DOCUMENT RESUME

ED 110 858 CG 008 834

AUTHOR Gast, Robert L.
TITLE Sixteen Questions and Responses Concerning the Development of a School District Standardized Testing Program.
PUB DATE 74
NOTE 17p.

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Achievement Tests; *Educational Testing; *Objectives; Program Development; Psychoeducational Processes; *Psychometrists; School Psychologists; *Testing Programs; *Test Results

ABSTRACT

The author, as a testing consultant, has considered and answered 16 questions on the various aspects of testing usually asked by school staff members. The questions and answers are primarily concerned with statistical aspects of achievement tests and the best way to interpret test results to clients, parents and teachers. The need for caution in the choice of particular tests is stressed; thus, tests should be neither too easy nor too difficult for students, and tests should be chosen by a committee who specifies the particular goals for which a test should be administered. The author notes the importance of informing faculty members about the meaning of test scores and how such scores might help them in their classrooms. Moreover, in interpreting a test score one should be aware that a score reflects a current situation resulting from the interaction of various complex variables, so that judgements about a student's academic future should not be based on test scores alone. Finally, the author presents a bibliography of books and journals, dealing with various aspects of testing, to assist testers in any difficulties they might encounter in their work. (SE)

SIXTEEN QUESTIONS AND RESPONSES CONCERNING THE DEVELOPMENT OF A SCHOOL DISTRICT STANDARDIZED TESTING PROGRAM

By

Robert L. Gast

Guidance Testing Consultant

Kansas State Department of Education

120 E. 10th

Topeka, Kansas

1974

SIXTEEN QUESTIONS AND RESPONSES CONCERNING THE DEVELOPMENT OF A SCHOOL DISTRICT STANDARDIZED TESTING PROGRAM

My role over the past six years as a Testing Consultant in the State Department of Education has brought to my attention some recurring questions concerning the development of school district testing programs. The questions have been presented by school staff members in a variety of settings and are usually the ones most often discussed.

As a reader of this paper you may not be in agreement with some of the responses. It is realized there may be no one best response for working with all situations. Even with this caution, the author believes the following responses can be practical in their application:

I. A. "Are we doing the correct amount of testing in our school district?"

B. "Are we over testing in our school district?"

C. "Are we under testing in our school district?"

The above interrelated questions are asking for ways to evaluate the effectiveness of an existing program. To answer any of these questions there are some further questions that need to be asked. Some of these are:

A. Is there a statement of objectives of the testing program in the local school district?

B. Are the stated objectives of the testing program based on current needs of everyone concerned? Do students, faculty, administrators and patrons receive the information they need?

C. Is the current testing program achieving the stated objectives? If so, which test score results have been used to achieve what objectives?

D. Are there tests being administered that serve none of the stated objectives of the testing program?

E. Have all test score results been adequately interpreted to the appropriate persons and groups?

It seems to the author that the statements of under testing or over testing become self-explanatory when these and any other pertinent questions are answered.

II. "Everyone of our 8th grade students who took the DAT was at the 99th percentile rank on the Clerical Speed and Accuracy Test. Can our students really be this good?"

This kind of performance by an entire class is always subject to suspicion; and therefore, more questions must be asked regarding the conditions surround-

ing this particular testing exercise. In this situation, it was found, after further study, that the person administering the test thought the pupils were doing so well on that specific test that it would be a shame not to give everyone just a few more minutes to complete the entire test. As a case in point, proper orientation of staff in the administration of tests could have prevented this testing fiasco.

III. "I have a student who did not try on a test, so what should I do with his test results?"

This should be a concern in any school district. Its occurrence is probably more wide spread than we would like to believe. Thanks to the adequate proctoring, it was observed, in this case, that the student was not responding as expected to the test. Of course, there needs to be follow-up with this particular student exploring the cause and correcting the conditions that caused the situation. The test score results themselves, in this case, should be marked invalid by the proctor and initialed, plus adding other pertinent information that would further describe the situation. If a record is not made of the incident and the test scores are not marked invalid, the test scores are most likely to be used in a manner prejudicial to the student.

IV. "What time of the year should you administer an achievement battery?"

The best time of year to administer an achievement test battery largely depends on when the information is needed.

V. "Too many of our faculty do not use the test score results. Why is this true and what can we do about the situation?"

A possible reason that faculty members do not use test score results may be because of their lack of sophistication in the whole area of testing. If you would check the transcripts of your faculty, you may find at the bachelors or masters degree level few faculty members who have had at least one course that would lead to their understanding of how standardized tests may assist them in their classrooms. To put it bluntly, "you don't work with that you don't know - no more than you can come back from where you ain't been." Because of this lack of understanding in testing, certain types of training opportunities should be made available to the staff. This should include some training in basic statistics, reporting characteristics of tests, knowledge of what tests measure and use of test results.

VI. "At what grade level should we start administering achievement tests?"

The answer to this question has to be based on a study of your local curriculum. For instance, you may measure reading at the primary level because it is taught at that level. The same concept would hold true for other subject matter areas. The answer depends on the grade level the subject matter is introduced. Subsequent measures could follow at appropriate grade levels.

VII. "When interpreting achievement test score results to the faculty for a given grade, should I involve teachers at other grade levels?"

It is crucial that you follow-up with all teachers, especially the former teachers who have contributed to the students' achievement to date. Teachers

of former students should have curriculum feedback, not only as a courtesy to them, but as a means of improving the instructional processes in their classrooms.

VIII. "We like the achievement battery we are administering in our school district; however, we find the test too difficult for our students. What do you suggest?"

Administer the next lower level test to your students. A test that is too difficult for the students taking it will lose some of its measurement qualities. In addition, the students will feel better about the test-taking experience if they can experience at least a modicum of success.

IX. "We have a highly mobile student population; therefore, securing comparable achievement test scores on these students is a great problem. Is there anyway that we can convert any of these raw scores or norms of one test to those of another test?"

Yes, there is a recent study called "The Anchor Test Study" which will soon make it possible to compare directly the results obtained on two or more different reading tests in grades 4, 5, and 6. A manual containing the equivalency and norms tables for all eight standardized reading tests will be available this spring from the U. S. Government Printing Office.

X. "We have our tests machine-scored. On one of our reports called the Frequency Distribution/Local Percentile Norms, we do not have a single student at the 99th percentile rank on any of the test areas. Could there be a mistake? Surely some of our students should be at the 99th percentile rank."

The most common error is that school people forget the definition of percentile rank. By definition, percentile rank is the percent of cases that fall below a selected point. For example, the 75th percentile rank means that 75% of the group fell below that point on the scale. Another way of saying it is that 25% of the group were at that point or above. Thus, the 75th percentile rank might be the highest percentile rank on the report, if indeed, 25% of the cases did fall in that interval or point. If only ten students were included in a local norm group the top scoring student would have a percentile rank of 90 since only 90% of the ten students scored lower.

XI. "What good does it do to go over results with a student who has a low test profile in all areas?"

You must interpret the results for the student in terms of a specific plan of action for his improvement. This should be based on the test score results and other supporting data. There is the fundamental assumption that information yielded will be used in affecting changes in areas needed.

XII. "In selecting an achievement battery, what are some recommended procedures?"

Selection of an achievement test battery is an assignment that should involve the testing committee. The committee should acquire specimen sets of several widely used tests for the appropriate grade levels under study. The test items, and especially the test objectives, should be reviewed by the appropriate committee member in each of the subject matter areas. It is most important that the committee have a systematic procedure for comparing tests so that the most appropriate test can be chosen. The committee should decide in advance what qualities or characteristics they want in a test. The test committee can develop its own form for evaluating an instrument or secure a standard form from most test publishers.

XIII. "Regarding teacher evaluation, would it be appropriate to use our students' standardized test score results for the purpose of retaining or releasing teachers from our system. What are some of the limitations, if any of this procedure?

The Kansas State Department of Education has a position paper on standardized testing, and I will draw the following comments from a portion of that paper:

A single test administration cannot be analytic and diagnostic of isolated factors which are confoundingly imbedded in multiple cause and effect. School achievement is a product of multiple causation; the single score on a test at any given time can only be a sample of specific behavior at that particular time. The score is a function of the interaction of many variables;

the pupil's ability, instructional materials, quality of instruction, environmental influences, previous development, previous experience, conditions of the testing situation, condition of health, and/or many other possible factors. Consequently, a single score cannot be representative of any single causative element, but it is representative of the sum and interaction of all. A score on a single instrument, therefore, cannot be used as a criterion for an imbedded causative element without carefully controlled conditions having been designed to isolate the cause.

In conclusion, if you are talking about accountability, you are talking about more than just testing. Accountability should be positive, and in this case, should be used to enhance teacher development and be of benefit to all concerned parties.

XIV. "What is the purpose of a test committee?" "Who should be on the committee?" "What should be the paramount criterion for selection?"

A test committee, because of its broadly based composition, can provide inputs for helping assure that the program delivers what is desired by all interested parties. The major role of the test committee is to recommend the most appropriate test instrument or battery for a specific purpose.

Every user of test data should be represented either directly or indirectly on the testing committee. This would include school staff members at the primary, elementary, and secondary levels.

Selection of committee members should be based on several criteria. One of the major criteria is the desire to serve on such a committee.

XV. "What are some resources, besides textbooks, that are available to assist us in developing our school district testing program?"

The following resources are listed for your consideration with a summary statement, when needed, describing the intent and purpose of each source:

A. "Nation's Schools" Volume 89, Number 4, April 1972.

There are several articles in this specific issue relating to the releasing of standardized test scores to the public.

B. "Evaluation Comment"

This newsletter routinely discusses controversial issues in the area of assessment. A copy is distributed free of charge to those on their mailing lists. To place your name on their mailing list, write to:

James Burry, Editor
Evaluation Comment
Center for the Study of Evaluation
145 Moore Hall
University of California
Los Angeles, California 90024

C. "Measurement and Evaluation of Guidance" (Journal) and the "Association for Measurement in Guidance Newsnotes" (Newsletter).

The Association for Measurement in Education and Guidance (AMEG) is a division of the American Personnel and Guidance Association. AMEG membership includes a subscription to the AMEG Journal (4 issues) and newsletter (4 issues). The Journal frequently provides a review of recently marketed instruments and the Newsletter provides information on current events in assessment.

Membership application forms are available from the Guidance Section, State Department of Education.

D. "Journal of Educational Measurement" and "National Council on Measurement in Education" (NCME Newsletter).

A membership fee of $10.00 to NCME provides members with four issues of the Journal and Newsletter.

The Journal routinely includes reviews and critiques of a specific standardized instrument.

NCME is an association of individuals that includes objectives to promote greater understanding and improved use of measurement techniques in education. Any person interested in promoting the objectives of the council is eligible for membership. Address your correspondence to:

NCME Secretary-Treasurer
Office of Evaluation Service
Michigan State University
East Lansing, Michigan 48823

E. "Kindergarten Test Evaluations" (1970). Cost $5.00.

"Elementary School Test Evaluations" (1970). Cost $5.00.

"Tests of Higher Order Cognitive, Affective, and Interpersonal Skills" (1972). Cost $8.50.

The above-mentioned test evaluation booklets evaluate commonly used school test instruments in terms of measurement validity, examinee appropriateness, administrative usability, and normed technical excellence. To order, send your request to:

Dissemination Office
Center for the Study of Evaluation
Graduate School of Education
University of California
405 Hilgard Avenue
Los Angeles, California 90024

F. "Guidelines for the Collection Maintenance and Dissemination of Pupil Records." (1970).

The Guidance Section, State Department of Education, has available on loan, the above publication. If you wish to purchase the document, address your correspondence to:

Russell Sage Foundation
New York

G. "A Guide for Keeping Student Records and Getting Rid of Them." American School Board Journal, April 1972.

H. Volume I "Righting the Balance" $2.00 (1970)

Volume II "Briefs" $3.00 (1970)

Price of both volumes ordered together, $4.50

The commission members appointed by the College Board were charged with undertaking a thorough and critical review of the College Board's testing function in American Education and to consider possible fundamental changes in the present college board tests and their uses, and with making recommendations based on their conclusions. Copies may be ordered from:

College Entrance Examination Board
Publications Order Office - Box 592
Princeton, New Jersey 08540

I. "Seventh Mental Measurement Yearbook" (1972) by Oscar K. Buros. Two Volumes. Cost $55.00. Address your correspondence to:

Oscar K. Buros
220 Gryphon Press
Highland Park, New Jersey 08904

J. "The Responsible Use of Tests: A Position Paper of AMEG, APGA, and NCME." This position paper is included in the Journal of Measurement and Evaluation in Guidance: Vol. 5, No. 2, July 1972.

Reproduced copies of the AMEG Journal article are available from APGA.

K. "A Position Paper on Standardized Testing" by the Kansas State Department of Education. Copies are available on request from the Guidance Section, State Department of Education, 120 E. 10th, Topeka, Kansas 66612.

L. "A School District Testing Program Guide" (1971) by the Guidance Section, State Department of Education.

A guide designed to help school staff members take a critical look at their testing program.

Copies available on request from the Guidance Section, State Department of Education, 120 E. 10th, Topeka, Kansas 66612.

M. "Testimony of Dr. Roger Lennon as an Expert Witness on Psychological Testing" (1966). Address your correspondence to:

Harcourt Brace Jovanovich, Inc.
Test Department
757 Third Avenue
New York, N. Y. 10017

Most of the above publications and articles are recent resource publications that should be of help to school staff members when they are confronted with the task of developing a school district testing program.

Resource people that are available upon request to the local school districts include the following:

1. State Department of Education
2. Colleges and Universities
3. Publishers

XVI. "Is there a rule of thumb regarding the relationship that should exist between a students' level of measured ability and of measured achievement?"

A positive correlation is normally expected. Also, it should be noted that the higher the correlation, the greater the chances that the student is achieving up to his potential.

If no such relationship seems to be present when ability and achievement test results are studied, something is wrong. It may be the test is at fault. The chances are, however, that for some reason the youngster is not being challenged to perform up to his potential; only the teacher can determine the real cause of this situation.

In some cases a student may score high on an achievement test and quite low on an ability test. In such a case, the ability measure is probably at fault. It would be highly unlikely that a youngsters' score on an achievement battery could be high due to error or chance factors. Many factors could cause him to score below his actual level on an achievement test, but few

would cause his score to be spuriously high. Furthermore, a youngster could not earn a high score on a standardized achievement test at the appropriate level without a correspondingly high ability potential. So if the ability score is low and the achievement score is high, the ability score is subject to suspect. Another ability measure would be recommended.

The converse is not necessarily true; if a student scores high on an ability test and low on an achievement test, both scores may be valid measures. It is likely that such a student is simply now working up to his potential. When such a situation occurs, the teacher should determine why this is happening and plan a corrective or remedial program for that student.