DOCUMENT RESUME

ED 124 587 TM 005 344

AUTHOR Estes, Gary D.; And Others
TITLE A Criterion-Referenced Basic Skills Assessment Program in a Large City School System.
PUB DATE [Apr 76]
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Standards; *Basic Skills; Behavioral Objectives; *City Wide Programs; *Criterion Referenced Tests; Educational Accountability; *Educational Assessment; Elementary Secondary Education; Graduation Requirements; Item Analysis; Mathematics; Public Schools; Reading; Student Needs; Student Testing; Test Construction; Testing Programs; Test Reliability; Test Validity
IDENTIFIERS Arizona (Phoenix); Minimal Reading Proficiency Assessment; Minimum Mathematics Proficiency Test; *Phoenix Union High School System

ABSTRACT

The Phoenix Union High School System has developed a basic skills assessment program in the areas of Reading and Mathematics. Procedures and considerations used in developing the system are discussed. Preliminary results on validity and reliability of the assessment instruments are presented. An example of criterion-referenced test development using traditional item analysis, reliability, and validity procedures is provided. (Author)

A Criterion-Referenced Basic Skills Assessment Program in a Large City School System

Gary D. Estes, Lloyd W. Colvin, and Coleen Goodwin
Phoenix Union High School System

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976

Session Number 13.04

ABSTRACT

A Criterion-Referenced Basic Skills Assessment Program in a Large City School System

GARY D. ESTES LLOYD W. COLVIN COLEEN GOODWIN
Phoenix Union High School System

The Phoenix Union High School System has developed a basic skills assessment program in the areas of Reading and Mathematics. Procedures and considerations used in developing the system are discussed. Preliminary results on validity and reliability of the assessment instruments are presented. An example of criterion-referenced test development using traditional item analysis, reliability, and validity procedures is provided.

A Criterion-Referenced Basic Skills Assessment Program in a Large City School System

An increased clamor for the accountability of public school systems is probably in part a result of reports which seem to indicate that the academic performance of students is decreasing. Reports put out by the American College Testing Program and College Entrance Examination Boards describe a steady decline in the overall performance level of high school students taking those exams. Such reports along with statements from the business community that high school graduates cannot do simple math calculations, fill out employment forms, or write have led to the increased emphasis given to the basic skills of reading, writing, and computing by boards of education and educators. An example of the emphasis given to the issue of accountability in the basic skills is found in the Arizona Revised Statute, 15-102 section 25, which states:

> The State Board of Education shall:
>
> By June 30, 1975, in cooperation with all local school districts, develop, establish, and direct the implementation of a continuous uniform evaluation system of pupil achievement in relation to measureable performance objectives in basic subjects. The Board shall assist in the development of alternative learning procedures to help pupils attain their individual learning expectancy levels based on intelligence factors, achievement factors and teacher evaluation. Basic subjects shall be defined for these purposes as reading, writing and computation skills. As amended, Laws 1972, Ch. 111 and Ch. 168.

The State Board of Education adopted the following policies in response to the legislative mandate,

> ...each student shall demonstrate ability to read at a ninth

grade level of proficiency as shall be established by the local district prior to graduation from high school. This policy becomes effective after January 1, 1976.

Mandates and policies such as the above can be criticized on the basis of poorly defined concepts or principles, but they are indications of the concerns of many groups about the effectiveness of current education.

In order to meet the concerns of the Board of Education, educators, and patrons, the Phoenix Union High School System has developed programs addressed to student needs, in the basic skill areas of reading, writing, and mathematics. An integral part of these basic skills programs has been the development of criterion-referenced assessment programs in the areas of reading and mathematics and the establishment of minimal proficiencies as requirements for graduation. The Phoenix Union High School System Board of Education approved the following policy dealing with basic skills on May 3, 1973.

> Twenty units of credit are required for graduation...
> In addition, students must demonstrate the ability to read at least a ninth-grade level of proficiency as established by the District beginning with the graduating class of 1975 and demonstrate proficiency in basic computational skills in mathematics beginning with the graduating class of 1977.

A basic skills program in writing will be initiated in the fall of 1976.

Establishing Goals and Objectives

Reading Goals and Objectives

The definition of ninth grade reading proficiency was to be established by the Phoenix Union High School System. The District Supervisor for Reading and a committee of reading specialists identified the goals and objectives for Minimal High School Reading Proficiency. An overall referant used in formulating the

goals and objectives was the question, "What basic reading skills does an adult member in our society need in order to function without handicaps?". Support for the goals and objectives was established from a review of the literature and opinions from leading reading educators. Twelve performance objectives were originally identified,

Comprehension Skills

1. A functional reader will correctly identify stated main ideas.
2. A functional reader will correctly identify inferred main ideas.
3. A functional reader will correctly identify conclusions drawn from a stated and/or inferred main idea.
4. A functional reader will correctly identify stated supportive details.
5. A functional reader will correctly identify conclusions drawn from stated and/or inferred supportive details.
6. A functional reader will correctly identify stated and/or inferred sequences within selected contents.
7. A functional reader will correctly answer relationship questions, such as cause and effect, fact and opinion, and/or time and space.

Study Skills

8. Given written directions, a functional reader will demonstrate the ability to follow those directions.
9. Given tasks using index and/or table of contents, a functional reader will demonstrate the ability to complete the given tasks.
10. Given tasks using a dictionary, a functional reader will demonstrate the ability to complete the given tasks.
11. Given tasks for extracting information from graphs, tables, maps, charts, diagrams, pictures, and/or cartoons, a functional reader will demonstrate the ability to complete the given tasks.

12. Given tasks for using context clues, a functional reader will select an appropriate meaning for a given unfamiliar word.

These twelve goals were expanded to fourteen in 1975. Modifications in the Comprehension Skills included combining objectives 3 and 5; separating the items in objective 7 into three objectives: stated cause/effect relationships, inferred cause/effect relationships, fact/opinion statements; and adding the identification of relevant/irrelevant data. In the Study Skills area, objectives 9 and 10 were combined and an objective added - identification of prefixes, suffixes, and roots.

Mathematics Goals and Objectives

A committee consisting of the District Supervisor of Mathematics and mathematics teachers identified 16 objectives as the minimal skills which would be required for graduation. These objectives are as follows:

1. add whole numbers.
2. subtract whole numbers.
3. multiply whole numbers.
4. divide whole numbers.
5. add fractions.
6. subtract fractions.
7. multiply fractions.
8. divide fractions.
9. add decimal numbers.
10. subtract decimal numbers.
11. multiply decimal numbers.
12. divide decimal numbers.
13. find an integral percent of a number.
14. read information from scales, graphs or charts.

15. select equivalent expressions for measurements.

16. give the numeral for the word representation of a number.

The objectives and goals listed have been approved by the Phoenix Union High School System as the minimal requirements necessary for graduation. The position taken is that these reading and mathematics skills are a minimal set of skills that students need to function in society without handicaps, keeping in mind that an important goal of the Phoenix Union High School System is to help students to achieve well beyond these minimal skills.

Development of Instruments

An important step after the goals and objectives were established was the selection or construction of instruments to measure student mastery of the objectives. A review of available instruments revealed that no tests currently published would measure precisely the goals and objectives established by the Phoenix Union High School System. The decision was made to develop criterion-referenced measures of the skills defined within the goals and objectives of the Phoenix Union High School System.

Several authors have proposed issues to be considered in the development of criterion-referenced tests (CRT) or domain-referenced tests (DRT), e.g., Cox and Vargas (1966), Hambleton and Novick (1973), Millman (1974), Popham and Husek (1969), and Woodson (1974). A prevalent theory is that the use of traditional item analysis and test development procedures will compromise the quality and usefulness of CRT or DRT. Haladyna (1974) obtained evidence that in at least some instances the potential limitations of using traditional procedures in developing CRT were not a problem.

The procedures used by the Phoenix Union High School System in developing the CRT in reading and mathematics were very similar to those normally used to develop

norm-referenced tests.

1. Committees of teachers and content area supervisors met in workshops and developed many test items for each objective.
2. The test items were divided into a number of field tests which were given to students in randomly selected freshman classes in every high school in the District.
3. Item difficulty and discrimination values were computed and used, along with teacher judgment, in identifying the test items to retain, revise, or delete from the item pool.
4. Three alternate forms of each test were constructed using the field testing results, in which the test items pertaining to each objective were mixed up.
5. Since then more tests have been devised, making available test forms in which the test items pertaining to each objective are grouped together.

Item Writing

The content area committees consisting of the supervisors and teachers used sources such as daily newspapers, application blanks, contracts, a driver's manual, and tax forms as a basis for the development of multiple choice questions. An attempt was made to develop test items which would be as closely related as possible to the actual performance students would be required to make outside of academic exercises. Attention was given to the construction of test items that were related to students interests as well, in order to facilitate the probability of obtaining an accurate assessment of each student's mastery or non-mastery of the skills. An attempt also was made to minimize ethnic biases. Each question had five possible responses.

Sufficient test items were written in order to select after the field testing

at least four questions per objective for each test form.

Field Testing of Test Items

Over 150 reading items placed in six pilot tests were field tested on random samples of ninth grade students. The 128 math items were subdivided into three subtests and were also administered to random samples of other ninth grade students.

Item difficulties and discrimination values were computed from the field test data. Item discrimination values were computed using the formula:

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

where,

H = number of students in the top 27% who answered the question correctly.

L = number of students in the bottom 27% who answered the question correctly.

N = number of students 27% represents.

Item difficulty values were the percent of all students who answered the items correctly.

Selection of Items

Based upon the results of the field testing, the committees met to select the items which would be used for the various forms of the Minimal Reading Proficiency Assessment and the Minimum Mathematics Proficiency Test.

The committee selected items based upon the discrimination values, item difficulties, and teacher judgments about the items. A general goal was to select items with discrimination values of at least .20 and difficulty values

from .40 to .80. If an item's discrimination value was less than .20, the committee examined the item to determine if the low value might be due to a very high or very low difficulty. If the discrimination value was low and the difficulty was very low or very high, the item was selected and/or revised if the committee determined that the item was needed to measure an important skill. Responses to item alternatives were also examined for clues to poorly or ambiguously worded items. In summary, traditional measures of item difficulty and discrimination were used with teacher judgments to select and revise the items to be included in the final test forms.

Operating Procedures

At the present time six forms of the Minimal Reading Proficiency Assessment and six forms of the Minimum Mathematics Proficiency Test are used in the basic skills assessment. The forms have the items grouped according to skill areas or randomly sequenced without regard to skill area. Items were grouped by skill area so that a student might take only those items within the skill areas not mastered. The tests contain four items for each skill area. Mastery is defined as at least three correct of four items for each skill area. As stated previously, a student must pass all the skill areas on the Minimal Reading Proficiency Assessment and Minimum Mathematics Proficiency Test as a requirement for graduation. Some "good" students have not passed the mimimal proficiency tests on the first try, thus revealing to them and their teachers a few weaknesses in their basic skills competencies.

Procedural manuals which describe procedures for exempting students from the minimal reading and mathematics requirements, e.g., students in special education and/or with learning disabilities may be exempted, have been developed for each assessment program. The procedural manuals state that each student

must be given a program of instruction designed to meet his/her specific needs. Provision for a continuous evaluation system also is included. Student records are maintained in the Phoenix Union High School System data processing Student Test History File and are updated periodically to reflect changes in student achievement of objectives.

A report to parents is provided twice a year to freshmen students and once a year to those sophomores, juniors, and seniors not passing the tests. The report includes a brief description of each skill area, a statement of the requirements for proficiency, and an indication of completion or noncompletion for each skill area for the Minimal Reading Proficiency Assessment and the Minimum Mathematics Proficiency Test.

Item Characteristics

The inter-quartile range for the Minimal Reading Proficiency Assessment item discrimination values is from .42 to .71 with a median of .56. The average Minimal Reading Proficiency Assessment item difficulties for entering ninth graders is approximately .70 for the four different forms of the Minimal Reading Proficiency Assessment analyzed so far. The inter-quartile range for the Minimum Mathematics Proficiency Test item discriminations is from .31 to .61 with a median of .45. The average Minimum Mathematics Proficiency Test item difficulty for entering ninth grade students is approximately .60 for the different forms of the Minimum Mathematics Proficiency Test.

Reliability

KR-21 reliability estimates have been calculated for the Minimal Reading Proficiency Assessment and Minimum Mathematics Proficiency Test. The KR-21 reliabilities for the different forms of the Minimal Reading Proficiency Assessment

range from .89 to .91 for random samples of at least 300 ninth grade students on initial testing results. KR-21 reliabilities for the Minimum Mathematics Proficiency Test forms range from .91 to .94 for random samples of at least 300 ninth grade students on initial testing results. A split-half reliability computed on Minimum Mathematics Proficiency Assessment data for a random sample of 479 ninth grade students was .90. Test-retest and equivalent form reliabilities have not been computed at this time.

Validity

Reading. McDonald and Moorman (1974) reported a concurrent validity estimate of .89 between the Minimal Reading Proficiency Assessment and the Davis Reading Test. A correlation of .76 was computed for raw scores on the Iowa Tests of Basic Skills and raw scores on the Minimal Reading Proficiency Assessment for a sample of 151 ninth grade students enrolled in a Title I program. For the same sample, a correlation of .80 was computed for grade equivalent scores on the Iowa Test of Basic Skills and the number of skill areas passed on the Minimal Reading Proficiency Assessment.

The readability level of the different forms of the Minimal Reading Proficiency Assessment range from 7.3 to 7.7 using the formula provided by Fry (1968).

Mathematics. Two estimates of the concurrent validity of the Minimum Mathematics Proficiency Test were obtained from a random sample of 479 ninth grade students who were administered the Minimum Mathematics Proficiency Test and the Shaw-Hiehle Individualized Computational Skills Test (Houghton-Mifflin, 1972). The first estimate (.84) was obtained by correlating total scores on the Shaw-Hiehle Test with total number of skill areas passed on the Minimum Mathematics Proficiency Test. The second estimate (.87) was obtained by computing the canonical correlation between subtests, i.e., addition of whole numbers,

addition of fractions, etc.

Using the percentile distributions for the two tests suggested that grade-equivalent scores on the Minimum Mathematics Proficiency Test range from approximately 3.8 for zero areas passed to approximately 10.8 for all 16 areas passed.

The content validity of the Minimal Reading Proficiency Assessment and Minimum Mathematics Proficiency Test was incorporated within the development of the tests. As described earlier, the objectives for minimal reading and computing proficiency were established, tables of specifications were developed, and items written to measure those objectives.

Student Performance

A summary of student performance on the Minimal Reading Proficiency Assessment and Minimum Mathematics Proficiency Test is presented in Table 1. The percent of entering ninth graders who have passed all the reading skill areas ranged from 15.5 percent for the class of 1974-78 to 19.1 percent for the class of 1972-76. The percent of entering ninth graders who have passed all the math skill areas has ranged from 10.3 percent for the class of 1974-78 to 11.7 percent for the class of 1975-79. Students have steadily progressed toward mastery of all areas such that 94.8 percent of the class of 1972-76 has mastered all the reading areas, and 80.2 percent of the class of 1973-77 has mastered all of the math areas with almost a year and one-half left before graduation.

Table 1

Percent of Students Passing All Areas on the Minimal Reading Proficiency Assessment (MRPA) and Minimum Mathematics Proficiency Test (M^2PT) by Class

Class	MRPA		M^2PT	
	Initial Testing As Freshmen	March 1976	Initial Testing As Freshmen	March 1976
1975-79	18.8%	43.2%	11.7%	44.0%
1974-78	15.5%	77.7%	10.3%	72.3%
1973-77	18.6%	87.8%	10.5%	80.2%
1972-76	19.1%	94.8%	Not Tested	Not Tested

Norms and Utility

Local norms have not been developed at this time. Prior to developing local norms a decision needs to be made as to what purpose the norms would serve, norms for what populations, and what norms could reasonably be developed.

Presently, the Phoenix Union High School System has used the Minimal Reading Proficiency Assessment and Minimum Mathematics Proficiency Test to plan instructional programs for individual students, to monitor student progress in achieving a well-defined set of objectives, and to provide parents and educators with an accountability of progress in meeting objectives. Additional comments, considerations, and recommendations related to the use of norms with these CRT are contained under Summary and Conclusions.

Impact of Basic Skills Assessment Program

The Phoenix Union High School System basic skills assessment program has affected the activities of other school systems in the State. Guidelines, outlined by the State Department of Education for the State of Arizona mandate,

to implement a continuous uniform evaluation system were very similar to those already developed by the Phoenix Union High School System. Thus, the Phoenix Union High School System's basic skills assessment program was available as a model which the Arizona State Department of Education used in developing continuous evaluation systems within the basic skill areas.

The thirteen elementary school districts which feed into the Phoenix Union High School System also have developed a continuous evaluation system similar to the Phoenix Union High School System, through the auspices of an intergovernmental agreement.[1] The Phoenix Union High School System under terms of the agreement provided leadership and expertise in the development of a "continuous uniform evaluation system" which would meet the State of Arizona mandate. This cooperative effort resulted in the development of goals and objectives and CRT in the areas of reading and mathematics which along with those developed by the Phoenix Union High School System covers student skill area development from K to 12.

The Phoenix Union High School System and the 13 elementary feeder districts have developed minimal writing objectives for kindergarten through the 12th grade and are in the process of developing measures of writing proficiency. Writing proficiency will be measured in two ways: student performance on specified writing tasks will be used primarily and CRT related to writing mechanics also will be used, if desired.

---

[1]The Phoenix Union High School System is a 9 to 12 system with approximately 27,000 students in 11 high schools. There are 13 separate and independent K to 8 elementary school districts which feed the 11 high schools in the Phoenix Union High School System.

Conclusion and Summary

The basic skills assessment program has made an impact on the educational systems within the Phoenix area and the State of Arizona. Sets of CRT have been developed. Implications can be derived about the use of traditional test development procedures with CRT.

Educational Effects

The Phoenix Union High School System basic skills assessment program has been effective in increasing the degree to which instruction is individualized. Each student is provided with the instructional program which meets his/her needs. Instruction in the areas of general math and reading is objective-based. A major benefit to the development of a local system, as opposed to purchasing a commercial system or using NRT, was that the teachers within the Phoenix Union High School System were involved in the identification and specification of objectives and consequently are committed to assisting students to accomplish those specific objectives. Often, NRT are used for program evaluations within the Phoenix Union High School System. Teachers frequently ask why a CRT like the Minimum Mathematics Proficiency Test or Minimal Reading Proficiency Assessment is not used "to measure what we teach." CRT are beginning to be used more frequently in program evaluations.

The basic skills assessment program within the Phoenix Union High School System has been effective in establishing a continuous evaluation system or accountability system, assisting teachers and others in assessing student progress in relation to a well defined set of objectives, and producing a coordinated effort among teachers to achieve a specified goal, i.e. assisting students to meet minimal standards in basic skills.

Measurement Implications

The use of traditional test development procedures has proven useful in the development of CRT for the assessment of basic skills in the Phoenix Union High School System. Specifically, traditional measures of item discrimination and difficulty assisted in the selection of items as did traditional procedures for obtaining reliability and validity estimates in assessing the validity of the CRT.

The Phoenix Union High School System has evidence that valid and reliable CRT have been developed for use in the basic skills assessment programs. The reliability and validity estimates cited earlier would probably be judged acceptable by most.

A condition which has enabled the traditional test development procedures to be useful is the fact that ninth grade students enter the high schools with varying states of mastery or non-mastery of specific skills. However, at the end of the ninth grade approximately 60 to 70 percent of the ninth grade students master the minimal reading and math skills. Thus, an attempt to correlate test scores or to compute item analysis based on end-of-year ninth grade performance will likely encounter the problem of limited variability and severely skewed distributions. Plans to conduct validity and reliability studies using end-of-year ninth grade data, and sophomore data are under consideration. These studies might use a dichotomy of mastery or non-mastery of all skill areas or use the raw scores on the tests instead of the number of skill areas passed. In any case, it is likely that limitations in variability of student performance will be noted when data other than ninth graders' test performance is used. It should be noted that while this might be a limitation in some respects, this is further evidence that these CRT are sensitive to what is taught.

The development of local norms is also underway. It is hoped that the development of local norms and that equivalency studies with NRT will yield information about the level of performance which is required for mastery of the minimal basic skills requirements.

The normative level of performance is not important when assessing achievement of objectives in a theoretical sense. In practice, however, school boards, educators, and parents invariably want to know how their students compare to students elsewhere. It might be commendable to state that a high school graduate should possess a specific set of skills. However, if few high school graduates have been achieving the required level of proficiency, it is unrealistic to expect that the implementation of a graduation requirement will alter student achievement immediately.

In summary, the basic skills assessment program within the Phoenix Union High School System has provided a basis for evaluating student progress in meeting minimal reading and math proficiencies requirements. A program will be implemented in the area of writing proficiency in 1976-77. These programs are directly related to the objectives classroom teachers have assisted in establishing. Validity and reliability data have been obtained for CRT used within these programs. Although much work remains, the progress to date suggests a school district can profit from the local development of an objectives-based program. The final assessment of the value of the programs described in this report must await further evidence of the impact these programs have on student achievement. The evidence to data is promising!

REFERENCES

Cox, R. C. and Vargas, J. S. A comparison of item selection techniques for norm-referenced and criterion-referenced tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, 1966.

Fry, E. B. The readability formula that saves time. Journal of Reading, 1968, 11, 513-516.

Haladyna, T. M. Effects of different samples on item and test characteristics of criterion-referenced tests. Journal of Educational Measurement, 1974, 11, 93-99.

Hambleton, R. K. and Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Helmstadter, G. C. A comparison of Bayesian and traditional indexes of test item effectiveness. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 1974.

McDonald, T. and Moorman, G. Criterion-referenced testing for functional literacy. Journal of Reading, 1974, 363-366.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, California: McCutchan Publishing Co., 1974.

Popham, W. J. and Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Woodson, M. I. C. E. The issue of item and test variance for criterion-referenced tests. Journal of Educational Measurement, 1974, 11, 63-64.