
Criterion and Norm-Referenced Testing: Approaches to Educational Measurement

Desalegn Chalchisa^{*}

Educational Measurement approaches are either criterion-referenced or norm-referenced. This paper presents a review of the fundamental concepts of these approaches. It contains definitions, elaborations and distinctions of both type of testing.

1. Norm-Referenced Testing (NRT)

Raw scores obtained by counting the number of right answers are hard to interpret. An examinee's performance would be better interpreted if it could be referenced to something outside of the test itself. As a result, tests require the use of some type of scores derived from raw scores in order to facilitate the interpretation of examinees' performance on them. Traditionally, derived scores such as ranks, percentiles, standard scores, age and grade equivalent scores have been used to report examinees' performance on tests. These types of derived scores indicate the status of the individual with respect to the performance of others. Tests developed to make such derived scores especially useful are called norm-referenced tests (Nitko,

1980). The following operational definition of NRT given by Popham (1981) is commonly used by test specialists.

NRT is used to ascertain an individual's status with respect to the performance of other individuals on the test.

In NRT, an individual's raw score is interpreted by comparing it to the scores of a defined group, often called the normative group (representative sample of testees who have been used as the basis for interpreting test scores). Norms are derived scores of the normative groups (Wiersma and Jurs, 1990). The intent is to compare the performance of an examinee on the test with that of the normative group, rather than to determine how proficient a student is in a particular subject or skill (Capper, 1994).

The following example illustrates NRT by the use of letter grading. Suppose that an English teacher in a

^{*} Hand, Measurement and Evaluation Unit, IER

Criterion and Norm-Referenced Testing: Approaches to Educational Measurement

Desalegn Chalchisa^{*}

Educational Measurement approaches are either criterion-referenced or norm-referenced. This paper presents a review of the fundamental concepts of these approaches. It contains definitions, elaborations and distinctions of both type of testing.

1. Norm-Referenced Testing (NRT)

Raw scores obtained by counting the number of right answers are hard to interpret. An examinee's performance would be better interpreted if it could be referenced to something outside of the test itself. As a result, tests require the use of some type of scores derived from raw scores in order to facilitate the interpretation of examinees' performance on them. Traditionally, derived scores such as ranks, percentiles, standard scores, age and grade equivalent scores have been used to report examinees' performance on tests. These types of derived scores indicate the status of the individual with respect to the performance of others. Tests developed to make such derived scores especially useful are called norm-referenced tests (Nitko,

1980). The following operational definition of NRT given by Popham (1981) is commonly used by test specialists.

NRT is used to ascertain an individual's status with respect to the performance of other individual's on the test.

In NRT, an individual's raw score is interpreted by comparing it to the scores of a defined group, often called the normative group (representative sample of testees who have been used as the basis for interpreting test scores). Norms are derived scores of the normative groups (Wiersma and Jurs, 1990). The intent is to compare the performance of an examinee on the test with that of the normative group, rather than to determine how proficient a student is in a particular subject or skill (Capper, 1994).

The following example illustrates NRT by the use of letter grading. Suppose that an English teacher in a

^{*} Hand, Measurement and Evaluation Unit, IER

college administers a final examination to a total of 180 students. The teacher decides that the distribution of letter grades assigned to the final examination result will be 5 percent A's, 15 percent B's, 60 percent C's, 15 percent D's and 5 percent F's. Suppose that a student obtains a score on the final exam such that 30 students have higher and 150 students have lower scores. His score on the exam will be B, because the top 9 students (5 percent of 180) will receive A's, the second top 27 students (15 percent of 180) will receive B's, the middle 108 students (60 percent of 180) will receive C's, the second bottom 27 students will receive D's and the last 9 students will receive F's. Counting from top to down, the student's score is located 31st, and he will, therefore, receive a grade of B.

In this example, the raw score of the student was not specified. The raw score would have been necessary to determine that the student's score is located 31st in relation to the 180 scores. However, in terms of interpretation of the score, it was based strictly on the student's relative position in the total group of students. Thus, this kind of interpretation specifies the performance of examinees in relative and not absolute terms.

2. Criterion-Referenced Testing (CRT)

CRT is relatively a new phenomenon in the history of testing. The concept of CRT was first introduced by Glases in 1963 (Poham, 1981; Wiersma and Jurs, 1990). He came up with the idea of derived scores that directly reflect the kind of performance that an individual can show or do rather than by his or her relative standing in a defined group examinees. He referred to this sort of notion as CRT (Nitho, 1980). The derived scores to be obtained from criterion-referenced test scores provide information about the degree of competence/mastery attained by a particular examinee along the continuum of achievement irrespective of the performance of others. A student who does not attain the criterion has not mastered the skill sufficiently to move to the next instructional level.

3. Fundamental Distinctions Between CRT and NRT

3.1. Breadth of the test

A norm-referenced test measures a more general category of behaviours like arithmetic skills whereas criterion-referenced test focuses on a more specific domain of behaviours such as solving addition problems with two

three-digit numbers or determining multiplication products of one and three digit numbers. CRT tends to focus on sub-skills more than on broad skills (Ebel, 1979; Popham 1981; Glaser and Nitko, 1971).

3.2. Way of interpretation of test scores

It is the kind of interpretations that basically distinguishes CRT and NRT. Norm-referenced interpretations are based on an individual's standing on the test relative to others. Any interpretations via ranks, percentiles, deciles, quartiles or standard scores (Z-scores, T-scores, Deviation I.Q scores, Stanines) is norm-referenced. CRT does not depend on comparisons of the performance of other examinees; rather, the score of an examinee on criterion-referenced test must yield direct information about the individual's performance on some criterion of interest independent of the test scores earned by any other examinee. There is a general agreement that criterion-referenced test scores must be referable to a well defined domain of behaviours (Hartel, 1985; Nitko, 1980; Popham, 1981) and to well defined domain of behaviours or defined performance level (Wiersa and Jurs, 1990). How well an examinee mastered a well defined behavioral domain or a defined performance level on a criterion-referenced test

is interpreted via percentages.

Both NRT and CRT approaches continue to exist and play important roles in accomplishing specific purposes in educational measurement and evaluation (Hamey, 1984).

REFERENCES

- Capper, J. 1994. Testing to Learning...Learning to Test: A Policy maker's Guide to Better Educational Testing.
- Ebel, R.L. 1979. Essentials of Educational Measurement (3rd ed.).
- Glaser, R; and Nitko, A.J. 1971. 'Measurement in learning and instruction'. In: R.L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, DC: American Council on Education.
- Gronlund, N.E. 1985. Measurement and Evaluation in Testing (5th ed.). New York: MacMillan Publishing Company.
- Hamey, W. 1984. 'Testing reasoning and reasoning about tests,' Review of Educational Research, 55, 597-654.
- Hartel, E. 1985. 'Construct validity and criterion-referenced testing,' Review of Educational Research, 55, (23-46).

Ministry of Education,
1980. Geography: Grade
8. Addis Ababa:
Educational Materials
production and
distribution Agency.

Nitko, A.J. 1980
'Distinguishing the many
varieties of criterion
referenced test'. Review
of Educational Research
50, (461-486).

Payne, D. 1992. Measuring
and Evaluating
Educational Outcomes.
New York: MacMillan.

Popham, W.J, (1981). Modern
Educational Measurement.
Englewood Cliffs:
prentice Hall.

Wiersma, W; & Jurs, S.G.
1990. Educational
Measurement and Testing
(2nd ed.). Boston:
Allyn and Bacon.

IER Conducted a Workshop on Measurement and Evaluation in Classroom Learning in Addis Ababa Schools

The Institute of Educational Research held a short-term training workshop from July 25 - 29/94 on Measurement and Evaluation in classroom learning in Addis Ababa schools. The training was conducted in response to

the need for improving teachers' skills in the area of testing. About 25 senior secondary school principals, unit leaders, department heads and teachers participated in the workshop.



Opening Session

The workshop specifically deals with the following.

- . student performance measurement;
- . achievement tests;
- . validity and reliability of test scores;
- . preparing and administering classroom tests in high schools and
- . the use and application of tests for guidance of students.

The workshop was opened by Dr. Makonnen Yimer, Vice President for Administration and Development of Addis Ababa University. In his opening remarks, the vice president stated that

Addis Ababa University had an increasing role in the development of the system of education through the direct involvement of its graduates and the sustained contributions of its researchers. He added that the training session was in line with the objectives of the university and in particular with that of the Institute of Educational Research.

The workshop was closed by Ato Tamirat Demessie, the former Head of the Education Bureau of Region 14. After giving certificates to the participants, Ato Tamirat made a closing speech in which he emphasized the need for similar workshops for the vast majority of teachers in outer schools. The same was expressed by the participants of the workshop.