

Criterion and Norm Referenced Tests in the Education System: Livening Up the Debate

Mashingaidze, S

*Faculty of Education, Applied Education
Midlands State University
Gweru, Zimbabwe*

Abstract

Any form of assessment should support purposes of learning and reflect the principles of curriculum for excellence. Curriculum policies ought to recognize the importance of assessment as an integral part of teaching and learning. Thus tests as forms of assessment are fundamental for effective teaching and learning to take place. Generally tests serve varied purposes depending on the type of test. The purposes of tests range from enabling teachers to identify their weaknesses and strengths of instruction for future adjustments to informing parents/guardians of their children's progress in learning. It is through tests only that the effectiveness of most education programmes can be determined. While tests can be classified into two categories, essay and objective tests, they can also be classified according to their purposes as norm referenced tests and criterion referenced tests. School teachers therefore need to acquaint themselves with the subject of testing, that is, learn more about testing systems, how to interpret test results, etc. It is in the thrust of this paper to present comparatively the epistemological dimensions of each of the two forms of tests.

Key Terms: Criterion referenced test (CRT), Norm referenced test (NRT), Discrimination index, Difficulty/facility index.

Introduction

Formal examinations or tests have not always been part of education. The first formal examination occurred at University of Bologna in 1219 (Bude & Lewin, 1997). Another significant figure in the history of educational testing is *E.L. Thorndike*. In 1900s Thorndike and his students at Columbia teachers' college contributed tests of arithmetic, handwriting, spelling and other academic abilities (Graham et al, 1984). Today achievement tests are a part of virtually every academic endeavor. Thus tests have become part and parcel of education systems.

A test result can provide varied meanings depending on, for instance, the purpose of testing. Some type of information tells us where a student stands compared to other students. Thus test data can help us determine a student's rank by comparing the

student's performance to a norm or average of performance by other, similar students (Kubiszyn and Borich, 2003). This kind of test is called a norm-referenced test.

Gay (1980) defines a norm-referenced test as any test that reports and interprets a score in terms of its relativity with other scores in an identical test. With a Norm-referenced test a representative group of students is given the test prior to its availability to the public. The scores of the students who take the test after publication are compared to those of the norm group (Bond, 1996). In other words, norms compare a pupil to other pupils in the same class or the same age in a school. In norm-referenced tests a student's performance is rated *visa -vis* the performance of his/her peers (Siann & Ugweugbu, 1984; Kubiszyn & Borich, 2003; Glaser, 1963).

Norm referenced tests thus compare a pupil's performance to other pupils in the same class or same age range in a school. In other words, the test provides information about how much more or less a pupil has the subject content in comparison to others. Therefore a norm-referenced test refers to a test, be it standardized or locally developed, which reports and interprets each score in terms of its relative position with other scores on the same test. Norm referenced tests entirely ends in grading of students. The basic principle on this kind of a test is to provide a continuum in performance ranging from the highest to the lowest in the group. The score or mark assigned to a student usually represents his/her place on the continuum. The group as a whole has a normative function of setting the standards. Such groups can be grade seven pupils proceeding to secondary education, O-level students proceeding to A-level or A-level proceeding to tertiary education. According to Siann and Ugweugbu (1984:236), "This procedure as a whole tells us nothing about the performance in absolute terms of an individual." In this view the form of testing cannot reveal actual weaknesses of a student, but only distinguishes the different levels of abilities among students. Thus the reference group may be bright or weak and this way only allows us to rank pupils.

According to Gay (1980:42), "Any test which reports and interprets each score in terms of an absolute standard is criterion -referenced." Zindi, Peresu and Mpofo (1998: 150) define a criterion -referenced test as, "a device which measures whether a student has not reached the criterion or specific level or achievement". Siann and Ugweugbu (1984:238) view criterion -referenced tests as measurement devices, "concerned with the individual's status with respect to some criterion -referenced performance standard". In other words, a criterion is what a test is designed to predict. For instance, considering the topic of Locus in mathematics, the criterion can be stated as, "ability to construct a shape given its sides and angles". An item which addresses this criterion can be put as, "construct a triangle ABC given that $AB = 5$ cm, $AC = 6$ cm and angle $BAC = 30$ degrees". In this example, a student is assessed against the criterion "to construct a shape given its sides and angles".

The three authorities cited above seem to agree that a criterion-referenced test allows educators to ascertain an individual's status with respect to a well-defined behavior domain. In agreement with this assertion, the author of this paper views a criterion-referenced test as a measurement device whose score does not depend on comparison with other scores. It instead flows directly from the connection between the items and the criterion.

A criterion-referenced test only reports and interprets each score in terms of an absolute standard. Therefore in a criterion-referenced test, interpretation of one person's score has nothing to do with anybody else's score (Kubiszyn and Borich, 2003). The comparison is with the standard of performance or objectives set not scores of other persons. Glaser (1970:23) sums it up by referring to it as, "a test that is deliberately constructed to yield measurement that is directly interpretable in terms of specified performance standards". It can however be inferred that while the two concepts *CRT* and *NRT* have different connotations they are used concomitantly within school systems.

Differences between Norm-Referenced and Criterion-Referenced Tests

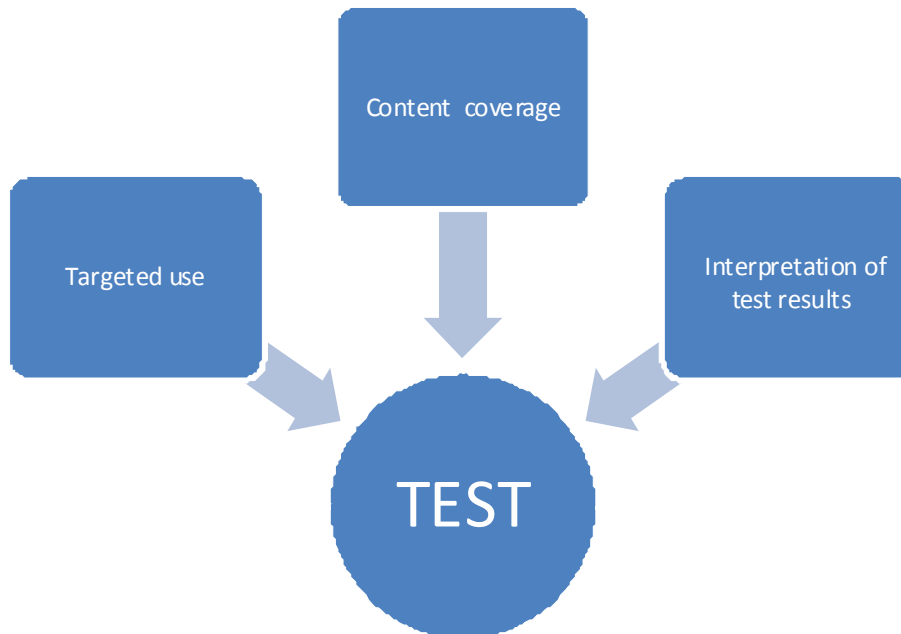


Fig. 1 **Three factors of a test** Adapted from Bond, (1996)

The differences between *NRTs* and *CRTs* can be looked at in terms of three vital strands, as shown in Fig. 1. These include; interpretation of test results, targeted use and content coverage. By *content coverage* (selection of content) it means the amount of content of the syllabus to be selected for testing. *Targeted use* means the driving force behind the use of a particular test, since any test constructed has to serve specific

purposes. Whilst *interpretation of test results* means using test results accordingly, since a test's interpretation is based on the purpose for which it should serve.

Content Coverage (Depth)

The content for *NRTs* is selected as to how well it can rank student performance, whereas the content for *CRTs* is selected as to how well it matches the learning outcome deemed most relevant (Bond, 1996). In other words, *CRTs* require that test items be selected on the basis of how well they reflect the behavior specified in the instructional objectives. Since *CRTs* are interpreted in terms of an absolute performance on some clearly defined set of learning tasks, it is important that the test items be a direct measure of expected learning outcomes. This means the behavior called forth by each test item should match a specific behavior described in the instructional objectives.

NRTs cover a larger domain of learning tasks with a few test items, whilst *CRTs* cover a delimited domain with a relatively large number of test items (Kubiszyn and Borich, 2003). *NRTs* emphasize on selecting the test items of average difficulty dismissing items which can easily be answered by all or most of the students and items that are too difficulty, yet *CRTs* match tasks with test items. The *facility/difficulty index*, obtained by dividing number who got item correct by total number of test takers, in *NRTs* is estimated to be in the range 0,40 to 0,70 since this test leaves out the test items which are too easy and too difficulty (Kubiszyn and Borich, 2003). "In *CRTs*, about 80% of students completing a unit of instruction are expected to answer each item correctly, whereas in *NRTs* about 50% are expected to do so", (Kubiszyn and Borich, 2003: 61).

The *discrimination index*, which defines how well test items discriminate between those pupils that are good and those regarded as weak, is expected to be positive in *NRTs*. In the *CRTs* the *facility index* may fluctuate from very high to very low since the level of difficulty of the items depends on the objectives. "In *CRTs*, items tend to be equivalent to each other in difficulty", (Kubiszyn and Borich, 2003: 61). The mixing of items is meant to attract and motivate the student by the easy questions first and then the challenging ones later. That is test items that are difficulty and test items that are easy may both be found in a *CRT*. When considering the *discrimination index* in *CRTs* it is not far from zero in the positive since *CRTs* do not need to discriminate amongst students at all. In other words, *NRTs* emphasize the discrimination in pupils while *CRTs* emphasize the description of what learning tasks individuals can and cannot perform.

Interpretation of Test Results

It is important to note that test interpretations differ in the two forms of testing. Most *CRTs* do not undergo a rigorous evaluation process for content or difficulty level (<http://www.casly.org>)

Norm-referenced tests use a relative standard while criterion-referenced tests use an absolute standard. Relative standard does not tell us much about an individual's absolute achievement. There is no indication of what one knows and can do or can not do. This form of test (NRT) does not indicate the degree of excellence or deficiency of student's performance in relation to peers. A pupil's score is compared to that of the persons in the norm group. On the other hand an absolute standard on which *CTR* is based is specific and objective-oriented, for example a pupil could be expected to multiply fractions in decimal form. Interpretation therefore is based on whether the student has achieved the objective or not and by how much.

In the case of norm referenced tests, a final grade received by a pupil in a subject obscures the pupil's misunderstandings, inadequate study skills and potential limitations in that subject. In other words, *NRTs* have to do with the ranking of scores and so do not tell us much about how much content pupils know in a subject. Contrary to *NRTs*, *CRTs* focus on the ability of the individual pupil to carry out specific tasks.

Nonetheless, it is helpful to both the individual pupil and society to know where a pupil's relative strengths and weaknesses lie. One way in which the weaknesses of an individual can be known is through comparing with the performance of his or her peers. While ranking does not tell us much about skills – oriented performance, it helps reflect on the individual pupil's position in relation to the performance of others. School progress reports and prize giving are basely centered on the relative standings. In other words, information about how well students are performing relative to others is, at least, valuable feedback to school systems. Thus while the results of an *NRT* does not tell much about actual performance, they provide information which school systems would use as basis for rewarding learners who excel in their groups. For instance the best three performers in a test in a class would be identified based on this system of testing.

NRTs interpret test results in relation to an identified group, which means no group no interpretation of *NRTs* whereas *CRTs* interpret results on the basis of an identified area of achievement. For example, acknowledging that the learner can or can not calculate area of a circle in mathematics. In other words, to interpret an *NRT* requires a clearly defined group of students, whereas with a *CRT* it requires a clearly defined and delimited area of achievement (domain).

Targeted Use

The two tests also differ in terms of the purpose they serve. The purpose of an *NRT* is to differentiate among individuals. Norm referenced tests tell us nothing about performance in absolute terms of an individual. The whole group providing the standard could be an above average or below average group and the scores only

help in ranking pupils. Whilst criterion referenced tests tell us more about the performance in absolute terms. *CRTs* hence provide information regarding to what an individual pupil can do, not how he or she stands in comparison to others. It focuses primarily on the uses to which knowledge or skill it is testing is to be put. Since criterion referenced test are more to do with performance of individuals its results are not necessarily ranked. Consolidating this view Gay (1980:146) says, "*CTRs* are not suitable for use in large scale evaluation. Many of them are designed for classroom use, not for progress evaluation". It follows therefore that score variability is very important in considering *NRTs* whilst irrelevant in *CRTs*. By variability is meant the spread of scores in a test and such information is of significance to the teacher as it indicates the differences in ability in a class.

Labeling and grading which are synonymous with norm referenced test tend to depress the performance of students who do not fall in the cream of the class, say the top ten. In support of this view Siann and Ugweugbu (1984:237) say, "while a certain amount of stress may be desirable to prompt optimum motivation greater levels of stress may lead to fairly severe emotional disturbances". Pupils who always come last in class performance are more likely to develop an inferiority complex, and such is likely to depress zeal to learn in most pupils. Grading results in hording pupils as regard their performance. Pupils are put into classes defining different performance levels such that, for instance, a '*U*' grade in Zimbabwe signifies an ungraded performance, a poor performance. So *NRTs* are group tests taken by whole classes in the case of schools so that comparison is between pupils in a given class/category.

In spite of its many limitations, norm- referenced tests prove to be quite handy when it comes to the selection of pupils for further learning. Siann and Ugweugbu (1984:237) say, "what ever the psychological disadvantage of norm- referenced testing, it is the only method of assessment when we are selecting candidates for a comparatively small number of places". For instance, in Zimbabwe progression to A-Level is based on the performance on the ZIMSEC –O level examinations (Kanyongo, 2005). Progression is on merit and schools set selection criteria. Only those students with good passes proceed to this level of education. Students, thus secure places for A-Level classes based on their norm-referenced performance. The criterion-referenced performances which hardly involve ranking or positioning are not an ideal where bottlenecks are involved.

The two tests also differ in paradigm approach. A criterion referenced test is more learner-centered as it can be used to address learner needs. "The purpose of testing and teaching are combined in the teacher, the results of testing are readily turned into instructional programmes." (Mpofu, 1991: 3). Its focus is to take up the learner through the learning ladder. It can be used to ensure learner is coping with learning, through diagnosing learner difficulties and insisting on remediation. Thus *CRTs*

can allow for full intellectual growth of individual learners. However this becomes more demanding on the part of the teacher. A good number of skills ought to be displayed by the teacher, in terms of assessment and evaluation techniques, accurate and consistent record keeping, and sacrificing time for quite a number of other things. Thus *NRTs* have a terminal effect, grading and ranking performance, of coming in late and failing to improve the individual learner. *NRTs'* traditional approach to learner assessment fails to correct misconceptions.

CRTs are somewhat used in formative evaluation where assessment occurs continuously during teaching and learning. In this view, there is frequent testing which is cumulative and meaningful. The testing is usually conducted during or after every lesson or unit. Thus it is used in the diagnostic evaluation where the teacher attempts to find out causes of learning difficulty. Essentially the teacher doesn't have to wait for a long period before testing for learning. The results of such testing prove quite useful to the teacher, students and the parents. They are enlightened on the students' strengths and weaknesses and so adjustments can be prepared accordingly. This provides vivid feedback to the school administration on curriculum areas that require their attention, for instance, organizing workshops for mathematics teachers to improve on their teaching.

Similarities between Norm and Criterion Referenced Tests

Although norm referenced tests and criterion referenced tests differ in many respects as highlighted above they have similarities. Both forms assess performance of individual learners. In other words, both *CRT* and *NRT* are important in the assessment process—they complement one another. "Both are essential for ensuring credibility in the assessment process and communicate information regarding the need for intervention and the type of intervention", (Montgomery, 1987: 1873). Such information is essential not only to the teacher, but even to the learner and the learner's parents.

A standardized test is conditionally taken to be a norm-referenced test, but criterion-referenced tests also may be standardized. *Standardization* is the process of administering a test under uniform conditions to each child who is to be tested (Montgomery and Connolly, 1984:1873). Thus a standardized test is one that employs uniform procedures for administration and scoring in order to assure results are comparable. A normative test is a test that has been administered to large numbers of groups of students. A test therefore may be standardized in its administration, but not normed as in some *CRTs* (Montgomery, 1987).

Both forms of tests review objectives that have been covered in teaching and learning and are usable as pointers of student capabilities, teacher capabilities in teaching and can be used as instruments to judge a students' worthiness to graduate in one course at the end of it and move on to another level.

Structure of Zimbabwe's Education System

The Zimbabwean education system consists of primary education, secondary education and tertiary education. The primary level is a seven-year cycle and it runs from Grade 1 through Grade 7 during which pupils will be largely exposed to criterion – referenced testing in preparation of the end of cycle examination (norm-referenced testing) (Kanyongo, 2005). The subjects taught in primary schools are: Mathematics; English; Shona and Ndebele (Indigenous languages); and General Paper covering Social Studies, Environmental Science, and Religious Education (largely based on Christianity) (Kanyongo, 2005). At the end of Grade 7, students are tested in the four subjects. Since primary education is compulsory and is guided by the policy of unimpeded progress, performance on the Grade 7 examination does not necessarily affect the progression of the students to secondary education. However, some secondary schools are selective and they set selection criteria based on the Grade 7 examinations.

Secondary education starts in Form 1. Pupils completing primary education enrol into Form 1. In most cases schools use Grade 7 national examination results to place students in remedial or gifted programmes (Kanyongo, 2005). This enables teachers to plan properly for their teaching knowing the class's level of competence. Like primary education, the secondary curriculum is centrally designed by the CDU in the Ministry of Education, Sport and Culture. Secondary education comprises a four-year O-Level cycle and a two-year Advanced Level (A-Level) cycle. The O-Level cycle covers a wide curriculum. Officially, a student should take a minimum of eight subjects in secondary education (Kanyongo, 2005). At the end of the four-year cycle, students sit for the Zimbabwe schools examination council Ordinary Level (ZIMSEC-O) examinations, which is a norm-referenced test. A student should pass a minimum of five subjects, which include Mathematics, English and Science (Kanyongo, 2005).

After O-Level, a student may choose to proceed to A-Level or go to any of the following: teacher's training college, technical college, agricultural college, polytechnic, and nursing training college. Progressing to A-Level is based on the performance on the ZIMSEC-O examinations (norm-referenced testing). Thus progression is largely based on norm-referenced test results. Only those students with good passes proceed to this level of education. At A-Level, students major in a minimum of three subjects. The choice of subjects is usually based on the students' long term career goals (Kanyongo, 2005). The subjects one picks at A-Level will determine the degree program one will study at the university level. For example, a student who wishes to study Law may consider subjects like English, English Literature, and History while a student who wishes to study Engineering may consider subjects like Chemistry, Physics and Mathematics.

The Zimbabwe Schools Examination Council (ZIMSEC) is responsible for all national examinations in primary and secondary education, which are Grade 7

examination, Zimbabwe Junior Certificate of Education (ZJC) [currently not being offered], Zimbabwe General Certificate of Education Ordinary Level (ZGCE O-Level) and Zimbabwe General Certificate of Education Advanced Level (ZGCE A-Level) examinations (Nziramasanga, 1999). The Grade 7 examination which takes place at the end of primary education tests students in four subjects: English, Mathematics, Shona or Ndebele, and General Paper (Kanyongo, 2005). The main purpose of the Grade 7 examination is certification of the students' level of educational achievement. It is also used for the selection of students to secondary education. Many schools, especially those in rural areas, have a 'mass admission' policy regardless of the students' results on the Grade 7 examination. This is because of the government policy of education for all, so that no student should be denied a place for whatever reason.

The next examination takes place at the end of Form four. Students pay examination fees to take this examination. This examination serves a number of purposes. First, it certifies students' level of educational achievement. Second, it is used for selection to go to A-Level. Third, it is also used by employers for hiring purposes as well as for admission to other institutions of higher learning like teachers' and nurses' training colleges (Kanyongo, 2005). A letter grading system is used as follows: A, B, C, D, E, F, and U; with A being the highest grade achieved and U (Unclassified) being the lowest. Grading used is largely based on students' performance nationally - thus norm-referenced. The grades are assigned for each subject, and C is the minimum acceptable passing grade. A student should get a minimum of 5 Cs including English to have successfully completed Ordinary level (Nziramasanga, 1999).

The final examination of the secondary school education system is the ZGCE A-L examination, taken at the end of Form 6. Results for this examination are used for: (a) certification of student's level of educational achievement, (b) selection to the university and other institutions of higher learning, and (c) employment purposes. Grading of this examination is based on a seven-letter grading system as follows: A, B, C, D, E, F (fail), and O (Kanyongo, 2005). The O indicates that the student produced work that is equivalent to O-Level standard. Criterion referenced testing is useful in preparing students for the final examination in that it helps in the building up of concepts in students. This is achieved largely through teacher made tests and exercises.

Critique

Norm-referenced and criterion referenced tests have both strengths and weaknesses as regards their use in the education system. Mississippi State University (2004) notes that norm-referenced scores derive their meaning only from comparisons with other scores - scores of other students, scores of the same students at different time periods, or scores from other colleges and universities.

Norm referenced testing is valued and seen as an indispensable part of teaching and learning (Bude & Lewin, 1997). This form of testing has the advantage of encouraging competition which brings out the best or the top candidates. In other words, norm referenced tests can serve the purpose of sorting students according to performance. One reason of assessing students is to be able to label them on the basis of their performance. But are we doing this in order to segregate students by ability and teach them separately?

This kind of sorting has been criticized but for different reasons. One of the criticisms is that sorting of individuals may result in putting them into wrong piles. Some studies suggest that student performance does not improve when teachers grade more stringently (Abraham et al, 1980). Are grades reliable enough to allow students to be sorted effectively? In reality any one teacher may give different grades to a single piece of work submitted at different times, for instance. Whatever use we make of sorting, the process itself is very different from and incompatible with the goal of helping students learn. Thus the purpose of assessing students should not be to sort them, but simply to provide feedback so they can learn more effectively in future than they are doing in the present. From this perspective, this is an entirely legitimate goal of assessment. Grades are an entirely inadequate means of reaching this goal. Learners should experience success and failure not as reward and punishment but as information (Bruner, 1961). This means a letter or number attached to a learner's work is not helpful, and tells a student little about what impression was his/her effort or how could it be improved.

However competition, engendered by ranking in the case of norm referenced tests, can serve as an extrinsic motivator. That is the student might want to learn for the sake of out competing colleagues. Whatever the psychological disadvantages norm referenced tests could have, they are the only method of assessment when candidates are being selected for comparatively small numbers of places. Norm referenced tests also have the advantage of going beyond the classroom borders. In other words, when communicating with parents and other interested parties through percentiles, percentages and grades information is passed on to stakeholders and the outside world at large.

Assessment may also serve to motivate students to work harder. Indeed given the extent to which letter grades function as rewards and punishments rather as useful feedback, a critical difference exist between intrinsic and extrinsic motivation, that is, between an interest in which one is learning for its own sake and a mindset in which learning is viewed as a means to an end.

On the other hand, norm referenced tests have limitations regarding their use in school. Comparing a student's performance against the scores achieved by others conceals the student's misunderstandings, inadequate skills and potential

limitations in the subject area. To be meaningful any individual 's score needs to be related to the tests. A mere mark does not signify a definite amount of knowledge and hence has very little relevance for content or absolute framework reference. Ranking creates unnecessary competition which can cause a lot of stress and anxiety and also can even cause disharmony in the class. It can even result in labeling which demotivates low achievers. Another drawback of NRTs has been a result of the attention that national examinations receive – even to the extent that national exams are announced over public radio and television (Bude & Lewin, 1997). Schools compete vigorously to top the lists national examinations. The incentive to cheat and find illegal means to pass the examinations has created many problems (Bude & Lewin, 1997: p 10). Heavy emphasis on examinations leads to a neglect of the broader pedagogical tasks of the schools.

Criterion referenced tests have the advantage that a score is compared with a standard of performance not with scores of other persons. In other words, when criterion referenced tests are used, everyone taking the test may do well or everyone may do poorly. The meaningfulness of an individual; with other candidates but with what an individual can do. In fact each item is chosen to represent a particular aspect of the skill or knowledge that a teacher wishes the candidate to master. What the examiner is interested in is whether the candidate has reached the criterion skills or not. Above all this method is the only one which pin-points where the strengths and weaknesses of both the student and the teacher lie for remedial purposes. Thus a major advantage with this form of testing is in that it provides much more frequent information than norm – referenced testing on student progress and help diagnose specific strengths and weaknesses. However they lack a national comparison and both test design and scoring may be less precise than on a norm – referenced tests (Bude & Lewin, 1997).

Educators may choose to use CRTs when they wish to see how well students have learned the knowledge and skills. A mathematics teacher, for instance, may give a test after a unit topic to establish what students have acquired and what they haven't acquired. Teachers know that they have to assess their students to build up a detailed understanding of each student's strength and weaknesses. This information is useful to the teacher in deciding whether to review or repeat or move on with teaching. Without such formative knowledge teachers cannot be in a position to assist.

However the use of criterion referenced tests in schools has its own weaknesses. It slows down the pace of teaching since the teacher is often forced to re-teach until acceptable levels are attained. Since criterion referenced tests are formative and ongoing they should not be overdone. Getting students to become preoccupied with how they are doing can undermine their interest in what they are doing. An excessive concern with performance can erode curiosity. Also, construction of the

test can be taxing for the teacher since it should cater for content validity and reliability. Ranking of performance scores per class which a *CRT* does not provide for is an essential component of the teaching-learning process especially when it comes to school progress reports and other essential documents.

The debate between the two forms of assessment also discerns on two models, the demand model and the support model, as shown in Fig b2

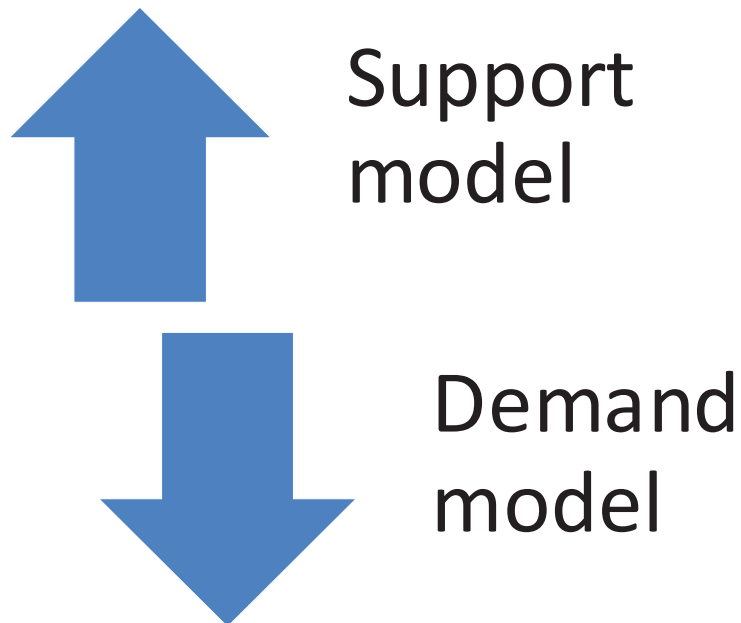


Fig. 2 The Demand model versus the Support model
Adapted from Schaps (1993)

Fig.2 above shows two opposing ideas or ideas that diverge from a central point. In this case the central point is student assessment. Schaps (1993) emphasizes separating between: focusing on what students ought to be able to do, that is, what we demand of them (pull effect), as contrasted with focusing on what we can do to support (push effect) and help them learn. In the demand model students are obligated to work for "it". Blame is leveled by saying students 'chose' not to work hard or earned a certain grade, apparently removing all responsibility from educators. In this mindset teachers report whether or not students did what they were supposed to do.

The support model, which by contrast assists and ensures learners of success, is supportive in nature, guiding and stimulating learners' natural inclination to explore

what is unfamiliar. This approach meshes with the so called learner – centered learning, whereby the point is to help students act on upgrading themselves.

The question of which form of test to use is largely a matter of personal or institutional decision based on set goals. However, the two forms of assessment, norm referenced and criterion referenced tests should be used concurrently in schools. In other words, they are complementary in nature as demonstrated in this paper. Fig. 3 below illustrates this notion.

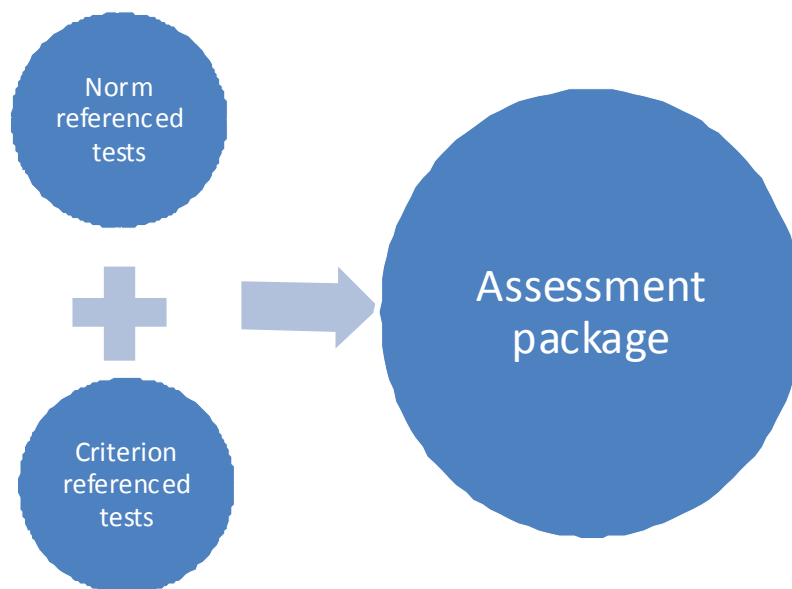


Fig. 3 Norm referenced test versus Criterion referenced test
[Adapted from Mpofu (1991)]

Neither of them is dispensable in the teaching learning process although criterion referenced testing is largely associated with formative evaluation while norm referenced testing is mostly used for summative evaluation. Formative testing refers to testing that is meant to guide teaching on a day to day basis so that the student's developing skills in an area of learning are apparent, whereas summative testing is to determine the pupil's overall achievement after a period of instruction (Mpofu, 1991).

The table below summarizes the debate.

Table 1. Comparison between *NRTs* and *CRTs*

Dimension	<i>CRT</i>	<i>NRT</i>
Purpose	<ul style="list-style-type: none"> • To determine whether each student has achieved specific skills or concepts. • To find out how much students know before instruction begins and after it has finished. • Diagnosis and prescription 	<ul style="list-style-type: none"> • To rank each student with respect to the achievement of others in broad areas of knowledge. • To discriminate between high and low achievers. • Classification and labelling
Content	<ul style="list-style-type: none"> • Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts. • Each skill is expressed as an instructional objective. • Should match specified curriculum goals. 	<ul style="list-style-type: none"> • Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts • Content which maximally discriminate between individuals, even though they may not represent important curriculum goals.
Score Interpretation	<ul style="list-style-type: none"> • Each individual is compared with a preset standard for acceptable achievement. The performance of other examinees is irrelevant. • A student's score is usually expressed as a percentage. • Student achievement is reported for individual skills. 	<ul style="list-style-type: none"> • Each individual is compared with other examinees and assigned a score--usually expressed as a percentile, a grade equivalent score, or a stanine. • Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills.

Adapted from Huitt(1996) & Mpofu (1991)

Conclusion

Is one test better than another? In this paper it was demonstrated that each type of test serves important but different purposes. Schools systems must therefore strive for well balanced education systems through employing both criterion and norm – referenced testing in their proper contexts. This is critical for helping students to become good learners and good people. Ironically, the environment created by an emphasis on grades, standardized testing makes it more difficult to know how well students understand. This paper has compared and contrasted norm referenced and criterion referenced tests. The strengths and limitations of each type of test were outlined and discussed. In a nutshell, norm referenced measures report the achievement of an individual in relation to the fixed standards. As discussed above, each of these two types of tests has a place and time in the evaluation and assessment of education programmes. In the case of norm-referenced tests the whole reference group providing the standard may be “bright or weak” and this way of looking at scores simply allows us to rank pupils. On the other hand when criterion referenced tests are used everyone can “do well or do bad” but there would still be remediation until everyone does well.

References

- Bond, L. A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 5(2).
- Bude, U & Lewin, K.(1997). Improving test design: Assessment of science and Agriculture in primary schools in Africa. Bonn: DSE.
- Graham, J.R & Roy, S. (1984). *Psychological testing*. New Jersey: Prentice Hall.
- Gray, L. R. (1980). *Educational Evaluation and Measurement*. London: Merrill Publishing Company.
- Glaser, R. (1963). *Instructional Technology and the Measurement of Learning Outcomes*. New York: Bell and Howell company .
- Glaser, R. (1970). *Measurement in Learning and Instruction*. Washington DC: American council on Education.
- Huitt, W. (1996). *Measurement and evaluation: Criterion- versus norm-referenced testing*. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. [<http://www.edpsycinteractive.org/topics/measeval/crnmref.html>]
- Kubiszyn, T. & Borich, G. (2003). *Educational testing and Measurement; Classroom Application and Practice*. New York: John Wiley & Sons, INC.

Mississippi State University.(2000). Office of Institutional Research, *Results of the Academic Profile Assessment*, [<http://www.msstate.edu/dept/oir/academicprofile.pdf>]

Kanyonga,G.K.(2005). Zimbabwe's public education system reforms: Successes and challenges. *International Education Journal*, 6(1), 65-74.

Mpofu, E. (1991). *Testing for Teaching*. Harare: Longman

Nunnally, J. C. (1972). *Education Measurement and Evaluation*. New York: McGraw hill.

Siann, G. & Ugwuegbu, D. C. E. (1984). *Education Psychology in a changing World*. London: Gorge Allen and UNWIN.

Zindi, F. Peresuh, M. & Mpofu, E. (1998). *Psychology for the classroom*. Harare: College Press. <http://www.casly.org>