ED 304 445   TM 012 639

AUTHOR	Gilman, David Alan
TITLE	Improving Schools through Inservice Test Construction: The Rossville Model.
PUB DATE	Nov 88
NOTE	35p.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Achievement Tests; Basic Skills; Computer Assisted Testing; Curriculum Development; Delphi Technique; *Educational Improvement; Elementary Secondary Education; *Faculty Development; *Inservice Teacher Education; *Minimum Competency Testing; Norm Referenced Tests; *Teacher Made Tests; *Test Construction; Testing Programs
IDENTIFIERS	Competency Based Assessment; *Competency Rossville Model; Rossville Alvin School District IL

ABSTRACT

A method for improving curriculum and schools through the local development of competency tests in basic skills--the Competency-Rossville Model (CRM)--is outlined. The method was originated in the school system of Rossville (Illinois) and has been tested in five other midwestern school systems. The approach leads the faculty of the school, with the guidance of a measurement consultant, in the development of a series of grade-level tests to measure mastery in basic skills achievement. This method of test development serves to articulate the curriculum; provides a useful, relevant, and appropriate achievement testing program; and provides a management system for the improvement of basic skills instruction. The CRM is compared favorably to the norm referenced testing model. The CRM program includes: (1) inservice instruction of faculty in an alternate model of evaluation of academic achievement; (2) a committee of school faculty formed around use of the Delphi method to confer with their fellow teachers; (3) grade-level representatives in charge of informal committees for each grade level; (4) informal teacher conferences, involving each grade level, to establish a list of skills to be mastered by students; (5) development of test items; and (6) computer-assisted interpretation of tests. This type of inservice development project has been successful in increasing faculty morale, improving basic skills instruction, and improving school achievement testing programs. The project involves faculty actively in curriculum development, results in valid and reliable tests, and provides information that is valuable and useful to teachers. Nine figures are provided, and a sample skill-referenced math test is appended. (TJH)

Improving Schools Through Inservice
Test Construction: The Rossville Model

by

David A. Gilman, Ph. D.

Professor of Education

Indiana State University

Terre Haute, Indiana.

November, 1988

ABSTRACT

Many articles and several books have been written which desribe the shortcomings of the classical psychometric approach to achievement testing.

Among the complaints concerning testing is that the measures are not valid or appropriate for the kinds of learning that is happening in schools today. Such tests are said to be biased, and the information that they provide is not useful because it is often filed and forgotten.

This presentation describes a tried and proven method for improvinig schools through the local development of competency testing in the basic skills. The method was originated in the school system of Rossville, Illinois and has been tested in five other midwestern school systems.

The plan is to lead the faculty of the school with the guidance of a measurement consultant in the development of a series of grade level tests to measure mastery in basic skills achievement. This method of test development serves to articulate the curriculum, to provide a useful, relevant, and appropriate achievement testing program, and to provide a management system for the improvement of basic skill instruction.

The program begins with the inservice instruction of faculty by teaching them an alternate model of the evaluation of educational achievement. A committee of school faculty is formed and then utilizes the Delphi method to confer with their fellow teachers. Grade level representatives are in charge of an informal committee for their respective grade level. Teachers confer informally within each grade level and establish a list of skills that they agree students should have mastered at the end of the grade level they are teaching. These skills lists are in the form of just the behavior part of the behavioral objectives for that grade level.

Test items are developed to test the skills. The test development phase of the project utilizes item pools. Items are developed from the item pool to form competency tests in the basic skills at each grade level. Test questions are designed with the desired objective that at least 70 per cent of the students will answer at least 70 per cent of the items correctly for each of the skills that are measured. When this goal is not met, each unmastered skill is examined to determine whether the test item, the objective, or the instruction needs to be improved for the next class at that grade level.

Tests are scored and results are interpreted by computer. Reports of each student's progress are provided to each teacher and to the child's parents.

This form of test development has the following advantages:

1. Faculty are actively involved in curriculum development.
2. The tests are valid, reliable, and appropriate.
3. The tests provide information that is valuable and useful to teachers rather than just filed forgotten, the way the results of standardized tests are.

Teachers enthusiastically support this method of testing. Parents support the wealth of information that the reports they receive provide about their children's progress in school. Of the six school systems that have attempted this program, all are presently continuing it and most are planning on expanding it.

This type of inservice development project has been successful where it has been tried in increasing faculty morale, improving basic skill instruction, and improving the achievement testing program in the school. Attempts will be made to extend the program to school subjects other than the basic skills and to provide additional benefits within the program.

ii 4

Improving Schools through Inservice Test

Construction: The Rossville Model

by David Alan Gilman, Ph.D.

Indiana State University

Education has its share of good guys (programs which generally merit positive comments) such as the school lunch program, basic skills instruction, and gifted-talented programs.   However, there is also an adequate supply of bad guys (programs which draw more blame than praise).   Among these are merit pay, vandalism, poor discipline, and standardized tests.

Although it is recognized that teachers could utilize the information that achievement testing provides, there has been a constant barrage of criticisms directed toward the classical psychometric model as it is applied to educational achievement testing. This model for testing provides a basis for the simple ranking of students from high to low.   The criticisms have not been directed toward the capability of the model to accomplish such a sorting, but rather have been directed toward the educational outcomes that occurs as a result of such rankings.   For some time now, educators have been questioning the amount of time that schools spend ordering and sorting students from high to low according to their various abilities.

The critics of this type of testing have been so vocal that they have succeeded in having all standardized intelligence testing removed from the New York City Public Schools and their protests has caused the National Education Association to recommend a complete moratorium on all standardized testing in U. S. schools.

Table I contains a summary of the criticisms that have been directed toward the classical psychometric model as it is applied in norm referenced achievement tests.

Table I

Criticisms of the Classical Psychometric Approach to Achievement
Testing

Substantive Issues

Invalid
Biased
Not useful to teachers
Tests define the curriculum
Filed and forgotten
Inaccurate
Wastes time
Misunderstood
Results are uninterpretable
No management system

Humanistic Issues

Degrading
Labels students
Destructive competition
Promotes dishonesty
Impersonal
Unfair
Puts pressure on students

Critics of standardized testing claim that such tests are invalid because skills measured on these tests are often not the ones taught in the classroom.

Although test publishers try to combat it, their tests are constantly criticized for being culturally biased against various minority groups.

The results of these tests only serve to rank students from high to low and consequently the results are not of any particular use to teachers or to school administrators in helping students overcome their specific learning difficulties.

Some educators believe that these tests serve to influence their school's curriculum in ways that infringe on the autonomy of the faculty and/or the local school board. The content of standardized tests is determined in such places as Iowa City, Iowa or Princeton, New Jersey and does, in some instances, exert a direct influence on what is taught in local school districts.

Since most teachers do not understand the intricate contingencies that are involved in the classical psychometric model of testing, they do not understand how the tests are to be utilized and do not know how the results are interpreted.

The scores of standardized tests are derived in such a way as to indicate a rather abstract relationship between each student's level of performance and the normal curve. Most teachers have never mastered the understanding of what these scores are trying to tell them. It is probably fair to say that some teachers do not possess the mathematical ability to analyze these results in a way that would cause them to benefit from what the scores are trying to tell them.

The process of standardized testing has no accompanying instructional management system that can direct educators to what can be done to solve the specific problems of an individual student.

Humanistic issues. Testing is said to be degrading because the constant threat of failure causes low achieving students to lose self esteem since they constantly expect to receive yet another low score each time they are tested.

Since students constantly compete to outdo each other in order to obtain a higher score than their fellow classmates, they enter into what psychologists refer to as destructive competition.

Standardized tests brand students with labels that cause their teachers to identify them as a 4.1 in reading, a 78 I.Q., learning disabled, or borderline retarded. In discussions among teachers, these labels are used frequently when referring to specific students.

Because tests provide no escape for those who have special temporary or persistent learning difficulties or have inadequate test taking aptitudes, tests are said to be impersonal.

Because of biases of tests, because of their impersonal nature, and because of the varying degree of test wiseness among students, tests are said to be unfair.

Tests cause students to be anxious and concerned about their performance in relation to other students and thus tests cause students to be pressured.

This list has provided a comprehensive although probably not a totally inclusive list of the criticisms of standardized tests. It is surprising that more has not been done to promote an alternate

model of educational evaluation. Although the criterion referenced
test model (CRT) was proposed a few years ago, it has received limited
acceptance because of the lack of understanding of the model by
teachers and by testing experts alike. This has brought about a
disagreement among measurement specialists as to the intent and purposes
of criterion referenced measurement. However, a new and different
approach to the measurement of basic skills achievement has evolved and
has been tried in six rural school districts in western Indiana and
eastern Illinois. It is not the currently accepted CRT model that has
evolved, but rather is a model that the originators of CRT envisioned.
For purposes of distinguishing the model under discussion with CRT, the
model described here will be referred to as the Competency-Rossville
Model or CRM.

The Trouble With Behavioral Objectives

Since 1960, behavioral objectives have been utilized by some
educators as a tool to specify test content, validate tests, and to
articulate curricula. Behavioral objectives require three components
to be specified in each objective. These are:

1. A behavior (something the learner must do to show that
learning has occurred).

2. Conditions (what the learner will be provided or denied
in the test situation).

3. A criterion (a minimum standard of ᵻptable
performance).

Because of the specificity required to articulate three parts
for each objective and the amount of verbal material that specification
of objectives for each grade level requires, books of objectives are
thick, cumbersome, and awkward to use. When teachers bring their
behavioral objectives to workshop sessions, it is amusing to watch them
blow the accumulated dust from the covers of their objectives books so
that they will not get themselves dusty when they use them.

An approach that has proven to be more beneficial is the
development of skills lists. These are lists of skills for each
subject tested at each grade level that teachers expect students to
have mastered. Each item from these skills lists is just the behavior
part of a behavioral objective.

Whereas a behavioral objective might be presented as:

When provided with a list of thirty long division problems
and without the use of a calculator, the student will solve at
least twenty four of them correctly.

The item from a skills list would appear as:

Solves long division problems

Although such verbs as "to know", "to understand", and "to appreciate" are usually not permitted in behavioral objectives, these words are permitted in skills lists.

Because of the economy of wording, teachers may have at their desks laminiated copies of the lists of skills that students are expected to master at their grade level. Teachers can refer to these lists as often as they need to and are constantly reminded of the basic skills curriculum.

One wrinkle with using the skills list approach is that there is no chance to specify different criteria for the various skills. Rather all objectives are specified and tests are designed so that the same agreed upon percentage of correctly answered skills test items serves as the criterion for each grade level. For purposes of these tests, the agreed upon percentage has been a criterion of 70 percent of the items answered correctly.

Copies of skills lists are provided as cover pages for the tests. Expamples of the tests and skills lists are contained in the Appendix of this report.

Norm Referenced Model (NRM) versus Competency-Rossville Model (CRM)

Any testing model has a philosophy which provides the basis for the procedures that are to be followed in the evaluative phase of instruction. Both NRM and CRM are supported by their respective philosophies. Furthermore, the philosophies of these two varieties of measurement are so fundamentally different that it seems virtually a coincidence that both are categorized as "educational measurement" and that the instruments of each are referred to by the same name "test".

In order for one to understand CRM, it is convenient to contrast it with the more common type of testing which is NRM.

A competency level or criterion in CRM is a standard of performance which serves as a minimum level to be used in a decision-making process. The competency level in CRM is the minimum score or rate that can be considered as an acceptable performance or as a minimal passing score.

In figure 1, the minimum standard of acceptable performance (the criterion) is that the student can answer 90% of the items correctly. Student P answered 95% of the items correctly. Since this score is above the criterion, Student P passed the test. Student F answered only 75% of the items correctly. His score is below criterion and thus Student F did not pass the test.

- -
Figure 1 goes about here.
- -

A norm may be thought of as an average. The mean, median, and mode are all examples of norms. Some of the types of scores derived from norm referenced information are percentiles, grade equivalent scores, age equivalent scores, I. Q scores, standard scores, and stanines. To obtain these types of scores for any student, it is

CRITERION: MINIMUM STANDARD
OF ACCEPTABLE
PERFORMANCE

100 %

STUDENT P's SCORE = 95

← (above criterion)

→ 90 % → CRITERION

80 %

STUDENT F's SCORE = 70

← (below criterion)

70 %

60 %

FIGURE 1
SCORES ON A CRITERION REFERENCED TEST

necessary to obtain the mean or some other type of norm for the group
that the student belongs to. Frequently the relative distance a
student scores from the mean is measured in units of standard
deviations. A standard score of -1.0 means the student's score is one
standard deviation below the mean while a standard score of +2.0
indicates the student's score is 2.0 standard deviations above the mean.
 (See Figure 2).

- -
Figure 2 goes about here.
- -

        Norm referenced tests are used to find out how each individual
performs in relationship to the performance of other individuals who
have taken the same test.   The meaning of a norm referenced test score
is derived from its comparison to the norm or average and consequently
with it comparison to the scores of other students.   Almost all
classroom tests and standardized intelligence tests are norm referenced
measures.   Because of the fact that they measure student's degree of
learning relative to the degree of learning of others and relative to
the normal curve, they are sometime referred to as relative tests.
        CRM is one example of what can be called an absolute form of
testing.   Absolute interpretation of test scores involves making a
judgment about the score of a student in terms of how his unique
individual performance on a test relates to a minimum standard.
However, recently a great amount of attention has been devoted to
absolute measurement by practitioners in a variety of areas.
        An absolute interpretation of test scores is advocated in such
diverse fields as individualized instruction, programmed instruction,
computer-assisted instruction, non-graded schools, governmental and
military education, performance based education, the systems approach
to education, minimum competency testing, early childhood education,
the British open school, competency based education, special education,
and physical education.
        CRM focuses attention on whether students are able to do certain
tasks acceptably.   It is because the learner is being compared to some·
established standard, rather than to other individuals, that causes
these measures to have educational value.   The meaningfulness of any
learner's score is not dependent on any comparison with scores of other
learners.

CRM, Behavioral Objectives, and Skills Lists

        The process of absolute testing has been closely tied to stating
goals of instruction in behavioral objectives.   However, as has been
stated earlier in this report, behavioral objectives are frequently
awkward to use.   The CRM model reported here utilizes only the
behavioral part of behavioral objectives.   These behaviors are listed
in what are called skills lists.
        Minimum standards vary depending on the task and its desired
degree of attainment.   Figure 3 shows some of the criterion levels that
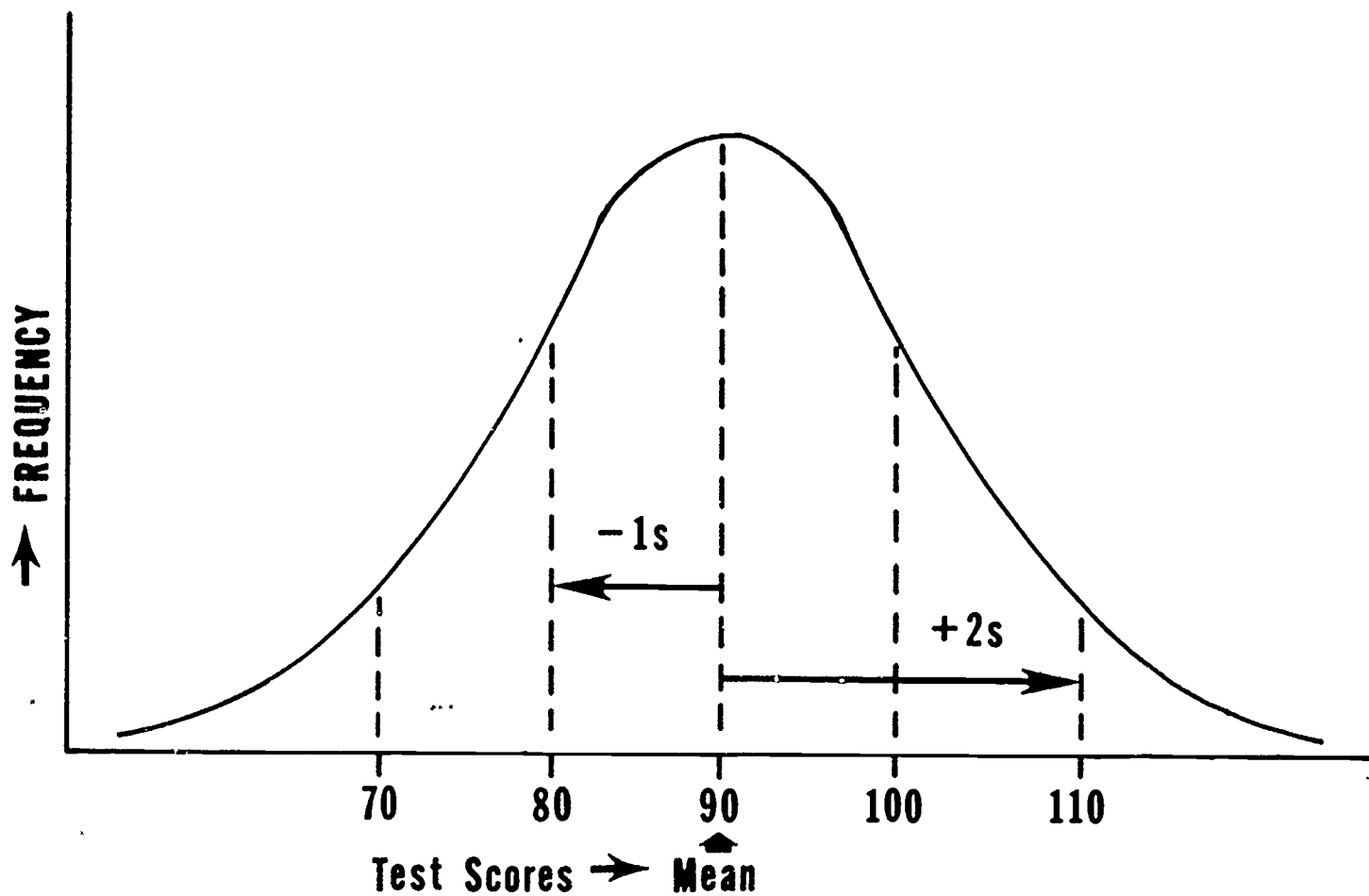may be specified for various performances.

FIGURE 2

A DISTRIBUTION OF TEST SCORES WITH A MEAN
OF 90 AND A STANDARD DEVIATION OF 10

11

- -
Figure 3 goes about here.
- -

An airline pilot will be expected to perform flawlessly on tests designed to measure piloting skills. A bright fourth grader may be expected to master all of the 100 multiplication facts. However, a social studies teacher may expect slow learning students to obtain a score of only at least 60% on a semester test. Consequently, the standard is set at 60%. A general education course taught at the college level may be taught in such a way that the instructor will consider that students have mastered the material if they score higher than the criterion of 90%.

A frequently specified minimum standard is 70%. When a teacher sets up objectives for a class, the instruction and the CRM exercises are designed and constructed in a way that explicity defines rules linking patterns of test performance to the skills lists. If 70% is the criterion score, then any student who scores above 70% will be considered by the teacher to have learned the material. Students who score lower than 70% are considered to be below the desired level of mastery.

A Double Criterion

Many instructors also use CRM to enable them to ascertain whether they are doing an effective job in teaching their classes by specifying a double criterion. The double criterion specifies the level of performance expected by each student in the class and also specifies the number of students that should meet this standard in order for the instructor to consider the instruction to be successful.

The double criterion specified in these tests is the 70-70 criterion. The 70-70 criterion means that the teacher will consider his/her work to be effective if 70% of the students are able to obtain a score of at least 70% on the test. Any student who scores above 70% will be considered as having satisfactorily mastered the material. If 70% or more of the students score above this minimum level, the instruction is considered to have been satisfactory.

The choice of the level of the criterion or the levels of the double criterion is usually determined by the instructor and is determined by the level of competency of the students, the importance of the task, and the level of the instructor's aspirations. However, in most educational circumstances, a reasonable and challenging goal for any instructional setting is the 70-70 criterion.

Steps in Constructing CRM Tests

The sequence for constructing NRM is typically to first teach, then design a test, and finally to administer it. The step-by-step procedure for utilizing CRM is a logical and rational methodology. However, some advocates of CRM feel that to follow the steps required for the construction of CRM instruments virtually ensures that the instruction will be effective.

SUITABLE CRITERIA

TEST	CRITERIA
AIRLINE PILOT	FLAWLESS
4TH GRADE MATH FACTS	100%
MILITARY TRAINING	95%
GRADUATE MEASUREMENT CLASS	90%
FREE THROW SHOOTING	70%
SLOW LEARNER, SOCIAL STUDIES	60%
BASEBALL HITTING	.250

FIGURE 3
ACCEPTABLE CRITERIA FOR VARIOUS SITUATIONS

The steps for constructing CRM are as follows. Before instruction begins and before the test is constructed, the desired - skills that are to be mastered are carefully specified in the skills lists. The situations are created in which performance of the skills is to be demonstrated by the students. These sample situations constitute the CRM instrument. Next instruction is planned so as to accomplish the mastery of the skills. After the instruction has been completed, the CRM instruments are administered to find (1) which student mastered which skills as demonstrated by their criterion scores on the various skills, and (2) whether instruction was accomplished effectively as demonstrated by the percentage of students who attain the criterion score.

Although the above sequence represents the sequential pattern that occurs in CRM, Figure 4 and Figure 5 represent a more practical representation of the sequence of CRM.

- -
Figure 4 goes about here.
- -

Figure 4 illustrates what actually occurs in CRM. First the objectives are stated in the form of a list of skills to be mastered.

Next, the test is constructed in such a manner as to determine if the student can demonstrate the accomplishment of the behaviors described in the skills list. It is interesting to note that in the sequence of CRM, test construction is the second step, while in NRM it is the next to last step.

Instruction is then performed in an attempt to master the skills on the skills list. Some critics of CRM have faulted this step of the procedure by asserting that at this point the instructor is "teaching to the test." It is a matter of individual perception as to whether that is happening or whether the skills are being taught, rather than the test. It is equally senseless to debate whether there is anything inherently wrong with teaching about the concepts that will be contained in test items.

After the instruction is completed, the CRM instrument is administered and scored. There are only two possible scores for each skill. Students who score above the criterion pass and those who score lower than the criterion do not pass.

The scores of all students are then evaluated to determine if the instruction was effective. If the desired percentage of students attain a passing score, the instructor may conclude that students are mastering the skills and that learning is being accomplished satisfactorily. If less than the desired percentage of students attain criterion, then the instructor must conclude that instruction has not been as effective as it was desired to be. The next step is to try to reason whether the objectives, the test, or the instruction should be changed on the next attempt at teaching the material.

Figure 5 demonstrates a step-by step procedure for CRM.

- -
Figure 5 goes about here.
- -

In figure 5, the criterion levels were not specified because

CRT CYCLE

① **STATE OBJECTIVES** → ② **DESIGN TEST**

③ **TEACH**

④ **ADMINISTER TEST**

⑤ **EVALUATE RESULTS**
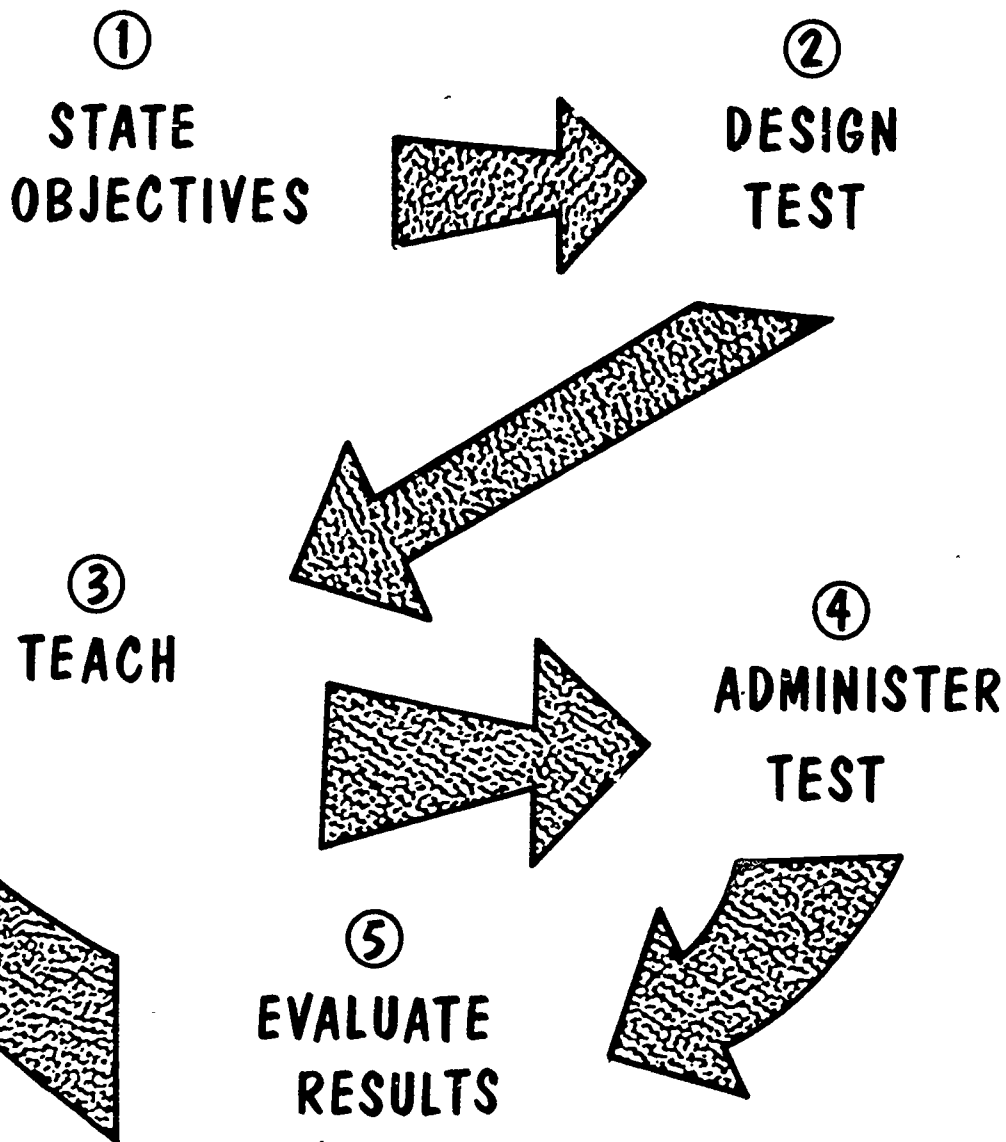
FIGURE 4

STEPS IN CRM

1. State objectives.

2. Prepare CRM instrument to measure objectives.

3. Teach to accomplish objectives.

4. Administer and score CRM instrument.

5. If any student scored above _____%, he has mastered the instruction.

6. If _____% of the students score above_____%, instruction is effective.

7. Decide if a change is needed in objectives, CRM instrument, or the instruction.

FIGURE 5

STEPS IN CRM

each instructional situation requires a decision as to the level that the learners should attain.

Figure 6 shows a model of the decision-making process associated with CRM and contrasts it with the process traditionally followed in NRM.

- -

Figure 6 goes about here.

- -

From Figure 6, it may be observed that there is no attempt made in NRM to revise instruction on the basis of the product results as measured by the NRM instrument. However, in the CRM process revisions occur in either the test, the instruction, or the objectives if results indicate that the skills are not being mastered.

Differences in NRM and CRM

It was noted earlier in this paper that both CRM and NRM are supported by their respective measurement philosophies and that the philosophies of the two varieties of measurement are strikingly dissimilar. The measurements in CRM and NRM each follow their respective measurement philosophies. Some of the differences are noted in the paragraphs below and are summarized in Figure 7.

- -

Figure 7 goes about here.

- -

Trait or ability to be measured. In NRM, the trait or ability to be measured is assumed to be present in varying degrees in different individuals. It is the purpose of NRM to order those individuals on a continuum ranging from highest to lowest in terms of the amount of that trait or ability that the learner possesses. In CRM, the trait or ability is assumed to be present in either a sufficient or an insufficient amount in different individuals. It is the purpose of CRM to separate those idividuals who have attained a prescribed level of mastery of the trait or ability from those who have not.

Previously acquired skills. Furthermore, CRM items are likely to be fashioned so that they focus on the measurement of the actual instruction, while controlling for or eliminating the measurement of previously learned traits, abilities, and prior achievements of the examinee.

Range of scores. In NRM, the test is designed so that students' test scores range from a low which is approximately equal to the chance level of the test to a high which may be equal to 100%.

CRM scores are considered to be passing if the student attains the criterion or above and are considered to not be passing if the student does not attain the criterion score. CRM scores can only take one of two possible values. The two values are variously specified as pass-not pass, pass-fail, go-no go, adequate-inadequate or yes-do over. The two value scoring of CRM is frequently referred to as producing dichotomous data. However, it could be logically argued that instruction is most effective when everyone receives the same score of
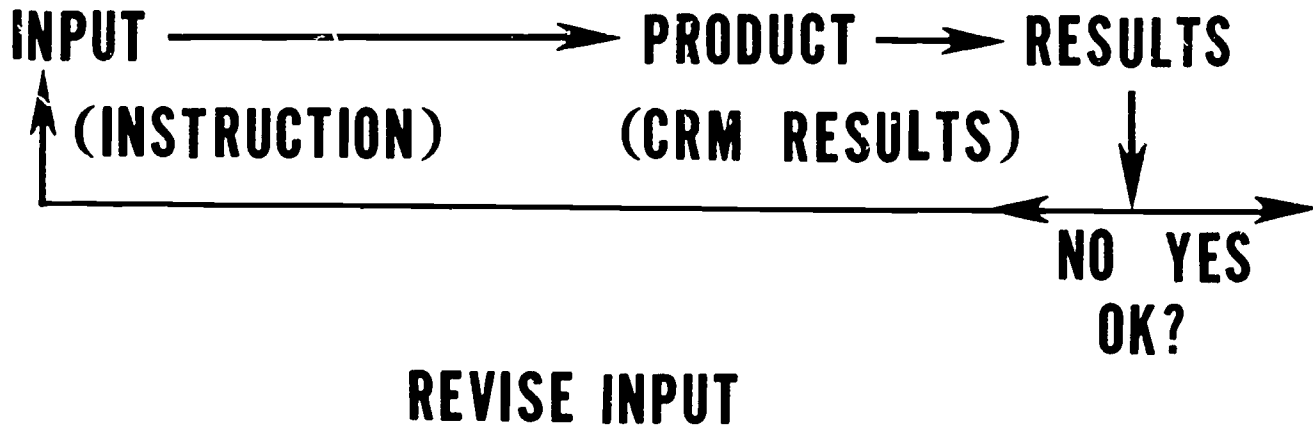
NRM MODEL

INPUT   ⟶   PRODUCT

(INSTRUCTION)   (NRM RESULTS)

CRM MODEL

INPUT   ⟶   PRODUCT ⟶ RESULTS

(INSTRUCTION)   (CRM RESULTS)

NO   YES

OK?

REVISE INPUT
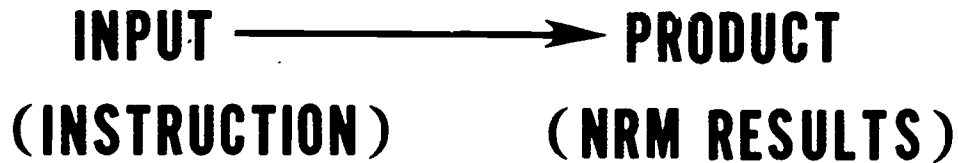
FIGURE 6

DECISION MAKING PROCESS IN NRM AND CRM

18

PHILOSOPHY OF TESTING

NRM		CRM
CONTENT	VALIDITY	CURRICULAR
TESTED	PREVIOUSLY ACQUIRED SKILLS	NOT CONSIDERED
STUDENT—STUDENT	COMPARISON	STUDENT—CRITERION
ORDERING STUDENTS	FUNCTION	EVALUATING INSTRUCTION
MAXIMIZE MATERIAL COVERED	INSTRUCTION FOR TEST	MAXIMIZE OBJECTIVES
COGNITIVE	DOMAIN OF INSTRUCTION	PSYCHOMOTOR OR COGNITIVE
MEDIUM	DIFFICULTY	EASY
HIGH	DISCRIMINATION	ZERO
IMPORTANT	RELIABILITY	UNIMPORTANT
HIGH	RANGE OF SCORES	LOW
IN VARYING DEGREES	TRAIT MEASURED	GO OR NO GO
NUMBER CORRECT	TYPE OF SCORES	DICOTOMOUS
NORMAL	DISTRIBUTION	RECTANGULAR
PARAMETRIC	STATISTICAL ANALYSIS	NONPARAMETRIC

FIGURE 7

SUMMARY OF DIFFERENCES BETWEEN NRM AND CRM

15

19

20

pass and therefore, there would only be one rectangular distribution.

Difficulty of items. Most test theorists believe that norm referenced test items of medium difficulty will produce the greatest discrimination, provide the most information, and will contribute most to the reliability of the test. Test experts specify that the best items on a norm referenced test are those for which the number of correct responses is approximately half way between chance and 100%. This means that for an essay or short answer completion test item, the ideal difficulty level would be for only half of the students to respond correctly.

Neither psychology nor common sense support motivating students by asking them questions that only half of them can answer correctly.

Although the actual difficulty level of CRM instruments depends on the ability of the group of students involved, the level of mastery required, and the objectives of the instructor, traditionally CRM items are relatively easy test items. Sometimes a criterion of 90% is specified. In this case, 90% of the students can be expected to anwer most items correctly.

Domain of instruction. It is difficult to make generalizations about the domain of instruction that is measured by the two types of tests, but it is fairly safe to say that NRM has most often been used for measuring learning of the factual information and concepts that is usually referred to as the cognitive domain. While CRM can be readily used for measurement in the cognitive domain, the nature of CRM also makes it especially useful for measuring the learning of physical skills that are included in the psychomotor domain.

Discrimination. NRM tests attempt to rank order groups of student from high to low. An NRM test is considered to be a good item if those who do well on the total test also do well on that item. Item analysis is a procedure through which a test constructor carefully evaluates each item to determine if the item discriminates between good and poor students. Items that do not have this quality are discarded and do not remain in the test.

In CRM, the best items are those that indicate that a large percentage of examinees have mastered the instruction. Therefore, good test items are found among those items that either show low or zero discrimination. It could be argued that the best educational situation occurs when everyone gets all of the items on a CRM test right. Thus, it might be an acceptable point of view to consider the best test items to be ones of zero discrimination.

Reliability. The reliability (the precision or accuracy of measurement) is a prime consideration for NRM. Mosts often, reliability estimates for NRM are obtained indirectly by correlational coefficients since reliability cannot be measured directly. Reliability is not considered to be such an overriding concern in CRM and most CRM instruments are constructed without much attention to reliability. NRM instruments are usually relatively long tests, since the degree of reliability is directly related to test length. Since reliability is not as important to most CRM constructors, CRM instruments are often shorter tests. It should be pointed out that the reliability of the skills tests that were developed in this project were all very high. Almost all of the tests developed have had reliability coefficients of above .90.

Validity. There are many methods for determining the validity

of an NRM instrument. Content validity is the most frequent means of
validity determination for NRM achievement tests. Content validity
attempts to demonstrate that the items covered on the test constitute a
representative sample of the material covered during instruction.
Since CRM items are based on the skills specified in skills lists,
curricular validity is used to determine the test content. Curricular
validity is established by keying a series of test items to each of the
skills in the skills list.

Previously acquired skills. In NRM, students must often use
previously acquired skills to respond to items so that they may
demonstrate the broad global understandings that are typically measured
by NRM instruments. CRM usually focuses on the learning specified in
the skills lists and consequently does not typically require the
learners to integrate as much of their previously acquired skills into
their test performance.

Comparisons. NRM measures a student's performance in relation
to that of the group norm and also to that of each of the other
students. CRM encourages competition with one's self to acquire
proficiencies. CRM merely attempts to find what each student can and
cannot do rather than attempting to find out who can do more of it than
other students can do. The student's score is compared to the criterion
rather than to the scores of other students.

Distribution of test scores. The distribution of test scores in
NRM is, ideally, a normal distribution. The CRM distribution of scores
consists of two rectangular distributions, passing and not passing.
These distributions are illustrated in Figure 8.

- -
                Figure 8 goes about here.
- -

        The NRM distribution is appropriate for the purpose of NRM which
is to order the group measured from high to low. The two rectangular
distributions illustrate the function of CRM which is to separate
students who have mastered skills from those who have not.

Instruction related to the test. Instructors who teach to an
NRM test try to maximize the amount of material covered. The goal of a
teacher teaching to an NRM test is to provide through a complete survey
of the field a thorough overview of the subject matter. Instructors
teaching with the anticipation of a CRM test try to maximize the
percentage of students who will master the skills.

Score. The score received in an NRM test is usually the number
of items answered correctly or the percentage of correct responses. As
previously indicated, the only score a student receives on a CRM test
is either of two scores, pass or non-pass.

Function. NRM measures the amount of knowledge individual
students have learned by ranking them from high to low. CRM evaluates
the effectiveness of instruction by determining how many of each class
have mastered the skills.

                Advantages of NRM and CRM

        Articles have be   written which have declared NRM to be immoral
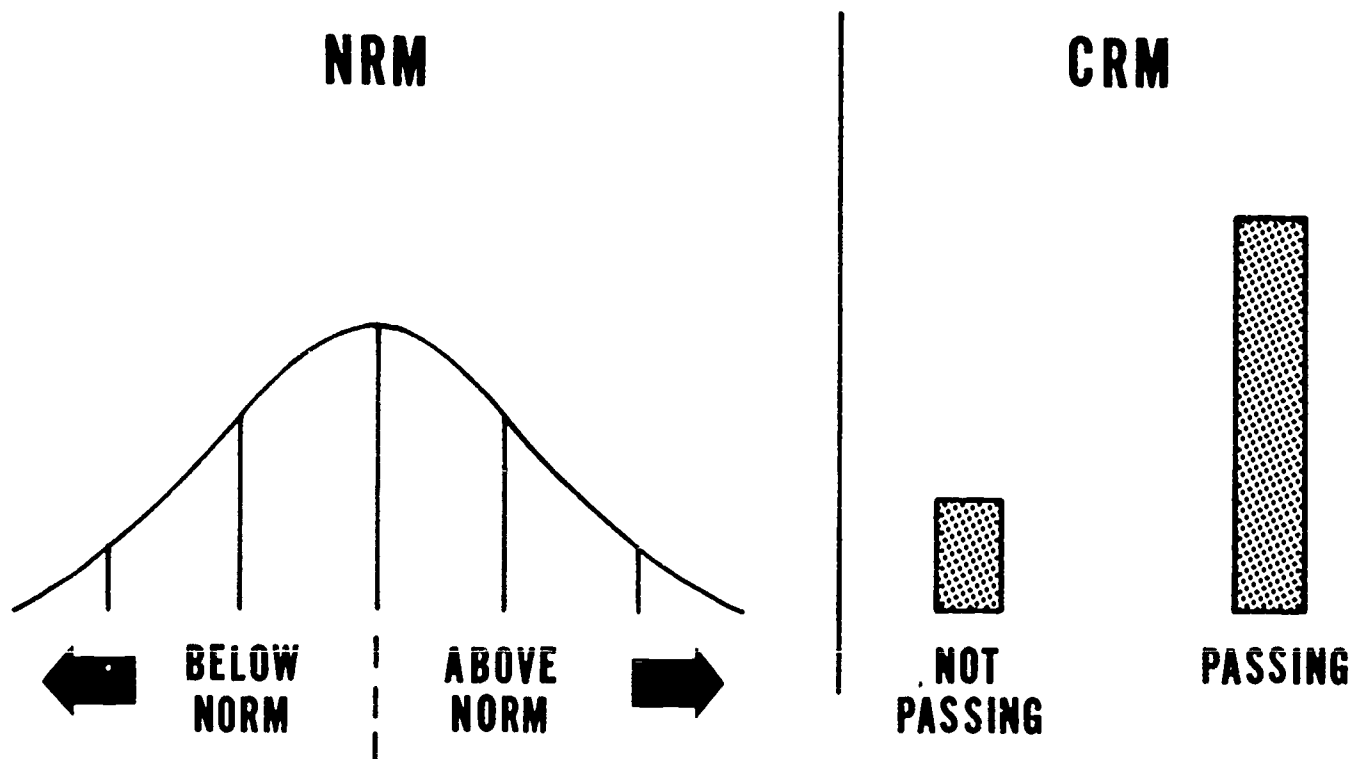
DISTRIBUTIONS

NRM   **CRM**

BELOW   ABOVE
NORM   NORM   NOT   PASSING
PASSING

FIGURE 8
DISTRIBUTIONS OF SCORES IN
NRM AND CRM

and have proclaimed CRM to be the only humane way to evaluaate
students. The rationale for these articles appears to be that it is
an inherent characteristic of NRM for half of the students to miss
each item and for half of the students to fall below the norm. This
approach does not serve to motivate students and consequently NRM
fails to encourage the type of success that enhances motivation and
learning. The critics of NRM also fault student vs. student
competition and consider the competition of students with themselves
or with a criterion to be healthier and non-destructive.

Certainly the potential for evaluating instruction is greater
in CRM than in NRM, because traditional NRM has never been concerned
with the evaluation of instructional effectiveness or with the
improvement of subsequent instruction. The NRM model has been one
that has been preoccupied with aptitude, selection, prediction, and
inference. The CRM model is concerned with evaluating and revising
instruction. CRM can lead to more meaningful information than that
provided with the NRM model when criteria are obvious and simple ones.
The information provided concerning the mastery or non-mastery of
skills is more useful for helping students who have specific learning
difficulties.

Method for School Improvement through Inservice Education

Although this method of improving schools through inservice
test construction has been tried in six school systems in rural
Indiana and Illinois, it was developed most extensively in Rossville,
Illinois and therefore is called the Rossville model.

Inservice training is involved in the training of faculty in
the testing model. After faculty are educated concerning the
inadequacies of standardized testing, they are very sensitive and
accepting of a new model. Inservice workshops are provided for
faculty to discuss and form drafts of the skills lists that will serve
as the basis for determining the content of the CRM skills lists.

A committee is then formed consisting of one teacher for each
subject to be tested at each grade level. The grade level
chairpersons utilize the Delphi method to confer with their fellow
teachers. Each grade level representative is in charge of an informal
committee for their assigned subject area and grade level. Teachers
at each grade level confer informally and complete the list of skills
that they agree students should have mastered at the end of each
semester of instruction.

The committee works informally with their grade level
chairpersons to develop test items to test the skills. The committee
of grade level chairpersons meets periodically with a testing
consultant or a curriculum specialist to develop the skills tests.
Figure 11 shows an organizational chart of the project.

- -
Figure 9 goes about here.
- -

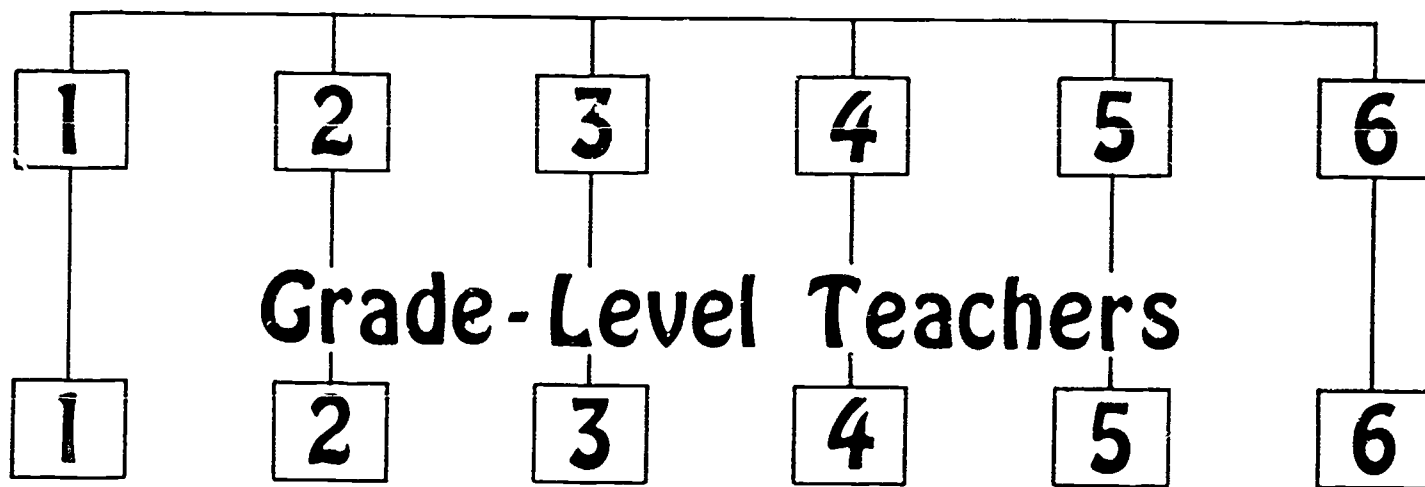It has been observed that many otherwise competent teachers do
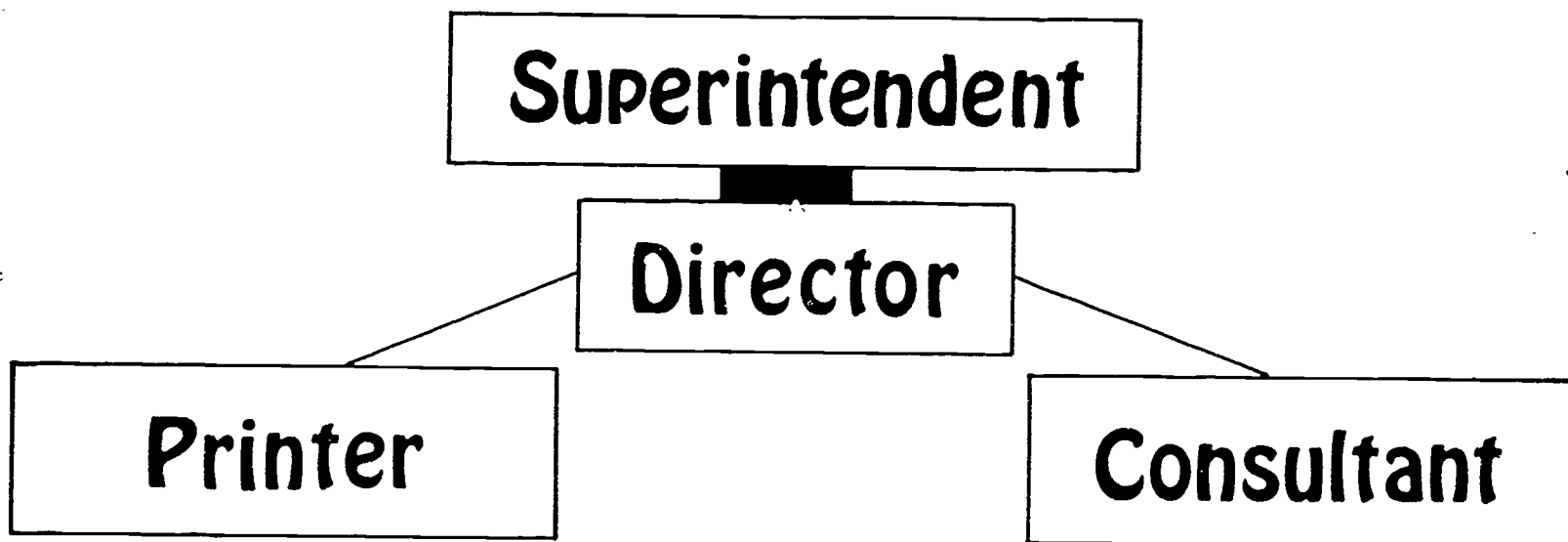
Figure 2

not write good test items. Therefore, the test development phase of
the project utilizes large item pools. Rather than the items being
randomly generated by a computer, they are carefully selected and
screened by the committee. Items are selected and screened to form
semester competency tests in the basic skills for each grade level.

The committee members are instructed to either select or write
items in such a way that they would estimate that at least 70% of the
students who have been instructed in the skill would respond to at least
70% of the items correctly.

After the tests have been administered and scored, each
unmastered skill will be analyzed to determine whether the test items,
the skills lists, or the instruction needs to be improved for groups of
students who will be learning at each grade level.

Tests are scored by a mark sense reader and results are
interpreted by a computer. After the results have been analyzed by the
testing consultant, a summary of the findings is presented to the
faculty and the school board.

Characteristics of the CRM

The Rossville model measures skills that teachers expect
students to have mastered. At least seventy percent of the students
answer at least 70 percent of the items correctly.

The Rossville Model possesses all of the advantages that were
described in an earlier part of this report and also goes a long way to
overcoming many of the disadvantages cited for NRM.

Among specific characteristics this method possesses are:

1. The tests are inexpensive and relatively easy to
    construct.

2. The tests monitor student progress and provide a
    diagnosis-remediation approach to learning.

3. The tests measure important basic skills.

4. The testing design procedure involves the total
    instructional staff.

5. The same testing pattern is integrated into all grades.

6. Tests are free from errors and contain clear and
    unambiguous items.

7. There are scores for students on each skill and for the
    total test.

8. There are grade summaries for each skill at each grade
    level.

Advantages of the CRM Model

Advantages for the CRM model over traditional standardized tests

are:

Time factor. High quality tests can be developed from start in a few months.

Skills lists. The skills lists serve to articulate the curriculum in the basic skills and to provide a basis for test development.

Management system. There is a diagnostic-remediation feature - clinical approach to learning- that is not available in most other testing methods.

Curriculum. The skills lists cause faculty to carefully examinea what they are trying to accomplish. In some school systems this has not been done in the past forty years.

Professional appearance. When tests are printed by a commercial printer, they have a professional appearance.

Item pool. The teachers may either write items or work with previously written items from a large item pool.

Scoring. Results are obtained by scoring with mark sense equipment and analysis by computer.

Test analysis. The tests and individual items on each test are constantly monitored by a computerized item analysis.

Revisions. Where test results indicate, items are revised to correct any editorial or statistical deficiencies.

Summary

f

This paper has described the logic and the procedures utilized in an inservice approach to curriculum and school improvement through the development of local tests in the basic skills. The approach has been attempted at six schools. Teachers at all of the schools are enthusiastic about the instructional advantages of this testing model over traditional testing. None of the schools has discontinued using i to this date. Teachers prefer the tests to traditional norm referenced tests and find that the information the tests provides assists them to work with individual students.

Future developments that are planned are to expand the testing method to other subjects than the basic skills and to computerize the entire testing process so that students can have tests scored and interpreted while working at a microcomputer.

Although this is a method which requires much work on the part of the teachers and consultants and much cooperation from the school administration, it is a method which has proven to eliminate many of the inadequacies and much of the unfairness that have been associated with standardized norm referenced tests.

APPENDIX

Name_____

Date_____

Score_____

Skill Referenced
MATH
3-2
Evaluation System

ITEMS	SKILLS	SCORE	MASTERY
1-5	Identifies value of all U.S. coins.	_____	_____
6-10	Interprets graph data.	_____	_____
11-15	Knows days of week and months of year in sequence.	_____	_____
16-20	Identifies operation needed (addition or subtraction) in story problems.	_____	_____
21-25	Writes time correctly to half-hour.	_____	_____
26-30	Knows meaning of halves, thirds, and fourths.	_____	_____
31-35	Measures to ½ inch.	_____	_____
36-40	Measures to nearest centimeter.	_____	_____
41-45	Understands number families.	_____	_____
46-50	Knows multiplication facts 1's, 2's, 3's, 5's, and 10's.	_____	_____

Revised 1983-84

Choose the best answer:

1.

4 quarters =

a) 90¢   c) 60¢

b) 100¢   d) 75¢

2. A penny is

a) $0.01   c) $0.10

b) $0.05   d) $0.25

3. A nickel is

a) $0.01   c) $0.10

b) $0.05   d) $0.25

4. A quarter is

a) 1¢   c) 10¢

b) 5¢   d) 25¢

5. A dime is

a) 1¢   c) 10¢

b) 5¢   d) 25¢

Answer items 6 - 8 using this graph.

NUMBER OF HOURS DAVID FISHED
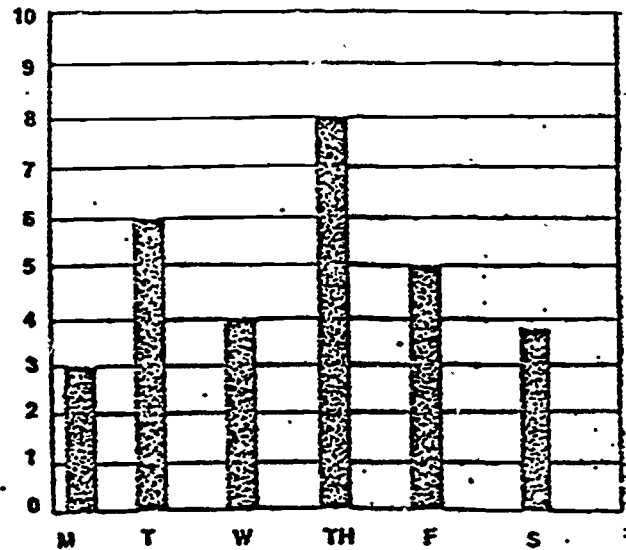
6. David fished _____ hours on Thursday.

a) 4   c) 6

b) 5   d) 8

7. David fished the shortest time on _____

a) Monday   c) Wednesday

b) Tuesday   d) Thursday

8. David fished the longest on _____

a) Monday   c) Thursday

b) Tuesday   d) Friday

TURN THE PAGE

page 1
Revised

M 3-2
1983-84

30

Answer items 9 and 10 using this graph.

BOXES OF COOKIES SOLD

SALLY	
RUTH	
TERRY	
CONNIE	

0 10 20 30 40 50 60
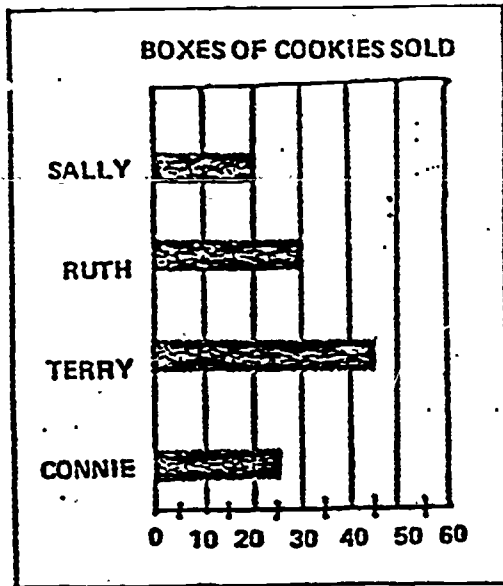
9. Who sold about 45 cookies?

a) Connie   b) Ruth   c) Terry

10. About how many cookies did Connie sell?

a) 15   b) 25   c) 35

MAY

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

Look at the calendar to answer these questions.

11. What day of the week is the fifth of May?

a) Monday   c) Wednesday
b) Tuesday   d) Thursday

12. What day does May begin on?

a) Saturday   c) Monday
b) Sunday   d) Friday

Which is missing?

13. June _____ August

a) May   c) September
b) April   d) July

Which is missing?
14. Tuesday _____ Thursday

a) Wednesday   c) Monday
b) Friday   d) Saturday

Which is missing?
15. February _____ April

a) January   c) December
b) March   d) July

For questions 16 - 20, read the story problem, and choose the correct answer.

16. Mary had 10 balloons.
    3 broke.
    How many balloons were left?

a) 7+3=10   c) 10-7=3

b) 10-3=7   d) 3+7=10

17. There were 14 cats and 22 dogs in the show. How many cats and dogs in all were in the pet show?

a) 36-22=14   c) 14+22=36

b) 36+22=14   d) 36-14=22

18. Bill had 20 cars.
    He lost 8 cars.
    How many did he have left?

a) 20-8=12   c) 8+12=20

b) 12+8=20   d) 20-12=8

19. Steve had 14 apples.
    Dave had 5 apples.
    How many more apples did Steve have than David?

a) 5+9=14   c) 14-9=5

b) 14-5=9   d) 9+5=14

TURN THE PAGE

page 2
Revised

M 3-2
1983-84

31

20. Polly found 13 butterflies.
John found 10 butterflies.
How many more did Polly find than John?

a) 13-3=10 c) 3+10=13

b) 10+3=13 d) 13-10=3

25.

a) 3:00 c) 9:00

b) 4:00 d) 10:00

Choose the correct time.

21.

a) 3:15 c) 10:15

b) 10:00 d) 10:30

Choose the correct answer.

26. Which circle is divided into fourths?

a) b) c) d)

22.

a) 6:15 c) 6:30

b) 2:30 d) 2:60
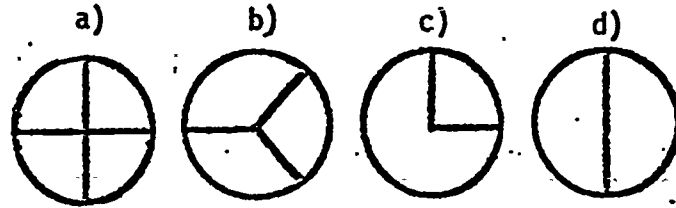
27. Which square is divided into thirds.

a) b) c) d)

23.

a) 12:00 c) 5:45

b) 6:00 d) 12:50

28. How much is shaded?

a) 1/8 c) 1/3

b) 1/4 d) 1/2

24.

a) 6:00 c) 9:00

b) 4:30 d) 12:00

TURN THE PAGE

page 3 M 3-2
Revised 1983-84

32
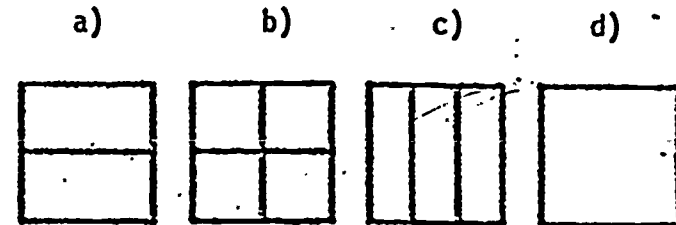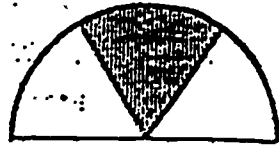
29. How much is shaded?

a) 1/8   c) 1/3
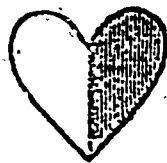
b) 1/4   d) 1/2

30. Which part of the object is shaded?

a) 1/2   c) 1/4
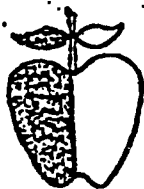
b) 1/3   d) 1/5

Choose the correct measurement.

31.
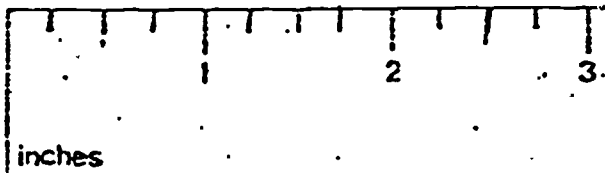
a) 2 inches   b) 2½ inches   c) 3 inches

32.
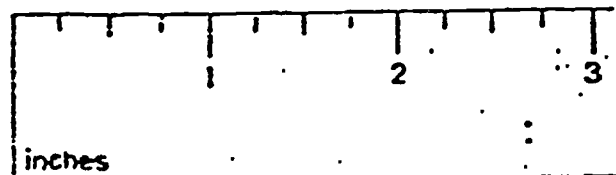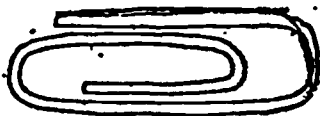
a) 1 inch   b) 2 inches   c) 1½ inches
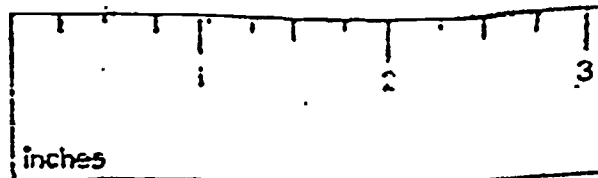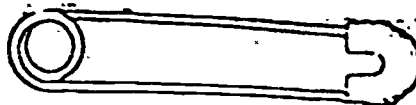
33.

a) 1½ inches   b) 2 inches   c) 2½ in

34.

a) 1½ inches   b) 2½ inches   c) 2 in

35.
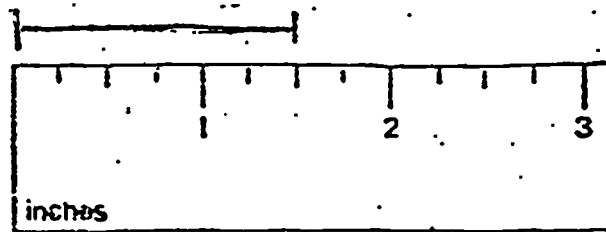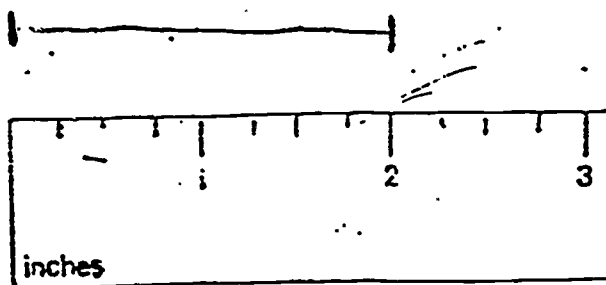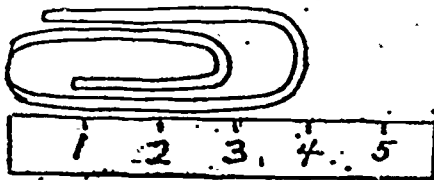
a) 1½ inches   b) 2 inches   c) 2½ ir

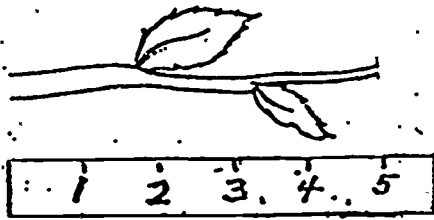Circle the correct measurements in centimete

36.
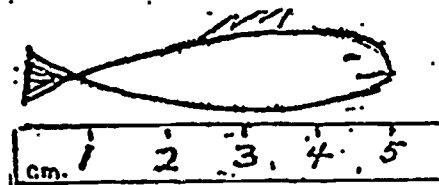
a) 3 centimeters   b) 2 centimeters

c) 5 centimeters
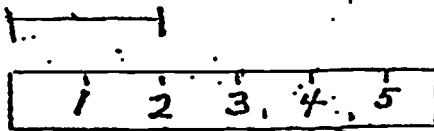
37.

a) 2 centimeters   b) 3 centimeters

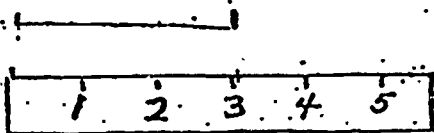c) 4 centimeters

38.

a) 3 centimeters   b) 5 centimeters

c) 4 centimeters

39.

a) 3 centimeters   b) 2 centimeters

c) 4 centimeters

40.

a) 3 centimeters   b) 2 centimeters

c) 4 centimeters

---

41. Choose the number family for the numbers
     (8, 2, 10)

a) 8+8=16   b) 10-5=5
    8+2=10   10-2=8
    2+8=10   8+2=10
    10-8= 2   2+8=10

        c) 8+2=10
            2+8=10
            10-8=2
            10-2=8

42. Choose the number family for the numbers
     (8, 9, 17)

a) 8+9=17   b) 17-6=11
    9+8=17   8+9=17
    17-9=8   9+8=17
    17-8=9   17-9=8

        c) 8+7=15
            8+9=17
            17-8=9
            17-9=8

43. Choose the number family for the numbers
     (3, 9, 12)

a) 3+3=6   b) 12-9=3
    9+3=12   12-3=9
    3+9=12   9+3=12
    12-9=3   3+9=12

        c) 12-6=6
            3+9=12
            12-3=9
            9+3=12

44. Choose the number family for the numbers
     (15, 4, 11)

a) 15-4=11   b) 15-5=10
    11+4=15   15-11=4
    15-11= 4   15-4=11
    4+11=15   11+4=15

        c) 4+11=15
            4+7=11
            15-4=11
            15-11=4

45. Choose the number family for the numbers
    (9, 5, 14)

a) 14-9=5   b) 9+5=14
    14-5=9   14-6=8
    5+9=14   14-9=5
    9+5=14   5+9=14

            c) 9+5=14
                4+9=13
                14-9=5
                14-5=9

CHOOSE THE CORRECT ANSWER

46. 3 X 2 = _____

a) 5   b) 6   c) 10

               d) 1

47. 6 X 1 = _____

a) 24   b) 22   c) 7

               d) 6

48. 8 X 5 = _____

a) 32   b) 13   c) 40

               d) 18

49. 7 X 3 = _____

a) 13   b) 42   c) 35

               d) 21

50. 9 X 8 = _____

a) 72   b) 17   c) 42

               d) 64

                    END