

DOCUMENT RESUME

ED 129 851

TM 005 471

AUTHOR Crew, John L., Sr.; Whitney, Edward N.
 TITLE Criterion/Objective Referenced Testing: Usages in Some Member Systems of the Council of the Great City Schools.
 PUB DATE [Apr 76]
 NOTE 22p.; 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 *Criterion Referenced Tests; Elementary Secondary Education; Questionnaires; *School Systems; Statistical Analysis; *Surveys

ABSTRACT

To gather some information about the extent of utilization of criterion-objective referenced tests, a survey was conducted among the 27 member school systems of the Council of the Great City Schools. The questionnaire was developed to solicit information in the following five areas: (1) local use of criterion-objective referenced tests; (2) research and evaluation activities related to locally-developed criterion-objective referenced tests; (3) the tendency to compare locally-developed criterion-objective referenced tests with other school systems; (4) the tendency to request other school systems to share their respective developments in the area of criterion-objective referenced tests; and (5) the inclination of teachers to use criterion-objective referenced tests in their instructional activities. The major finding was the indication of considerable interest and usage of criterion-objective referenced testing in the member systems. An examination of the individual items revealed that there is generally limited understanding on the part of classroom teachers, and that little attention is given to the technical characteristics of these tests, such as reliability, validity, and item analyses. (Author/RC) e

 * Documents acquired by ERIC include many informal unpublished *
 * materials not available from other sources. ERIC makes every effort *
 * to obtain the best copy available. Nevertheless, items of marginal *
 * reproducibility are often encountered and this affects the quality *
 * of the microfiche and hardcopy reproductions ERIC makes available *
 * via the ERIC Document Reproduction Service (EDRS). EDRS is not *
 * responsible for the quality of the original document. Reproductions *
 * supplied by EDRS are the best that can be made from the original. *

ED129851

**Criterion/Objective Referenced Testing:
Usages In Some Member Systems
The Council of Great City Schools**

by
John L. Crew Sr.
Superintendent
Baltimore City Public Schools

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

Edward N. Whitney
Staff Director
Office of Pupil and Program
Monitoring and Appraisal

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

an address prepared for the
American Educational Research Association
April 1976

M005 471

**CRITERION/OBJECTIVE REFERENCED TESTING:
USAGES IN SOME MEMBER SYSTEMS OF
THE COUNCIL OF THE GREAT CITY SCHOOLS**
American Educational Research Association, Session No. 14.08

by
John L. Crew, Sr.
Superintendent
Baltimore City Public Schools

and
Edward N. Whitney
Staff Director
Office of Pupil and Program
Monitoring and Appraisal

ABSTRACT

The topics of objective referenced tests and criterion referenced tests have been discussed quite frequently in literature and among school people for the past several years. The extent of understanding and usage across different educational levels is a topic of discussion at a symposium for the annual meeting of the American Educational Research Association. To gather some information about the extent of utilization, a survey was conducted among the twenty-seven member school systems of the Council of the Great City Schools. The questionnaire was developed to solicit information in the following five areas:

1. local use of criterion-objective referenced tests;

2. research and evaluation activities related to locally-developed criterion-objective referenced tests;
3. the tendency to compare locally-developed criterion-objective referenced tests with other school systems;
4. the tendency to request other school systems to share their respective developments in the area of criterion-objective referenced tests;
5. inclination of teachers to use criterion-objective referenced tests in their instructional activities.

A response rate of 70.4% was received to the questionnaire; 19.5% of the respondents sent narrative letters of explanation, and 11.1% did not reply. A five-point scale was used for each of the questions above with one reflecting extensive usage and five reflecting little or no usage. Each of the response sets was defined as including all values between zero and one with values of .50 and larger being placed in the succeeding categories. Thus, it was possible to categorize all responses on the one to five scale and obtain an average for each question.

The major finding was that the questionnaire yielded an average scale value of 2.50. This indicates considerable interest but limited usage of criterion-objective referenced testing in the seven member systems of the Council of the Great City Schools.

An examination of the individual items revealed that there is generally limited understanding on the part of classroom teachers, and that little attention is given to the technical characteristics of these tests, such as reliability, validity and item analysis.

1

**CRITERION/OBJECTIVE REFERENCED TESTING:
USAGES IN SOME MEMBER SYSTEMS OF
THE COUNCIL OF THE GREAT CITY SCHOOLS**

PURPOSE OF THE STUDY

Professional literature reflects extensive interest in the area of criterion-objective referenced tests, proficiency tests, and competency-based assessment tests in the public schools. The primary developmental initiative for such tests appears to be associated with colleges and universities. Since there seems to be a need for an exploration study designed to assess the degree to which local schools systems are developing and using criterion-objective referenced tests, the present study was undertaken.

An ideal population of school systems for determining relative use and development of criterion-objective referenced tests appeared to be the member systems of the Council of the Great City Schools. The Council is an association of 27 urban school districts which looks after city education interests in Washington, D.C.; Atlanta; Baltimore; New York City; Boston; Chicago; Cleveland; Dallas; Denver; Detroit; Long Beach, Ca.; Los Angeles; Memphis; Miami; Milwaukee; Minneapolis; Nashville; New Orleans; Oakland, Ca.; Pittsburgh; Philadelphia; Portland; St. Louis; San Diego; Sacramento; and Toledo. These 27 school systems enroll, collectively, approximately one-fifth of all pupils attending schools in the United States.

Professional journals reflect neither extensive availability of criterion-objective referenced test materials nor studies produced by local school systems identified for this study. Also, two other concerns called for attention; (1) the observation made by staff in local school systems that there is a general apprehension among teachers concerning the utilization of criterion-objective referenced tests results once available; and (2) the need for a national perspective on the position taken by some curriculum decision-makers that teachers can design learning hierarchies needed for appropriate instructional activities, given criterion-objective referenced test results.

METHODOLOGY AND PROCEDURES

A ten-item questionnaire, with "closed-end" response sets was designed by the investigators. A letter of explanation of the purpose of the study was also developed and copies sent to the superintendent and research director in each of the school systems identified. (Please see Appendix A for a copy of this communique.)

The "closed-end" questionnaire was designed to conform to a five-point Likert-type instrument that ranged from a positive value of one (1) which reflected extensive usage; to a negative value of five (5) which reflected little, if any, usage. The computational procedure employed to determine whether the average number of response frequencies should be assigned to one or the other of the five possible response sets is described by the following. Each of the response sets is defined as including all values within the range of .50 smaller than the particular response set value, to .49 larger than the particular response set value (23).

Example:

a. Possible response set values (RSV) are 1 2 3 4 5.

b. The range of possible scores for a particular response set value would include those values that are .50 smaller than (RSV) to those values up to .49 which are larger than the (RSV).

∴ that set of

$$\{ .50 - RSV + .49 = \}$$
 confidence interval for average scores covered by one RSV

c. All frequency averages falling outside the above range would fall into the next response set value that is being approached.

i.e.,	<u>Observed Average</u>	<u>Appropriate Response Set</u>
	1.30	1
	1.71	2
	2.14	2
	3.09	3
	4.75	5

Areas of inquiry as reflected in the survey questionnaire concern:

1. The extent that criterion-objective referenced tests are used locally.
2. The degree to which research and evaluation have been applied to locally-developed criterion-objective referenced testing.



3. The tendency to compare local criterion-objective referenced test developments with other school systems.
4. The tendency to request of other school systems their development in the area of criterion-objective referenced testing.
5. Inclination on the part of the teachers to use criterion-objective referenced tests in their instructional activities.

Statistical analysis used in this survey shall be restricted to summary computations reflecting average response by item, by total questionnaire for local school system, by grand average for items and total test.

RESULTS AND DISCUSSION

Survey questionnaires were sent to the 27 member systems of the Council of the Great City Schools. The returned questionnaires amounted to 19 in number along with various other communications explaining why some forms had not been returned. Reasons given by the five school systems, not returning their questionnaires, ranged from categorically not using criterion-objective referenced tests, to not enough time to fill out the questionnaire by the due date, to sending a copy of a prior-sent correspondence explaining the use of a special language linkage system. A total of three systems did not respond at all.

A special effort was made to assure anonymity of the different systems by not requiring systems to be identified by name. Pre-addressed envelopes with stamps were sent to all systems. The consequences of this approach resulted in a response percentage of 70.4% return of usable data. Figure 1, reflects the extent of returned questionnaires, related correspondence, and no replies.

Figure 1
Percentage of Returned Questionnaires
and Other Information

<u>Type of Correspondence</u>	<u>Frequency</u>	<u>Percent Represented</u>
Usable Questionnaire	19	70.4%
Letters of Explanation	5	18.5%
No Replies	3	11.1%
	27	100.0%

The following data represent summary of item responses, reflecting frequency counts for each response set and an interpretation of the results:

TABLE A
Item Response Frequency and
Interpretation Charts

1. To what extent is criterion-objective referenced testing used in your school system?

- 1** 1. CRTs are currently used in most subject areas.
- 15** 2. CRTs are used on a limited basis in a few subject areas.
- 1** 3. Uncertain as to what extent CRTing is used.
- 0** 4. CRTs may be used to some extent, but I have not heard anything in terms of their instructional usefulness.
- 2** 5. To my knowledge, CRTs are not being used at all in our school system.

Interpretation:

The frequency of response clusters in response set number two (2). The computed average 2.3 reflects a value that indicates limited use of CRTs in a few subject areas.

2. Is the local effort to use CRTs more a function of commercially acquired or locally-developed tests?

- 3** 1. CRTs being used in our schools have primarily been developed by staff in our system.
- 5** 2. While CRTs are used in our system that have been commercially acquired, the majority of CRTs that are being used have been developed locally.
- 1** 3. CRTs, as used in our system, have been equal between commercially and locally-produced tests.

* — = Frequency of Responses

- 3 4. While CRTs are used in our system that have been locally-developed, the majority of CRTs that are being used have been commercially acquired.
- 6 5. CRTs being used in our schools have primarily been acquired through commercial sources.

Interpretation:

The pattern of distribution across the five response categories reflects a clustering of near even proportion above and below response set number three (3). As such, the averaged response of 3.05 for participating respondents would reflect equal usage of criterion-objective referenced tests as acquired locally or commercially.

3. To what extent have local efforts been made to determine the appropriateness or the sufficiency for which pupils can use CRT materials?

- 3 1. We have analyzed pupil performances on our CRTs and have calculated both reliability and/or validity data on these tests.
- 5 2. Based on reactions by our curriculum Specialists and/or teachers, the CRTs being used in our system is adequate for our pupils.
- 9 3. We have yet to start a formal study as concern the adequacy of the CRTs being used in our system.
- 1 4. Based on reactions by our curriculum Specialists and/or teachers, the CRTs being used in our system are generally considered inadequate.
- 0 5. Conventional procedures for determining, statistically, the adequacy or sufficiency of CRTs as used in our system are questionable as to the meaningfulness of their interpretation.

Interpretation:

The response set having the highest number of responses is number three (3), which reflects a lack of formal studies being initiated to determine the degree of appropriateness or sufficiency. In contrast to this, the computed average of 2.32 reflects an adequate level of appropriateness and sufficiency as intuitively perceived by teachers/curriculum Specialists.

4. To what degree are developments in the area of CRTing, as used by your local school system, shared with other school systems?

- 0 1. We regularly disseminate reports of recent program developments to other school systems.
- 11 2. As inquiries are made from other systems concerning program developments such as CRTing, appropriate reports, if available, are sent to these school systems.
- 2 3. I am not certain as to the degree that program developments, such as our CRTs, are shared with other school systems.
- 0 4. While we share program development with other school systems, I am not certain of the usefulness of such sharing, given the local nature of such information.
- 5 5. We rarely disseminate reports of recent program developments such as our CRTing to other school systems.

Interpretation:

A significant clustering of responses has been entered in number two (2). Such a response indicates that local developments are shared readily to the extent available. The averaged response of 2.79 indicates an unawareness on the part of respondents of the extent of sharing that occurs between school systems.

5. To what degree do developments in the area of CRTing, as used by your local school system come about as the result of shared developments as received from other local school systems?

- 1 1. We regularly receive disseminated reports on topics such as CRTing from other local school systems.
- 8 2. Upon notification of available relevant reports, we frequently request such reports.
- 2 3. I am not certain as to the degree that our program developments, in the area of CRTing, have come about as the result of information received from other school systems.

- 0 4. While we have received CRT-type program development reports from other school systems, I am not certain of the usefulness of such sharing, given the local nature of such information.
- 7 5. We rarely receive disseminated reports of recent program developments which would relate to our CRTs from other school systems.

Interpretation:

The pattern of responses are close, being evenly distributed above and below response set three (3). The computed average of 3.05 reflects a lack of certainty as to the degree local CRT developments having come about as the result of shared information received from other school systems.

6. What would be the best indicator of reactions by classroom teachers to existing criterion-objective referenced testing as used in your school system?

- 1 1. The volume of requests for help in developing other CRTs that can be used in other subjects.
- 5 2. An increase in acknowledgement of the usefulness, in the classroom, of statistically analysed CRTs.
- 8 3. I'm not sure what would be the best indicator of teacher reactions to the role of CRTs.
- 4 4. It is quite possible that the current level of teachers using the CRTs is the best indicator.
- 0 5. It is quite likely that, given the controversy surrounding the use of CRTs, such teacher reactions should be kept to a minimum.

Interpretation:

An overwhelming concentration of responses has been observed in those response sets covering two (2) to four (4). Such a range would include attitudes reflecting satisfaction based on overt teacher acknowledgement to capitulation as to what might constitute the best indicator; to, finally, current level of teacher usage as reflecting the optimal indicator. The computed average response of 2.68 reflects a lack of certainty as to what might be the best indicator of teacher reactions.

7. Can the use of the results of criterion-objective referenced testing, by teachers in your school system, be described more as a diagnostic tool or as a product-achievement measure?

- 7 1. It has been the practice of our system to use CRTs both as diagnostic and product achievement tools.
- 12 2. CRTing has been primarily used as a diagnostic tool, while product achievement has been determined by standardized achievement testing.
- 1 3. I'm not certain as how CRTing is used in an instructional capacity by classroom teachers.
- 0 4. While CRTs are used in both a diagnostic and achievement assessing way, there is some question as to how such information is to be meaningfully interpreted.
- 1 5. It has not been the practice by our system to use CRTs both as diagnostic and product achievement tools.

Interpretation:

The response pattern for this item reflects a strong disposition toward separating and restricting the purpose of criterion-objective referenced testing to diagnostic testing, while product achievement testing has been determined through standardized achievement tests. The computed average of 1.89 is well within the range appropriate for response set number two (2).

8. Are criterion-objective referenced tests, as used in your schools, scored by teachers or are they machine scored?

- 4 1. CRTs as used in our system are scored by machine.
- 10 2. CRTs as used in our system are primarily scored by hand, but some machine scoring is done on a few tests.
- 0 3. I'm not certain as to the primary way CRTs are scored in the schools.

- 0 4. While machine scoring of CRTs is available in our system the amount of "turn-around" time involved tends to discourage the use of this service.
- 4 5. Our system does not have a formal CRTing program which might require centralized test processing.

Interpretation:

The response pattern for this item indicates that criterion-objective referenced tests are primarily scored by hand. To the extent machine scoring of criterion-objective referenced testing is done, an equal number of participating systems indicated a lack of facility for centralized test processing. the computed average for this item is 2.32 which falls well within the range appropriate for response set number two (2).

9. What relationships have been determined between pupil performance on achievement test scores and criterion-objective referenced tests in your school?
- 3 1. A high statistical relationship has been observed between performance of our pupils on CRTs and standardized achievement test scores.
- 3 2. While we have not initiated a formal study to make such a determination of relationship, it is felt that pupil performance on such tests is comparable.
- 10 3. I'm not certain as to what the degree of relationship between CRTs and achievement test score performances of our pupils.
- 2 4. While a relationship may exist between pupil performance levels on these two tests, the problem of how such tests compliment each other has yet to be resolved.

- 0 5. Our formal studies have not shown a sufficiently (significant) high relationship between pupil performance levels on these two tests.

Interpretation:

A major clustering of responses has been observed for response set three (3). While this response set reflects indefiniteness as concern the degree of relationship between these two types of tests, the computed average of 2.47 does fall just outside the range appropriate for response set number three (3). As such, the prevailing disposition to this item is that pupil performances are comparable on both tests as described by response set number two (2).

10. Is criterion-objective referenced testing on-going continuously throughout the year in your local school system?

- 6 1. CRTing is used by classroom teachers continuously throughout the year.
- 5 2. CRTing, while used on a voluntary basis by teachers, is primarily used as a diagnostic tool and used at the beginning and ending of the school year.
- 2 3. I'm not certain as to the degree that CRTing occurs in the schools.
- 2 4. CRTing is rarely, if ever, used in our school system.
- 2 5. CRTing, as defined, is not used in our school system.

Interpretation:

Indicated responses were observed in all possible sets. the response set reflecting continuous use (i.e., number one), had the greatest frequency but the computed average of 2.10 falls well within the range of response set number two (2). This response set stresses the voluntary basis, diagnostic tool and specified times during the year for testing.

A generalized reaction to the overall average, Appendix B, which had a scale value of 2.50, is that most of the participating school systems were uncertain as to the extent of criterion-objective referenced testing at the time of the questionnaire. Shaw (19), in his text characterized such a score as ambiguous when reflected on a five-point Likert scale. This overall score tends to reflect a similar level of criterion-objective referenced test activities commensurate with developments and prevailing activities as observed in Baltimore City. In this respect, specific applications for such tests have been produced for use in the area of ESEA Title I pre-K programs affecting three-and-four year olds. Tests have also been developed for use by the Office of Reading and Right to Read for grades Kindergarten-12; and finally, tests have been developed for application in the area of music education. Current efforts in test development include mathematics and proficiency testing for graduation. Formal studies concerning reliability and validity analysis (1) have been completed for the Office of Reading and Right to Read, and formal presentations regarding the Pre-K ESEA Title I program (16) have been presented in 1974 before the Jean Piaget Society in Los Angeles, California.

CONCLUSION

Participating school systems, apparently, have used criterion-objective referenced tests in their systems, but, generally, on a voluntary basis and in limited subject areas.

Generally, such testing and development activities have been restricted to face validity procedures for determining sufficiency. The average response to Item 3, which related to efforts being made to determine the appropriateness or sufficiency of the locally-developed criterion-objective referenced test, reflected an average scale value of 2.32. Such an average falls within the range covered by Response Set 2, which relates to developments based on reactions of curriculum Specialists and teachers. Baltimore City, to this extent, has incorporated the statistical procedure for determining reliability as espoused by Samuel A. Livingston (12) of the Johns Hopkins University in Baltimore, Maryland.

The general lack of understanding on the part of classroom teachers to use criterion-objective referenced testing results reflects similarly to observations made in Baltimore City. The average range for Item 6, which concerns teacher usage, had a scale value of 2.68, well within the range of indecisiveness of the extent of use.

In conclusion, the primary implication for criterion-objective referenced testing is the general lack of research for determining the role for prediction based on criterion-objective referenced parameters similar to Livingston's approach. If the assumption is correct, the progressive performance by a pupil on a statistically-fair criterion-objective referenced test, that has been developed locally, will reflect a corollary achievement rate on nationally standardized tests commensurate with levels of proficiency and increments of growth.

REFERENCES

- ¹An Empirical Application of Livingston's Method of Estimating Criterion-Referenced Reliability, Grades One through Eleven, Baltimore City Public Schools, 1974.
- ²Berube, M. R. and Gittell, M., Editors, Confrontation At Ocean Hill-Brownsville, Draeger, New York, 1969.
- ³Boyd, F. V. and Sciolino, E., "The Decline In SAT Scores", Newsweek, March 8, 1976.
- ⁴Brazziel, W. F., "Criterion-Referenced Tests: Some Trends and Prospects", Today's Education, NEA Journal, 1973.
- ⁵Davis, F. B., "Reporting Test Data In The Media: Two Case Studies", The Reading Teacher, December, 1972.
- ⁶Dyer, H. S., Untitled speech delivered March 22, 1971 to the ETS Conference On Educational Accountability, Washington, D.C.
- ⁷Eaton, D., Editor, "NIE Probes Decline In Test Scores", NIE Information, Fall 1975, Vol. 1, No. 4, p. 1.
- ⁸Education Commission of the States, "Society 'Key' to Writing Decline", NAEP Newsletter, February 1976, Vol. IX, No. 1.
- ⁹Ensor, David, Editor, "Psychologist Urges Closer Look At Testing", Report On Pre-school Education, 1976, Vol. 8, p. 6.
- ¹⁰Gardner, J. W., Self-Renewal: The Individual and the Innovative Society, New York: Harper Colophon Books, 1965.
- ¹¹International Reading Association, "Working Towards Literacy", Right to Read '76, January 1976, Vol. 2, No. 5.
- ¹²Livingston, S. A., "Criterion-Referenced Applications of Classical Test Theory", Journal of Educational Measurement, Spring, 1972, Vol. 9, No. 1.
- ¹³Music Study Evaluation, Baltimore City Public Schools, Baltimore, Maryland, 1975.
- ¹⁴Neil, A. S., Summerhill, Hart, 1960.
- ¹⁵Popham, W. J., Editor, Criterion-Referenced Measurement, Educational Technology Publications, Englewood Cliffs, New Jersey, 1973.

REFERENCES

¹⁶Reid, E. M., Whitney, E. N., Crew, J. L., and Patterson, R. N., The Use of a Piagetian Scale of Development as a part of Growth Measure of Pre-Kindergarten Children Participating in the Baltimore City Public Schools, Baltimore City Public Schools, 1975, ERIC, Ed. 104 961.

¹⁷Robinson, V., Editor, "Value of Title I Reading Programs Is Debatable. Says GAO", Report on Education of the Disadvantaged, January 2, 1976, Vol. 8. No. 27, p. 3.

¹⁸Sexton, P. D., The American School: A Sociological Analysis, Englewood Cliffs, New Jersey, 1967.

¹⁹Shaw, M. E., and Wright, J. M., Scales for the Measurement of Attitudes, New York, McGraw-Hill Book Company, 1967.

²⁰Williams, P. V., and Whitney, E. N., "Disputed Measures of Pupil Progress", The Sunpaper, Baltimore, January 3, 1976, p. A12.

²¹Winter, B., "California official sites 'dramatic' results of experimental basic education plan", The Sunpaper, Baltimore, April 1975, p. 10.

²²Wrightstone, J. W., Hogen, T. P., and Abbott, M. M., "Accountability in Education and Associated Measurement Problems", Test Service Notebook 33, Harcourt Brace Jovanovich, Inc.

APPENDIX A

Cover Letter For Questionnaire

CITY OF BALTIMORE

WILLIAM DONALD SCHAEFER, Mayor

**DEPARTMENT OF EDUCATION**
DEPUTY SUPERINTENDENT
PLANNING, RESEARCH AND EVALUATION
3 East 25th Street, East Wing
Baltimore, Maryland 21218**Dear**

Baltimore City Public Schools have been engaged in departmental activities as regard criterion-objective referenced testing. To this point, our system has endeavored to provide a product and insight to applications and investigate studies as concerns the reliability and validity of such testing materials. Our efforts have, by and large, been restricted to reading diagnostic efforts and curriculum evaluation as regard pre-kindergarten programs. Our interest and purpose for this inquiry is to determine the extent of similar interests, usages and effects of member school districts of the Council of Great City Schools.

We would appreciate being sent information on these activities in your school system. In addition, when a subsequent analysis concerning such activities (interests, usages and effects) is produced by this office, it will be forwarded to participating school systems. It is hoped that samples of materials can be forwarded to us, where possible; and in addition, it is requested that the enclosed questionnaire be responded to and returned to Dr. Edward N. Whitney, Staff Director, Office of Pupil and Program Monitoring and Appraisal, Baltimore City Public Schools, 2519 N. Charles Street, Baltimore, Maryland 21218, no later than December 15, 1975.

Sincerely, I remain**Edward N. Whitney, Ph.D.**
Staff Director**ENW/csj**
Enclosure

APPENDIX B

**Summary of Responses By Participating
Member School Systems**

