of a word, a number, or as long as a sentence in the case of a definition. This form is good for the measurement of recall, and it is better than the alternate-response and multiple-choice forms in that the student cannot guess the correct response unless the item is poorly constructed. Short-answer tests tend to stress rote memory, and the teacher who uses verbatim quoting will further emphasize this weakness. This can be avoided by rephrasing the statement, which will serve to check understanding.

The alternate-response form generally consists of a statement to be judged true or false. Since there are only two choices, the student has a 50 per cent chance of guessing the correct response; and since he will most likely remember something about each statement, his chances for guessing correctly may be increased. Hence it is very important that items be without unnecessary clues. This form can measure both factual information and understanding of concepts, but in order to be valid and reliable, it must consist of no fewer than about forty items and be carefully constructed. The test-wise pupil may further increase his chances of guessing correctly since teachers have a tendency to select most of the items directly from the book and to include more true than false items.<sup>1</sup> The true-false type of question

---

<sup>1</sup>Ibid., p. 31.

tends to penalize the brighter student since that student most frequently thinks of the exception or conditional factors that change the meaning of the question.<sup>1</sup> It is therefore very important that the test-maker make each statement entirely true or entirely false. The test-maker can avoid these ambiguous statements by modifying the test such that when the student marks a statement false, he must tell what would make the statement true. The true-false item is adaptable and relatively easy to prepare, but it is less discriminating, item for item, than the multiple-choice type, and somewhat more subject to ambiguity and misinterpretation.<sup>2</sup>

The multiple-choice test is a good test for measuring a variety of educational objectives. According to Robert L. Ebel, "Almost any understanding or ability that can be tested by means of any other item form--short answer, completion, true-false, matching or essay--can also be tested by means of multiple-choice test items."<sup>3</sup> Multiple-choice items have an incomplete statement or a question followed by several choices which include one correct answer and several distracters, all of which are plausible.

The matching examination is most useful for measuring recognition and recall. In this test, a number of homogeneous questions or statements are given and a list of between five and fifteen homogeneous

---

<sup>1</sup>Denis Baron and Harold W. Bernard, Evaluation Techniques for Classroom Teachers (New York: McGraw-Hill Book Company, 1958), p. 207.

<sup>2</sup>Ebel. op. cit., pp. 60-61.

<sup>3</sup>Ibid., p. 149.

responses arranged for maximum clarity and convenience to the examinee is provided to match with the premises. The number of responses is limited, since for a larger list, much time would be spent searching for the correct response, especially if the list were to be divided between two pages. A one-to-one matching between premises and responses should not be provided, as a student may be able to eliminate what he knows and have a good chance of guessing the rest correctly.

The essay test. The essay test is the most effective means of evaluating creativeness on the part of a student. It measures such non-structured types of learning as problem-solving, creative thinking, writing, and organizational ability.<sup>1</sup> If the pupil is to write a proof of a statement, he must be able to supply the steps from his own background of knowledge, organize his steps until they are logical, and supply reasons for his solution. Thus the mathematical proof is necessarily classified as an essay test.

The essay test is weak in that it is inadequate for sampling a great breadth of knowledge. Due to the amount of time required for the student to organize and present his answer, only a very few questions can be asked. Another disadvantage is the subjectivity of grading. In order to avoid subjectivity, it is necessary that all possible points be listed on the examination key, and these must be adhered to very closely. This practice also reduces the tendency on the part of the teacher to allow points where a student has tried to bluff and mislead.

---

<sup>1</sup>Green, op. cit., p. 6.

The performance test. Knowledge and the successful application of knowledge do not always correlate highly. Those who apply knowledge in successful performance have achieved a higher degree of learning than that represented by the mere retention of knowledge. Performance testing is valuable for any subject in which pupils are being taught to follow specific procedures or to create some product, with its strength being in its ability to examine the application of classroom learnings in actual job performance. Because the product and the procedure may be closely related, as in Industrial Arts, the teacher interested in performance measurement may choose to concentrate on the procedure, the product, or on any combination of the two.<sup>1</sup>

#### Methods of Construction

In developing a high-quality test, the teacher is compelled to observe many rules and many cautions, most of which apply to all types of tests. The job of preparing a test need not be laborious, but it is obvious that some advance attention must be given to the preparation if each test is to be of educational value. There are essentially three stages to developing a high-quality test: the planning stage, the construction stage, and the evaluation stage.

The planning stage. Every teacher should be directing his instruction toward a number of specific and general educational objectives. They should be stated in precise, operational terms to form the basis for planning a test, since the measurement objectives should be

---

<sup>1</sup>Ibid., p. 43.



the same as the instructional objectives. In testing, the instructional objectives determine the choice of test form and the type and number of items. These objectives, along with the content areas which have been taught, form the outline for constructing a good test. Since all objectives do not have the same value, weights in the form of percentages should be assigned to these to indicate emphasis. Perhaps the best way to insure proper emphasis in both the objectives and the content areas is to construct a test-specification chart.<sup>1</sup>

A test-specification chart is prepared by listing the content areas and the objectives for the unit to be tested. To form a chart, the content areas are put on the left-hand side and the objectives across the top, making a two-way table. If such a test-specification chart is used at the beginning of each chapter, the teacher will have in written form, his objectives, content areas, and an outline for quizzes and the chapter test. He can see at a glance the relative value of each area and guide his instruction accordingly. If during a chapter, the weights must be changed, this should be done on the chart also, so that when it is time for a test the measurement objectives and the instructional objectives will still be the same. In Table II it can be noted that content area W had only one objective, D, which is the objective for no other content area. This 5 per cent of the material might then be tested separately in a short quiz to avoid having numerous different kinds of questions on the unit test. In the same manner, content area R has only one objective, B, and again a quiz

---

<sup>1</sup>Ibid., p. 16.

might be more practical for evaluation than an additional area on a unit test. Another important observation to make about Table II is that content area X constitutes 25 per cent of the total content, and objective B at 40 per cent, appears to be the most important objective in the unit. Thus Table II has, in general, outlined the types of tests and the relative values of each instructional goal.

TABLE II  
TEST-SPECIFICATION CHART

CONTENT AREAS	OBJECTIVES				Totals
	A	B	C	D	
R	0	10	0	0	10
S	10	0	10	0	20
T	5	10	5	0	20
U	5	5	0	0	10
V	0	5	5	0	10
W	0	0	0	5	5
X	5	10	10	0	25
Totals	25	40	30	5	100

The task of the test-maker is greatly reduced if he can write several test items for each content area, directed towards the specific objectives, as he plans a lesson. This will serve a dual purpose of guiding his instruction towards the objectives, and eliminating the need to compose a test from the beginning at examination time.

If more items than needed are written each time and kept on file, he can choose his questions to best suit the particular class.

The construction stage. When the test has been planned in this manner the construction of a test is in two phases: the item-writing phase and the compiling phase. Item-writing is done when the lessons are being planned and compiling is done when the test is to be given.

In writing items for tests, as far as possible, test for principles and generalizations rather than isolated facts. The items should be clear and concise and worded in such a way that pupils are not unduly penalized by reading speed and comprehension. Inaccurate wording often prevents the pupil from understanding the questions asked, and he may thus be unable to answer even those questions to which he would otherwise have a ready answer.

When striving for objectivity, it is particularly important that there be no clues to the correct answers other than the meaning of the questions. In the same manner, information used in one problem should not provide a lead to another's answer. The items must also be worded such that an answer which is supposed to be correct is not only unquestionably correct, but also the only correct answer. The teacher must be careful to omit only significant words in the incomplete statement items, and must be sure to leave enough clues so that the pupil who knows the answer can give the correct response. Grammatical clues to the correct answer and over-mutilated statements should be avoided. If too many clues are given, most students will answer correctly, while if too few clues are given there may be a great variety of answers due to failure to understand the type of response desired.

In multiple-choice items, the central problem of the item should be stated so as to make only one choice justifiable, but all choices should seem plausible. Increasing the homogeneity of the choices makes an item more difficult. The negative multiple-choice item, as well as the negative true-false item, tends to measure reading ability, since poor readers often respond to such items incorrectly, even when they know the correct answer.<sup>1</sup>

In compiling a test, there are several factors which must be considered. The test must be planned to fit the fifty- to sixty-minute class period, but should be long enough to be valid and reliable. The length necessary for validity depends upon the number and complexity of the objectives being measured. Reliability, on the other hand, is related more directly to the length of a test than to the objectives, and increasing the length of a test increases the reliability provided the items are of equal quality. Poorly written items added solely to lengthen a test may have the effect of actually decreasing the reliability of the instrument.

A test should generally consist of no more than two or three types of items, and all the items of one type should be included in one section. If numerous instructional objectives are being measured and if they need the use of a number of different kinds of test items, several different tests should be given at different times. Frequent tests can sample smaller units of instruction and thus give a more reliable basis for evaluation.

---

<sup>1</sup>Ibid., p. 36.

The first items on a test should be easy enough for all students to answer. Success in these first items will give some security to a student who approaches the test with a poor test psychology.<sup>1</sup> The student who comes to the test in this frame of mind may very well miss items which he would answer easily were he not in a test situation.

The teacher should take care to make the directions clear and explicit, and there should be a simple method of indicating answers to make them easy to score. Each test should be typed and reproduced by some means so that each pupil can have a copy.

A test should be completely constructed about one week ahead of the date that it is to be administered. It should be left for several days and then examined again for construction errors in the form of: one problem suggesting the answer to another, ambiguous questions or statements, and more than one correct answer to a question.

The evaluation stage. All measuring instruments possess to some degree three important qualities. These are validity, which refers to an instrument's truthfulness; reliability, which refers to its consistency; and usability, which pertains to its practicality.<sup>2</sup>

Validity is extremely important. A test is valid when it actually measures what it is designed to measure.<sup>3</sup> Robert L. Ebel lists ten different types of validity which "are not all distinctly different

---

<sup>1</sup>Ibid., p. 18.

<sup>2</sup>Ibid., p. 84.

<sup>3</sup>Baron, op. cit., p. 15.

from each other."<sup>1</sup> The type of validity which is of most concern to the teacher with each test is the content validity. For classroom tests, this means checking the contents to determine the extent of agreement between the material in the test and the instructional objectives being measured. The best means of determining the content validity of objective tests is to compare the test content to the table of specifications. If the table has been correctly used to construct the test, there should be very high validity.

Another important step in the evaluation of objective tests is item analysis, a check on the quality of each item on the test. Since it requires time, an item analysis should not be done for tests which are carelessly prepared, and which, the teacher feels, do not discriminate between good and poor pupils. In addition to indicating the contribution of separate items to over-all validity, an item analysis also gives information on the difficulty of each item and the discrimination of each item between good and poor pupils. If a higher percentage of good pupils than poor pupils answer an item correctly, that item shows positive discrimination.

For teacher-made tests the following simple procedure gives adequate evidence for determining the quality and difficulty of the items:

1. Select the top 20 to 30 percent and the bottom 20 to 30 percent of the papers.
2. Working with one group at a time, place the papers on a large table overlapping each other so that only the response column of each paper is visible.

---

<sup>1</sup>Ebel, op. cit., p. 381.

3. Count the number of correct responses on each item and convert to percents, which are entered on the record sheet. For example, if 18 out of 20 pupils in the high group get the item correct, 90 percent should be entered on the record sheet for the item.
4. When the correct responses have been counted for both the high and low groups on all items, the power of the item can be determined by calculating the percent in both groups who responded correctly. For example, when 32 of the 40 pupils have succeeded on the item, the ease index is 80 percent. Or, stated another way, the difficulty level is 20 percent because 20 percent failed the item.
5. Calculate the discrimination of each item by using the following formula:

$$D = \frac{U - L}{N}$$

where D = index of item-discriminating power.

U = number of pupils in upper group answering item correctly.

L = number of pupils in lower group answering item correctly.

N = number of pupils in each group.

The discrimination index will range from +1.00 to -1.00, and only items which show positive scores will be retained. A discrimination index above +0.40 is desirable.<sup>1</sup>

After the item analysis is made, the very easy items which were answered correctly by all pupils should be discarded, although a few may be retained to use at the beginning of future tests to help establish a good test psychology for the pupils. Other items which should be discarded are those which do not discriminate between good and poor pupils or those which show negative discrimination, where the poorest pupils are more successful than the best pupils. On multiple-choice test items, there should be a record made of the number of pupils choosing each response. Items which discriminate properly, but have weak distracters can often be rewritten for future use.

---

<sup>1</sup>Green, *op. cit.*, p. 88.



The consistency with which a test measures what it is measuring is called reliability.<sup>1</sup> If the test is reliable, subsequent measurements with the test give approximately the same score each time. As the amount of change in individual pupils will vary, it is difficult to tell whether score differences are a result of low test reliability or of marked pupil change.

Some controllable factors which might increase or decrease the reliability of a test would be the length of the test, the level of difficulty of the test, and the objectivity of the scoring. Within limits, lengthening a test increases its reliability, provided the items added are of equal or better quality than the original items. There is a point, however, where the test becomes so long that fatigue, boredom, or other factors tend to reduce the reliability.<sup>2</sup> Short daily quizzes are completely unreliable unless the scores are accumulated to give one score, so that the cumulative test is long enough to give measurement reliability. In the same manner, a true-false test of less than about forty items is too unreliable to have much measurement use.<sup>3</sup>

The objectivity of the scoring affects reliability since the objectivity in a test eliminates the personal opinion or judgement of the scorer. If the scorer's personal opinion were to enter in, the reliability would decrease since his opinion may differ greatly between test papers, being influenced by such factors as the personality of the examinee, the neatness of the paper, or the method used to arrive at

---

<sup>1</sup>Baron, op. cit., p. 20.

<sup>2</sup>Green, op. cit., p. 96.

<sup>3</sup>Ibid., p. 97.



the answer to the question. The reliability is thus greatest with the objective-type test.

Three different reliability coefficients are often obtained to check the reliability of a test. These are the coefficient of stability, the coefficient of equivalency, and the coefficient of internal consistency.<sup>1</sup> For each of these coefficients, the Pearson method is used to determine the correlation coefficient.

Correlation coefficient, Pearson method:<sup>2</sup>

$$r = \frac{\sum x_1 x_2}{\sqrt{(\sum x_1^2)(\sum x_2^2)}}$$

where:  $r$  = sample correlation coefficient.

$x_1$  = deviation from the mean of the first variable.

$x_2$  = deviation from the mean of the second variable.

Reliability may be checked using the test-retest method, where the same test is administered a second time after a lapse of time of from one to six months. This method gives a coefficient of stability since it is based on the stability of the performance of a group in a retest situation. The scores from the first and second administrations are then correlated to determine reliability. Using the Pearson formula,  $x_1$  is the deviation of each student's score from the mean of the first test, and  $x_2$  is the deviation from the mean of the second test. The coefficient should be at least 0.85 for good test reliability.

---

<sup>1</sup>Ibid., p. 93.

<sup>2</sup>George W. Snedecor, Statistical Methods (Ames, Iowa: Iowa State University Press, 1956), p. 160.

A coefficient of equivalence is determined by administering two equivalent forms of the same test to a group at the same time, with one-half of the group taking each form. With this alternate-forms method both the equivalency of the test forms and the stability of the group may be checked if the second form is administered some time later to the same group. The alternate-forms method is most useful to the person administering the test. It is important to have more than one form of a test with high reliability because, in many instances, pupils have to make up the test for having missed it due to illness or other circumstances.

In the split-half method of determining reliability there is one group, one test form, and one test administration. With this method the Spearman-Brown formula is commonly used, and the pupils' scores on the odd-numbered test items are correlated against their scores on the even-numbered items. The split-half method gives a coefficient of internal consistency in that it compares the rank and score on one half of the test against their rank and score on the other half. Reliability is affected adversely by the shortening to one-half.

Spearman-Brown formula:<sup>1</sup>

$$r_{xx} = \frac{2r_{oe}}{1 + r_{oe}}$$

where:  $r_{xx}$  = coefficient of internal consistency of the total test.

$r_{oe}$  = coefficient of correlation between pupils' odd-half scores and their even-half scores.

---

<sup>1</sup>Green, op. cit., p. 95.

When an item analysis has been run on a test and the difficulty value or the proportion of pupils responding correctly to each item has been obtained, the Kuder-Richardson formula may be used for calculating internal consistency. This will simplify the number of calculations necessary overall.

Kuder-Richardson formula:<sup>1</sup>

$$r_{xx} = \left( \frac{N}{N-1} \right) \left( 1 - \frac{pq}{s_t^2} \right)$$

where: N = number of items.

p = proportion (per cent of pupils answering item correctly).

q = 1 - p.

$s_t^2$  = standard deviation of total test squared.

For a test to have a high degree of usability, it should be easy to administer and score, economical in time and material, and have a clear format. The usability of a test is determined by inspection, and is done by the teacher before, during, and after the administration of each examination.

#### THE APPLICATION OF TEST CONSTRUCTION TO HIGH SCHOOL MODERN ALGEBRA

Modern algebra in the high school has included several topics which are unfamiliar to many students. Set theory, proofs, and fundamental axioms are used as tools in developing the familiar concepts of algebra, and the teacher's testing program must be designed with this in mind.

---

<sup>1</sup>Ibid., p. 96.

### Planning the Test

The developing of a test begins the first time a lesson is planned for the particular unit. In following the rules for test construction, the teacher must have a number of specific and general educational objectives toward which the instruction is aimed. For the tests to measure these objectives, test questions should be written each time a lesson is planned, following the test-specification chart to determine the approximate number of questions needed.

Objectives. The teacher's instructional objectives should be stated in specific, operational terms, to form the basis for planning a test. A different set of objectives would probably be developed by every mathematics teacher, with no list being entirely accepted by all. In order to have a frame of reference, Max A. Sobel and Donovan A. Johnson have presented the following sample list of objectives for use in any mathematics course.

The student should:

- have a knowledge and understanding of mathematical processes, facts, and concepts;
- have a skill in computing with understanding, accuracy, and efficiency;
- have the ability to use a general problem-solving technique;
- understand the logical structure of mathematics and the nature of proof;
- use mathematical concepts and processes to discover new generalizations and applications;
- recognize and appreciate the role of mathematics in society;
- develop study habits essential for independent progress in mathematics;
- develop reading skill and vocabulary essential for progress in mathematics;

- demonstrate such mental traits as creativity, imagination, curiosity, and visualization;
- develop attitudes that lead to appreciation, confidence, respect, initiative, and independence.<sup>1</sup>

This list, of course, is not exhaustive. It may be possible to set up objectives which do not fall into any of the preceding categories, but this list does indicate the possibilities. As an example, the objectives for the first chapter in a course in modern algebra might be the following:

The student should:

- have a knowledge and understanding of mathematical processes, facts, and concepts;
- recognize the practical application of specific mathematical processes, facts, and concepts;
- have skill in computing with understanding, accuracy, and efficiency.

Having these three general objectives now stated in writing, they must be used in conjunction with the content areas, to determine the specific, operational objectives. For the purpose of this example, the content areas of the first chapter in a course in modern algebra will be the following:

<u>UNIT</u>	<u>CONTENT AREA</u>
1	Operations, relations, and symbols
2	Axioms and properties of numbers
3	Order of operations

---

<sup>1</sup>Max A. Sobel and Donovan A. Johnson, "Analysis of Illustrative Test Items," Evaluation in Mathematics, Twenty-sixth Yearbook of the National Council of Teachers of Mathematics (Washington, D. C.: National Council of Teachers of Mathematics, 1961), p. 72.

4	Inverses
5	Algebraic sentences and equations
6	Inequalities
7	Formulas
8	Definitions
9	Sets

With these areas in mind the three general objectives could be rewritten in more precise terms as follows:

The student should:

- have a working knowledge and an understanding of the definitions;
- be able to recognize the application of the given facts and concepts;
- be able to use mathematical symbols in the application to problems;
- be able to apply mathematical processes to discover solutions to problems;
- have skill in computing with accuracy and efficiency.

In this revised list, the first objective is a restatement of the first general objective. The second and third objectives are taken from the second general objective, and the remaining two objectives of this list are from the third general objective.

The test-specification chart. With the objectives now stated in precise, operational terms, the content areas can be combined with the objectives to establish the test-specification chart. Using the content areas in the left-hand column and the objectives across the top,

the two-way table is formed. The chart is then completed by choosing values for each pair of objectives and content areas. Table III is the test-specification chart thus formed.

A teacher planning for this chapter, would be able to see immediately that units two, three, four, and five are going to require the most emphasis, since the total percentages are greatest in these units. He can also notice in the same manner, that "be able to apply mathematical processes to discover solutions to problems" is the most important objective. Unit eight, definitions, has only one objective and can thus be easily made into a short measurement test, or quiz, avoiding the necessity to have an additional unit on the chapter test. Unit seven and unit one comprising only 4 per cent and 7 per cent, respectively, of the total might also be better tested with short measurement tests. In the same manner, the objective, "have a working knowledge and an understanding of the definitions" has only 10 per cent of the total and, in order to limit the size required for the chapter test, this objective might be tested with one or more short measurement tests. Choices for each of these possibilities rest solely upon the teacher's judgement in the particular situation.

In previewing the chapter, the teacher might expect the students to have extra difficulty, for instance, in units two, four, six, and nine. With this in mind, he could then plan to give instructional tests during the study of any or all of these units. Since the grades would not count heavily, this test would not preclude the necessity for a measurement test over the same material.

At the end of units two and five a mastery test is a possibility since the pupils will have to know certain parts of each of these to

TABLE III

TEST-SPECIFICATION CHART FOR THE FIRST CHAPTER OF  
A COURSE IN HIGH SCHOOL MODERN ALGEBRA

UNIT	CONTENT AREAS	OBJECTIVES					Total
		The student should:					
		have a working knowledge and an understanding of the definitions	be able to recognize the application of the given facts and concepts	be able to use mathematical symbols in the application to problems	be able to apply mathematical processes to discover solutions to problems	have skill in computing with accuracy and efficiency	
1	Operations, Relations, and Symbols	0	2	5	0	0	7
2	Axioms and Properties of Numbers	0	10	0	10	0	20
3	Order of Operations	0	0	0	5	10	15
4	Inverses	1	2	5	5	0	13
5	Algebraic Sentences and Equations	2	0	5	5	5	17
6	Inequalities	2	0	5	0	2	9
7	Formulas	0	2	0	2	0	4
8	Definitions	5	0	0	0	0	5
9	Sets	0	5	0	5	0	10
	<b>Total</b>	10	21	20	32	17	100



proceed to the units which follow. This test would most likely be a short test, covering very little material, and one in which scores are high. Since the first four units constitute about half the chapter, according to the relative weights, the teacher might want to have two large measurement tests, depending upon the importance of this chapter in relation to the entire course. Here, again, a test-specification chart for the entire course would prove instrumental in determining the relative importance of each chapter.

At this point, it would be good to have a list giving a clear and brief outline to the teacher of the tests he is considering, with the relative weights of each unit listed as a guide to determining the value of the tests. The following list would be representative of the example being used.

<u>UNIT</u>	<u>CONTENT AREA</u>	<u>TEST</u>	<u>RELATIVE WEIGHT</u>
1	Operations, relations, and symbols	Measurement	7
2	Axioms and properties of numbers	Instructional and/or mastery	20
3	Order of operations	None	15
4	Inverses	Instructional	13
5	Algebraic sentences and equations	Mastery	17
6	Inequalities	Instructional	9
7	Formulas	Measurement	4
8	Definitions	Measurement	5
9	Sets	Instructional	10

Since too many tests are indicated here to be able to include all of them, it is necessary for the teacher to judge which tests to include largely by the specific material and the particular classes involved.

### Constructing the Test

When the lessons have been planned by using the objectives and content areas to determine the relative values of each, the task of the test construction is simplified considerably. Items for the test are written as each lesson is planned, using the test-specification chart as a guide to the number and type of questions needed. At testing time, then, all that remains is to compile the questions into the necessary test and reexamine each item for item-writing errors.

The item-writing phase. Three types of items are written for tests, corresponding to the three types of examinations. These are of the oral, objective, and essay types, as previously mentioned. The performance test which was also mentioned, is primarily used in the fields of Home Economics, Industrial Arts, and Business, where a coordinated performance connected with a product is involved. It is not intended to replace a pencil-and-paper test, but is intended rather to test those areas which a pencil-and-paper test will not adequately evaluate.

In analyzing the test-specification chart, it can be seen that neither the content areas nor the objectives require the particular characteristics of an oral examination. The oral questioning technique, on the other hand, would be an excellent instructional device where the instructional test has been indicated on the list of tests. With the use of the oral questioning technique, grades are not counted heavily, since the purpose is that of instruction and not measurement. Referring again to Table III, the objectives, "have a working knowledge and an understanding of the definitions," and "be able to recognize the

application of the given facts and concepts," would appear to fit most readily into the oral questioning technique. Having these two references as guides, oral questioning is indicated for units two, four, six, and nine, with only one objective being tested in units two, six, and nine, and both objectives being tested in unit four.

Unit two, axioms and properties of numbers, would lend itself very well to oral questioning. For instance, an axiom which says, "if the same quantity is added to equal quantities of the same kind, the sums are equal," can be tested orally with problems which are aimed at having the student recognize the axiom's applications. The axiom can be illustrated with this problem:

EXAMPLE 1.

If two girls each have five dollars and they receive two dollars apiece for babysitting, they each now have seven dollars. This is an example of which axiom?

Another item for oral questioning would be the following:

EXAMPLE 2.

If the side of one square is equal to the side of another square, do the squares have equal areas? If two squares have equal areas, is the side of one square equal to the side of the other? If one number equals another, are their squares equal? Are their cubes equal? What is the axiom illustrated by these problems?

Questions such as this on axioms and properties of numbers will permit the teacher to draw out the full meaning of the axiom when the student's

answer is vague or incomplete. Oral questioning in this area will assist the teacher in diagnosing misunderstandings which the pupils have about particular rules. If the axiom, "the same powers of equal quantities are equal," were being sought by the teacher, and the answer a pupil gave was, "the powers of equal quantities are equal," a basic misunderstanding is apparent. The teacher, at this point, could give a counter-example and probably elicit the correct answer. Some obvious problems which could be used to point out the error in the student's statement would be, "Is  $a^2$  equal to  $a^5$ ?" and "Is  $2^2$  equal to  $2^5$ ?"

For this unit, on axioms and properties of numbers, the "discovery method" could be quite profitably used. In Example 2, if the final question were changed to "What rule could you make from these illustrations?", the pupils would be trying to discover an axiom which they would encounter in the next lesson in the text.

The use of lengthy oral examinations in mathematics would necessarily be limited, but as a questioning technique, some measurement is possible. The greatest value is as an instructional device, where grades are not counted heavily.

Referring to Table III, the test-specification chart, any of the content areas and any of the objectives could be tested using the objective type test. Each of the four common forms of the objective test, the short-answer, the alternate-response, the multiple-choice, and the matching form is applicable at some point in the chart and can be used at the discretion of the teacher.

The short-answer form may include questions, definitions, incomplete sentences, or identification items. An example of the definition item would be the following illustration:

## EXAMPLE 3.

Define:

- A. null set
- B. solution set of an algebraic sentence

This type of item borders very closely to the essay test in the difficulty of scoring objectively. A written answer of this nature would have to have a strict grading system, in order to have objective scoring; there would have to be one point for each thought the teacher expected the student to express. For instance, in Example 3, for the null set, there might be two points sought:

<u>Point</u>	<u>Definition</u>
1	A set
1	having no members.

and for the solution set:

<u>Point</u>	<u>Definition</u>
1	The set
1	of all numbers
1	which make an algebraic sentence true.

These definition items could easily be considered as question items, by rewriting them in this manner:

## EXAMPLE 4.

What is the definition of a "null set"?

## EXAMPLE 5.

What is the "solution set" of any algebraic sentence?

Another example of the question form is the type of question which asks for only one or two specific words. Illustrations of this would be:

EXAMPLE 6.

What property says that the sum of any two numbers is a number?

EXAMPLE 7.

What is the inverse of the operation "squaring"?

EXAMPLE 8.

What is the name of the set having no members?

The incomplete sentence form of the objective test is quite similar to the question form, in that there is usually only one or two specific words which will correctly answer the item. The incomplete sentence can be used also in each of the above questions in the following manner:

EXAMPLE 9.

"The sum of any two numbers is a number" is an example of the \_\_\_\_\_ property.

EXAMPLE 10.

The inverse of squaring is called \_\_\_\_\_.

EXAMPLE 11.

The name of the set having no members is the \_\_\_\_\_ set.

Or in another example:

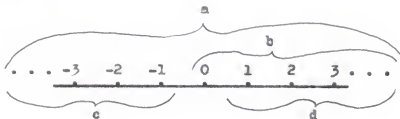
EXAMPLE 12.

By the transitive axiom, if  $a = b$  and  $b = c$ , then \_\_\_\_\_.

Another short-answer type, closely related to the incomplete sentence form, is the identification form. This form is most practical in items where a figure can be shown, with each of the parts labeled which are to be identified. There is one particular case where an identification item would be useful in the chapter outlined on Table III: the number line. An identification item would permit the teacher to test the student's knowledge of the names of portions of the number line without having to write separate questions for each. It could be designed like this:

**EXAMPLE 13.**

Identify:



where the answers are:

- a. integers
- b. whole numbers or non-negative integers
- c. negative integers
- d. natural numbers or positive integers

Also coming under the classification of the short-answer form is the familiar problem-solving, where the student is to derive a specific numerical answer from a given problem. Again, as in the essay question, the problem arises as to how much partial credit to allow for a partially correct solution. If the problem is short and relatively

simple, there might be one point given, that being for the correct answer. However, when the problem is more complex, a point should be allowed for specific steps of the problem which the student must obtain to have the correct answer. Consider the problem  $5x = 15$ . In this problem, there would be little value in allowing more than one point. In the following problem, however, there would be points given for certain steps.

EXAMPLE 14.

Let  $Z = \{\text{integers}\}$ .

If  $x \in Z$ , find  $\{x \mid 2(x + 4) = 5x + 2 - x\}$ .

Points for this problem could be distributed as follows:

<u>Points</u>	<u>Steps</u>	<u>Solution</u>
		$2(x + 4) = 5x + 2 - x$
1	$2(x + 4) = 2x + 8$	$2x + 8 = 5x + 2 - x$
1	$5x - x - 2x = 2x$	$8 = 2x + 2$
1	$8 - 2 = 6$	$6 = 2x$
1	$\frac{6}{2} = 3$	$3 = x$ , or $x = 3$

Thus, this problem may become a four-point problem. The student must expand the term with parentheses, he must collect both the  $x$ -terms and the constants, and he must divide by two to obtain the result. This type of problem-solving would probably be the best examination form to use for objectives having to do with skill in computing.

The alternate-response form, since it is normally used where there are only two possible answers, will find very limited use in mathematics. From Table III, a unit which might be tested by means of



true-false questions is unit two, axioms and properties of numbers. In this case, a modified true-false item where the student must correct the underlined word if the statement is false, would be best since there are numerous axioms and properties, and not just two choices. Examples of the modified true-false questions would be:

EXAMPLE 15.

The statement  $2 = 2$  is an example of the symmetry axiom.

EXAMPLE 16.

If  $2^2 = 4$ , and  $4 = \frac{8}{2}$ , then  $2^2 = \frac{8}{2}$ . This statement is an example of the replacement axiom.

Example 15 would be false until "symmetry" is changed to "reflexive." Example 16 would be corrected to the "transitive" axiom. It is important to notice that only the important words have been underlined, since these are the words the student must learn and remember. Modified alternate-response items, such as these, require the student to be able to recognize the falsity of the statement and to supply the word which will correct it. True-false items are not as discriminating, item for item, as the other objective test forms and are somewhat more subject to ambiguity and misinterpretation.

The multiple-choice item is more difficult to prepare than the alternate-response form but is good for measuring a variety of educational objectives. The multiple-choice items have the most instructional value if the choices are the answers a student would obtain by making certain common errors. These choices can be illustrated in this manner:

## EXAMPLE 17.

The area (in square inches) of the region bounded by a circle with a radius of 3 inches is:

- A.  $6\pi$
- B.  $9\pi$
- C.  $18\pi$
- D.  $36\pi$

Answer A is the circumference of this circle,  $2\pi r$ , which could not be in square inches. This fact is not likely to be a specific determiner that would eliminate this choice, since students frequently confuse the circumference formula and the area formula and tend to ignore the units of measure. Choice B is  $2\pi r^2$ , which is a cross between the area formula and the circumfe. enc. formula. Choice C is correct, and D is  $\pi d^2$  where  $d^2$  has been substituted for  $r^2$  in the correct formula. An example in a unit on sets would be the following:

## EXAMPLE 18.

If  $M = \{1, 2, 3, 4\}$  and  $N = \{3, 4\}$ , what does  $M \cup N$  equal?

- A.  $\{1, 2, 3, 4\}$
- B.  $\{3, 4\}$
- C.  $\{1, 2, 3, 4, 3, 4\}$
- D.  $\{1, 2, 3, 4\} + \{3, 4\}$
- E.  $\{17\}$

Choice A is correct; B is  $M \cap N$ ; C has a duplication of like terms, a mistake often made by students when first studying modern algebra; D is

$M + N$ ; and  $E$  is found by adding  $1 + 2 + 3 + 4 + 3 + 4$ . Another example would be the following, where the wrong choice indicated the student's misunderstanding of a definition.

EXAMPLE 19.

If one quantity equals a second quantity, then the second equals the first. This statement is the \_\_\_\_\_ axiom.

- A. power
- B. reflexive
- C. symmetry
- D. transitive

The matching form of objective test is most useful for measuring recognition and recall. This can best be illustrated by using part of a list of names and symbols found in the unit on sets from Algebra One by Glen D. Vannatta.<sup>1</sup> It should be noted that the names and symbols of Example 20 are for use in the specific text cited and are not universal in every instance.

Several distracters are included in this list of symbols, such as "not contained in," "contains," and "universal set." Thus a student may choose G as the answer to 2 and H as the answer to 3, or he may choose I as the answer to 2 and J as the answer to 3, since either of these pairs is consistent in form.

---

<sup>1</sup>Glen D. Vannatta, A. Wilson Goodwin, and Harold P. Fawcett, Algebra One: A Modern Course (Columbus, Ohio: Charles E. Merrill Books, Inc., 1962), p. 50.

## EXAMPLE 20.

## Matching:

<u>Name</u>	<u>Symbol</u>
1. Set	A. S
2. Member	B. U
3. Not a member	C. $\cup$
4. Is contained in	D. $\cap$
5. Set notation	E. $\{   \}$
6. Solution set	F. $\{ \}$
7. Union	G. $\in$
8. Intersection	H. $\notin$
	I. C
	J. $\phi$
	K. $\supset$

Generally speaking, the objective test will be the most commonly used test in any mathematics, due to the adaptability to the subject and the ease in scoring. However, the essay form is becoming of more importance in modern algebra due to the increasing emphasis on proofs. A proof is necessarily classified as an essay item since the student must supply the steps from his own background of knowledge and organize them until they are logical.

## EXAMPLE 21.

Prove that  $(a + b) + [(-a) + (-b)] = 0$ .

Proof:

$(a + b) + [(-a) + (-b)] = 0$	Given
$[(a + b) + (-a)] + (-b) = 0$	Associative property
$[(-a) + (a + b)] + (-b) = 0$	Commutative property
$\{[(-a) + a] + b\} + (-b) = 0$	Associative property
$(0 + b) + (-b) = 0$	Inverse property
$(0) + [b + (-b)] = 0$	Associative property
$[b + (-b)] = 0$	Identity property
$0 = 0$	Inverse property

The compiling phase. If the test items have been written when lessons were planned, the task remaining is to compile these items into a test. Since testing is done for various purposes, the compiling must be aimed at satisfying one of the three test functions: instruction, mastery, or measurement.

In compiling the instructional test for unit nine outlined in Table III, ten multiple-choice items might be best. These items should be specifically designed to include the wrong answers which pupils will obtain if they make any of the most common errors, in addition to the one correct answer. This one example is not intended to show the only possible type of test to be used here, but rather to illustrate the fact that the items must be chosen to satisfy a particular purpose.

A mastery test for unit five, since the students are expected to answer 90 per cent correctly, would probably include only problem-solving items. Students would be given equations and asked to solve for

the value of the variable, or given written problems and asked to apply equations to the solving of the problems. In either case, the intended purpose of the test would be to assure the teacher as well as the students that the skills in the unit had been mastered.

The measurement test is the test which is the primary grade-determiner, since it purports to measure a student's learning for the grading period. The short measurement test will not be very reliable unless the scores are accumulated throughout the period to give one total score approximately equivalent to a one-hour test. Test items on the measurement tests encompass all of the types discussed in the item-writing phase.

#### Evaluating the Test

Once the test has been administered, the results should be used to evaluate its effectiveness. All measuring instruments possess to some degree three important qualities: validity, reliability, and usability. These qualities apply whether the test is in mathematics or any other subject, and the procedures and formulas are also the same.

If a teacher's testing program is to improve from year to year, the items which have been used on each test must be analyzed and either revised or discarded. Items which should be discarded are those which do not discriminate between good and poor pupils or those which show negative discrimination, where the poorest pupils are more successful than the best pupils.

## SUMMARY

The teacher's instructional purposes will generally fall into three categories of test functions. A test may function as an instructional test, which checks skill; as a mastery test, which evaluates a pupil's knowledge of fundamentals; or as a measurement test which measures his understanding of concepts. These test functions are each satisfied by using any of four types of tests: the oral, objective, essay, or performance test.

There are three stages necessary to develop a high-quality test: the planning stage, the construction stage, and the evaluation stage. In planning a test, the instructional objectives and the content areas together make up the test-specification chart, which will determine the choice of test form and the type and number of items. The construction of a test is usually handled in two phases: the item-writing phase and the compiling phase. The evaluation stage involves an analysis of the test results to serve as a basis for improving the items on the test. This is necessary in order that they may be revised on tests which are given in subsequent years. The tools which are used for the analysis are validity, which refers to the test's truthfulness; reliability, which refers to its consistency; and usability, which pertains to its practicality.

The general principles of test construction apply to essentially all subjects being taught in high school. The specific application of these principles to high school modern algebra is not a matter of developing new principles, but rather one of using the new approaches of the modern algebra in writing individual items.

Modern algebra in the high school has included set theory, proofs, and axioms from which the familiar concepts of algebra are developed. The discovery method is used, where a student will solve problems and from these state a rule, and this same method can be used on instructional tests to give extra aid to the student.

A carefully planned testing program will not only determine a student's grade, but will also aid the teacher in his test-making task in future years. With careful analysis and revision of items a teacher's classroom tests will contribute additional instruction, and will continue to increase in effectiveness from year to year.



## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Adams, Georgia S. Measurement and Evaluation. New York: Holt, Rinehart and Winston, 1964. Pp. 321-362.
- Baron, Denis, and Harold W. Bernard. Evaluation Techniques for Classroom Teachers. New York: McGraw-Hill Book Company, 1958.
- Bean, Kenneth L. Construction of Educational and Personnel Tests. New York: McGraw-Hill Book Company, 1953. Pp. 37-106.
- Bradfield, James M. Measurement and Evaluation in Education. New York: The Macmillan Company, 1957. Pp. 309-327.
- Brown, Kenneth E. "Research in the Teaching of High School Mathematics," The Mathematics Teacher, 59:53-56, January, 1966.
- Cahan, Leonard S. "An Interim Report on the National Longitudinal Study of Mathematical Abilities," The Mathematics Teacher, 58:522-526, October, 1965.
- Doll, Ronald C. Curriculum Improvement: Decision-making and Process. Boston: Allyn and Bacon, Inc., 1964. P. 24.
- Ebel, Robert L. Measuring Educational Achievement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.
- Furst, Edward J. Constructing Evaluation Instruments. New York: Longmans, Green, Courtesy McKay, 1958. Pp. 203-327.
- Green, John A. Teacher-made Tests. New York: Harper and Row, Publishers, 1963.
- Hart, William L. Algebra, Elementary Functions, and Probability. Boston: D. C. Heath and Co., 1965.
- Linquist, E. F. (ed.). Educational Measurement. Washington, D. C.: American Council on Education, 1941. Pp. 146-151.
- Mayor, John R., and Marie S. Wilcox. Algebra, First Course. Second edition. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1961.
- Mayor, John R., and Marie S. Wilcox. Algebra, Second Course. Second edition. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1961.
- Nelson, Martin J., E. C. Denny, and Arthur P. Coladarci. Statistics for Teachers. New York: Henry Holt and Company, 1956.
- Payne, Holland. "What about Modern Programs in Mathematics?" The Mathematics Teacher, 58:422-424, May, 1965.

- School Mathematics Study Group. Intermediate Mathematics, Part I.  
Unit seventeen. New Haven: Yale University Press, 1961.
- School Mathematics Study Group. Intermediate Mathematics, Part II.  
Unit eighteen. New Haven: Yale University Press, 1961.
- Smith, Rolland R., Francis G. Lankford, Jr., and Joseph N. Payne.  
Contemporary Algebra, Book One. New York: Harcourt, Brace, and  
World, Inc., 1962.
- Smith, Rolland R., Francis G. Lankford, Jr., and Joseph N. Payne.  
Contemporary Algebra, Book Two. New York: Harcourt, Brace, and  
World, Inc., 1963.
- Snedecor, George W. Statistical Methods. Ames, Iowa: Iowa State  
University Press, 1956. P. 160.
- National Council of Teachers of Mathematics. Evaluation in Mathematics.  
Twenty-sixth Yearbook. Washington, D. C.: National Council of  
Teachers of Mathematics, 1961.
- Vannatta, Glen D., A. Wilson Goodwin, and Harold P. Fawcett. Algebra  
One: A Modern Course. Columbus, Ohio: Charles E. Merrill Books,  
Inc., 1962.
- Vannatta, Glen D., A. Wilson Goodwin, and Harold P. Fawcett. Algebra  
Two: A Modern Course. Columbus, Ohio: Charles E. Merrill Books,  
Inc., 1962.
- Wood, Dorothy A. Test Construction. Columbus, Ohio: Charles E. Merrill  
Books, Inc., 1960.

75

DEVELOPING CLASSROOM TESTS FOR HIGH  
SCHOOL MODERN ALGEBRA

by

RICHARD JAY JEWETT

B. S., Kansas State University, 1964

---

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

College of Education

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

1966

It was the purpose of this study to (1) present the general principles of test construction as they are commonly found in texts on measurement and evaluation, and (2) offer methods of applying these principles to high school modern algebra.

The method of carrying out this study of test construction for modern algebra was reading and library research. Only literature pertaining to classroom tests in mathematics and references on general measurement techniques which were available in the Kansas State University Library were investigated. The general information was then applied to high school modern algebra by the author.

There is no one best test; it is for the test-maker to construct the best test for his instructional purposes. The teacher's instructional purposes will generally fall into three categories of test functions. A test may function as an instructional test, which checks skill; as a mastery test, which evaluates a pupil's knowledge of fundamentals; or as a measurement test which measures his understanding of concepts.

These test functions are each satisfied by using any of four types of tests: the oral, objective, essay, or performance test. Each of these types of tests must be considered for its strengths and weaknesses as well as for the function it is to serve. The oral test is valuable in that it improves test rapport for pupils who fear written examinations, but its weakness is that it is too time-consuming. The objective test can be adapted to several teaching objectives, and still be made highly valid and reliable. Its disadvantage is that it is difficult to prepare. The essay test, although it favors the verbally

inclined student and is difficult to grade, is easy to construct and it does promote the proper type of study. Where some skills and abstract abilities not measured by other conventional test forms need to be measured, the performance test is in order. This test is not adaptable to many fields of learning and it is often difficult to grade.

There are three stages necessary to develop a high-quality test: the planning stage, the construction stage, and the evaluation stage. In planning a test, the instructional objectives and the content areas together make up the test-specification chart, which will determine the choice of test form and the type and number of items. The construction of a test is usually handled in two phases: the item-writing phase and the compiling phase. Item-writing is not such a tedious task if the teacher will write several test items every time he plans a lesson. This serves a dual purpose of directing the testing toward the instructional objectives, and of easing the difficulty in developing a test at the end of a unit. The test can then be compiled from these previously written items and examined for common item-writing errors. The evaluation stage involves an analysis of the test results to serve as a basis for improving the items on the test. This is necessary in order that they may be revised on tests which are given in subsequent years. The tools which are used for the analysis are validity, which refers to the test's truthfulness; reliability, which refers to its consistency; and usability, which pertains to its practicality.

The application of these principles to modern algebra involves developing the instructional objectives in specific, operational terms. These objectives, together with the content areas of a unit form the

basis for the test-specification chart, which will determine the type of test.

It should be noted that in modern algebra special attention is given to sets and to the proofs, axioms, and postulates which formerly have been limited to geometry. The test items in modern algebra are aimed at fostering these concepts by testing the applications where possible. Modern algebra has presented a new approach to an old high school subject and it follows that the classroom tests must present a corresponding approach if the testing program is to aid in the total education of the individual pupils.

A carefully planned testing program will both determine the pupils' grades, and assist the teacher in improving his test items for future years. With a good analysis and revision of items a teacher's classroom tests will contribute additional instruction, and will continue to increase in effectiveness from year to year.