DOCUMENT RESUME

ED 176 214 CS 004 955

AUTHOR Dziuban, Charles D.; And Others
TITLE The Application of an Index of Efficiency to a Reading Mastery Test.
PUB DATE [75]
NOTE 11p.; Study prepared at the University of Central Florida

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; *Cutting Scores; *Efficiency; Elementary Education; *Evaluation Criteria; Evaluation Methods; Grade 3; Grade 5; Mastery Tests; Measurement Techniques; *Reading Tests; Testing; *Test Interpretation

ABSTRACT

The Harris Index of Efficiency was used to evaluate the criterion setting (66%) of the Prescriptive Reading Inventory (PRI), a criterion referenced reading test. The subjects, who were involved in a larger experimental study that used the PRI as a criterion measure, were 92 third and fifth grade students. Initial assessment with the Comprehensive Tests of Basic Skills revealed varying student achievement that necessitated the administration of various levels of the PRI on a pre/posttest basis. After these tests, the students were classified as masters/nonmasters using three separate criterion scores (66%, 75%, and 90%). The Harris Index of Efficiency was computed for each criterion score to determine which score proved most efficient at sorting the pupils into two separate groups. The results suggest that the criterion score of 66% was the most efficient cutting point with the PRI; a much less efficient sorting of the subjects was evident when using the 75% and 90% levels. The results also indicate that the Harris Index is an effective descriptive measure that can help to determine at what cutting point a test might yield the most information for instructional purposes. (RL)

THE APPLICATION OF AN INDEX OF EFFICIENCY TO A READING MASTERY TEST

Charles D. Dziuban
Richard A. Thompson
Cavlyne Ussery McGregor

University of Central Florida

The notion of a criterion referenced test has kindled considerable enthusiasm among those concerned primarily with reading assessment (James, 1975). Simultaneous attempts, however, to implement those concepts have given rise to a good deal of controversy. See, for instance, Dole's (1974) response to a recent Thompson and Dziuban (1973) paper and their rebuttal (Thompson and Dziuban, 1974). The futility of trying to "sort out" various claims regarding the use and nature of a so-called criterion referenced test is becoming more and more obvious. A serious student of the movement might be easily overwhelmed by the currently available literature, large portions of which are something less than understandable. Because of this we shall not outline the definitions which abound in the literature, since here we are not so much concerned with a criterion test as with a criterion score.

The problem addressed involves working with an assessment technique (test) which can be used to sort students into a minimum of two groups--mastery and non-mastery. Obviously this necessitates the selection of some criterion or cutting score. A more realistic situation might involve the setting of two scores one which designates mastery when students score above it and a separate score below which signifies non-mastery (Harris, 1974). Those falling between the two scores would be designated neither and for them judgment might be suspended. For this paper we shall consider only the one criterion score case. The problem is to pick the smallest proportion of items passed which might indicate that the pupil has mastered the objective(s).

An Index of Efficiency

Recently Harris (1974) developed an Index of Efficiency for a fixed length mastery test (μ_c^2). By the selection of a cutting score on a K

item device we are able to sort a sample of pupils into two groups and evaluate the efficiency of the decision. Harris (1974) built upon Fisher's (1936) two group discriminant function where the column vector weights were chosen to be 1's, thus creating a special case equivalent to incorporating the proportion of items passed as the "criterion" for discriminating between masters and non-masters. He showed that this was related to the squared canonical correlation, μ^2, which in this case was the squared person product moment correlation between the total test scores and a "dummy" variable indicating the mastery or non-mastery membership of the pupils.(1) The index is positive lying between 0 and 1 with the upper limit being achieved when only two scores occur in a group. Harris (1974) showed that it has two highly desirable features. First, it can be construed as a ratio of true to observed score variance relating it to classic reliability theory. Secondly, the largest μ_c^2 for a particular test is the upper limit to validity when such validity is defined in a like manner. It is the purpose of this paper to report the results of the application of the index to a "criterion referenced reading" test.

The Test, Subjects, and Methods

The Prescriptive Reading Inventory (PRI) is published by California Test Bureau/McGraw-Hill (1972) under the term criterion referenced test. The instrument is organized into four levels which have associated grade level equivalents. It is an untimed test and is purported to facilitate

(1) The reader will recognize this as the squared point-biserial correlation.

the diagnosis of reading skill deficiencies. Mastery for a particular objective is based upon a criterion score usually sixty-six per cent of the items passed--customarily two out of three questions per objective.

The subjects were involved in a larger experimental study in which the PRI was used as a criterion measure. They were ninety-two students in grades three and five attending a central Florida elementary school. Of this number thirty-five were boys and fifty-seven were girls. Initial assessment with the Comprehensive Test of Basic Skills revealed discrepancies in the achievement ranges of the groups and necessitated administration of various levels of the PRI on a pretest/post test basis:

Test Level	Appropriate Grade Level of the Test	Administration Groups
Red (A)	1.5-2.5	3rd Grade-One Group
Green (B)	2.0-3.5	3rd Grade-One Group
Blue (C)	3.0-4.5	5th Grade-Two Groups
Orange (E)	4.0-6.5	5th Grade-Two Groups

Upon completion of the administration of the pre and post test for each subject, a percentage of items correct (P.C.) for the total test was determined:

$$P.C. = P/T$$

where P equals the number of items passed and T is the total number of items on the test. Based upon these scores each student was classified as a master or non-master using three separate criterion scores: 66%, 75%, 90%. For a given student if his P.C. score equaled or exceeded the criterion, he was assigned a 1 (master) and a 0 (non-master) if his score was below the cutoff point. For each group the Harris Index of Efficiency was computed for each criterion score for both the pre and post tests. The objective was to

determine which criterion score proved most efficient for sorting pupils into two separate groups.

Results

The results of the analysis are presented in Table One.

Insert Table One About Here

It may be observed for the third grade red group that the most efficient criterion score for both the pre and post test was at the 66% level (.56 and .81). Substantial decrease in the index was noted as the criterion score was raised. The coefficient for the pretest at the 90% level was undefined. This simply means that no student achieved the 90% level; therefore, all must be classified as non-masters, so at this score the test is completely inefficient for two group purposes. A similar pattern may be observed for the third grade green group with the largest index obtained for both the pre and post test at the 66% criterion score (.68 and .76). The fifth grade blue groups exhibited a similar pattern in that the 66% level proved to be the most efficient cutting point of the three. Again a substantial drop-off was noted as the criterion score was raised to 75% and 90%. The highest indices for the first blue group were .60 and .73 and again .60 and .73 for the second blue group. The fifth grade orange groups revealed somewhat different results in that all indices for the pretests were undefined as they were for all tests at the 90% level. Once again maximum efficiency was achieved at the 66% cutting score (μ_c^2 = .71, .71). Another interesting trend is observable

in that there appeared to be a substantial increase in the indices from the pre to the post tests in all instances where such determination could be made.

Discussion

The results of this study seem to suggest that with this sample of subjects the most efficient cutting point with the PRI was achieved at a criterion score of 66%. This corresponds to the level recommended by the test publishers. If one chose to use the 75% or 90% levels, a much less efficient sorting of subjects would be realized. These scores were determined using full scale scores (based upon all items) and were not related to specific objectives. It is our speculation these shorter tests would be much less efficient in providing information as to which students were masters and non-masters.

It seems clear that the Harris Index as a descriptive measure can be used effectively to help determine at what cutting point the test might yield the most information for instructional purposes. A strategy might involve for a particular group first determining where the test is most efficient and subsequently deciding whether or not that score can be accepted as mastery level. For instance, it is conceivable that a particular test might achieve its maximum efficiency at a score of say 45% but that teachers or curriculum specialists are unwilling to accept that as a minimum mastery score. Further consideration would be necessary regarding the use of the test and design of the reading program. Additionally it is very likely that the same criterion score will not be equally as efficient for

all ability levels, so that the combination of the difficulty of the test and the skill level of students can have a marked effect on the efficiency and thus diagnostic power of an instrument. We point out that we have used the term criterion referenced test very sparingly. We have followed Harris' (1974) example, dealt with a criterion score, and demonstrated that his Index provides valuable information regarding such scores. Accordingly, those concerned with reading mastery and instructional decisions related to it might find these procedures of considerable assistance.

TABLE ONE

Harris Indices (μ_c^2) for Each of the Test Administrations*

	3rd Grade				5th Grade							
	RED		GREEN		BLUE				ORANGE			
	Group	N = 21	Group	N = 22	Group 1 N = 14		Group 2 N = 14		Group 1 N = 10		Group 2 N = 11	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
66%	.56	.81	.68	.76	.60	.73	.60	.73	—	.71	—	.71
75%	.28	.51	.52	.71	.26	.46	.27	.46	—	.52	—	.51
90%	—	.24	—	.43	—	.07	—	.06	—	—	—	—

*— indicates that the index was undefined

REFERENCES

Dole, Phyllis W. "Letter to Crossfire: Criterion Referenced Reading Test in Perspective." The Reading Teacher, (April, 1974), Vol. 27, No. 7, 716-719.

Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems." Annals of Eugenics, 1936, Vol. 111, Part II, 179-188.

Harris, C. W. "Some Technical Characteristics of Mastery Tests." In C. W. Harris, Ed., Problems in Criterion-Referenced Measurement (Los Angeles: Center for the Study of Evaluation, 1974), 98-115.

James, Shirley M. "CTBE and Trends and Practices in the Reading Program." The Reading Teacher, (December, 1975), Vol. 29, No. 3, 267-271.

Thompson, Richard A. and Charles D. Dziuban. "Criterion Referenced Reading Tests in Perspective." The Reading Teacher, (December, 1973), Vol. 27, No. 3, 292-294.

Thompson, Richard A. and Charles D. Dziuban. "Letter to Crossfire: Criterion Referenced Reading Tests in Perspective." The Reading Teacher, (April, 1974), Vol. 27, No. 7, 719-721.