*International Journal of Humanities and Social Science Invention*
*ISSN (Online): 2319 – 7722, ISSN (Print): 2319 – 7714*
*www.ijhssi.org ||Volume 4 Issue 10 || October. 2015 || PP.24-30*

# Norm-Referenced and Criterion-Referenced Test in EFL Classroom

Shafaat Hussain[1], Tessema Tadesse [2], Sumaiya Sajid[3]
[1](*Department of Journalism and Communication, Madawalabu University, Ethiopia)*
[2](*Department of English Language and Literature, Madawalabu University, Ethiopia)*
[3](*Department of English, Falahe Ummat Girls PG College,UP, India)*

**ABSTRACT** *: Language teaching and testing are inseparable and complementary to one another. Beginning from intuitive via scientific testing, today we are in communicative era of testing. Recently, we have plethora of language testing approaches and one of them is norm-referenced versus criterion-referenced testing. Although no testing is flawless and some degree of subjectivity is inevitable norm-referenced and criterion-referenced tests give us different types of information regarding the performance of students in the classroom and outside. Both tests inform us how well students are performing. This paper attempts to clear the dust from these two concepts -- norm-referenced and criterion-referenced tests; their comparison and contrast; the relationship between them, how items are constructed and score is interpreted; and finally their critical appreciation.*

**KEYWORDS -** *criterion-referenced test., EFL, language assessment, language testing approaches, norm-referenced test*

## I. INTRODUCTION

A test is a method of measuring a person's ability, knowledge, or performance in a given domain (Brown 2003; Bachman 1995). Language teaching and testing are intertwined as it helps students create positive attitude, competitive temperament and mastery of language; it helps teachers in raising morale, getting reflections, diagnosing error, identifying thrust areas; enhancing effectiveness, and knowing future course of action (Kubiszyn & Borich 2007, Madsen 1983, Salvia & Ysseldyke 2007). Language testing approaches are axioms or correlative assumptions which provide method (a set of testing style), design (norm and domain) and procedure (technique and administration) to follow (Richards & Rodgers 2001). There are six major dichotomies in the literatures of language testing and they are – formative versus summative, direct versus indirect, discrete versus integrative, objective versus subjective, traditional versus alternative and norm-referenced versus criterion-referenced.

***Formative assessment*** is testing that is part of the developmental or ongoing teaching/learning process. It should include delivery of feedback to the student. It is an evaluation of student learning that aids understanding and development of knowledge, skills and abilities without passing any final judgment (via recorded grade) on the level of learning. *Summative assessment* is the process of evaluating (and grading) the learning of students at a point in time. It is testing which often occurs at the end of a term or course, used primarily to provide information about how much the student has learned and how well the course was taught (Brown 2003; Hughes 2003; Wojtczak 2002). Testing is said to be ***direct*** when it requires the candidate to perform precisely the skill that is to be measured whereas ***indirect testing*** attempts to measure the abilities which underlie the skill in which we are interested (Henning 1987; Hughes 2003; Wojtczak 2002). ***Discrete point testing*** refers to the testing of one element at a time, item by item. ***Integrative testing***, by contrast, requires the candidate to combine many language elements in the completion of the task (Brown 2003; Henning 1987; Hughes 2003; Wojtczak 2002). A test which is scored by comparing the response of the student with an established set of correct response is known as ***objective test***. In contrast, a ***subjective test*** is scored by opinion and personal judgment (Bachman 1995; Hughes 2003). ***Traditional test*** ask students what they can recall and produced while ***alternate test*** ask students what they can do with the language, how they integrate and produce it (Brown 2003). ***High-stakes tests*** are those which impact large number of people or program whereas ***low-stakes tests*** have relatively minor impact on people or program (Wojtczak 2002).

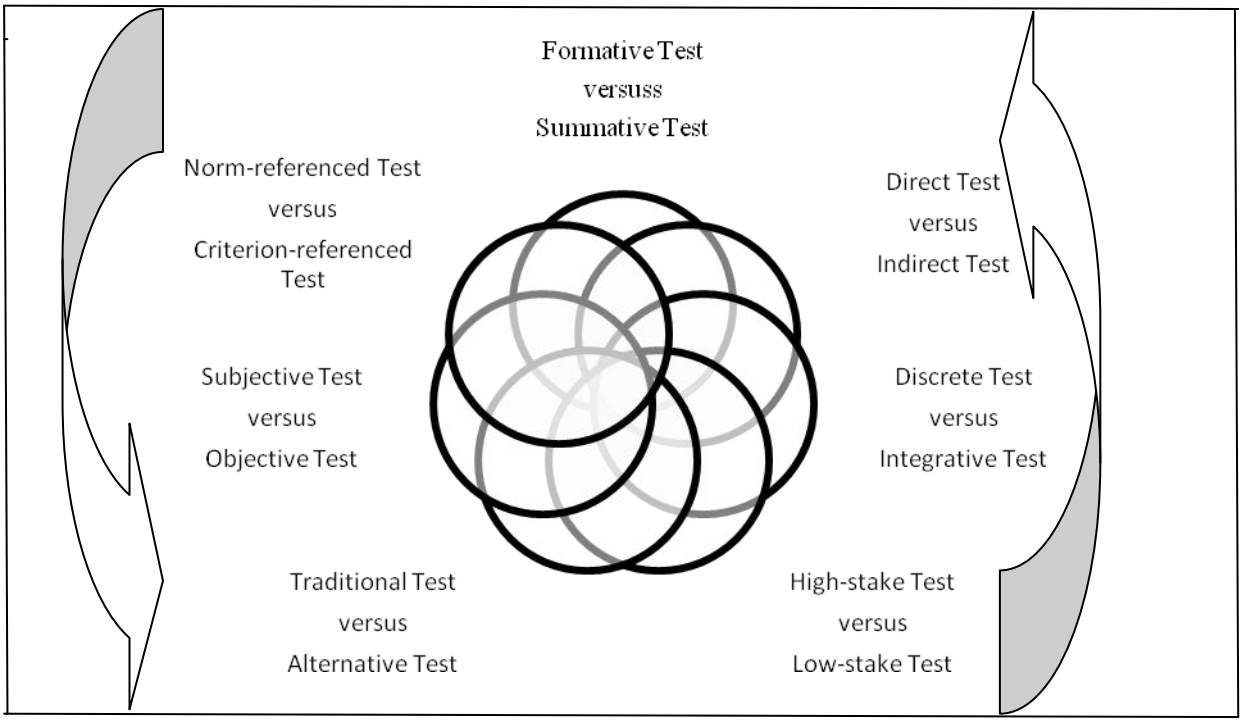


Figure 1: Approaches to language testing

Norm-referenced test and criterion-referenced test are the language testing approaches that provide information about the knowledge and skills of the students tested. ***Norm-referenced test*** is the process of evaluating (and grading) the learning of students by judging (and ranking) them against the performance of their peers. ***Criteria-referenced test*** is the process of evaluating (and grading) the learning of students against a set of pre-specified criteria (Brown 2003; Hughes 2003; Huitt 1996; Wojtczak 2002).

There is another way to look at different kinds of language tests which is summarized in figure 2 below:

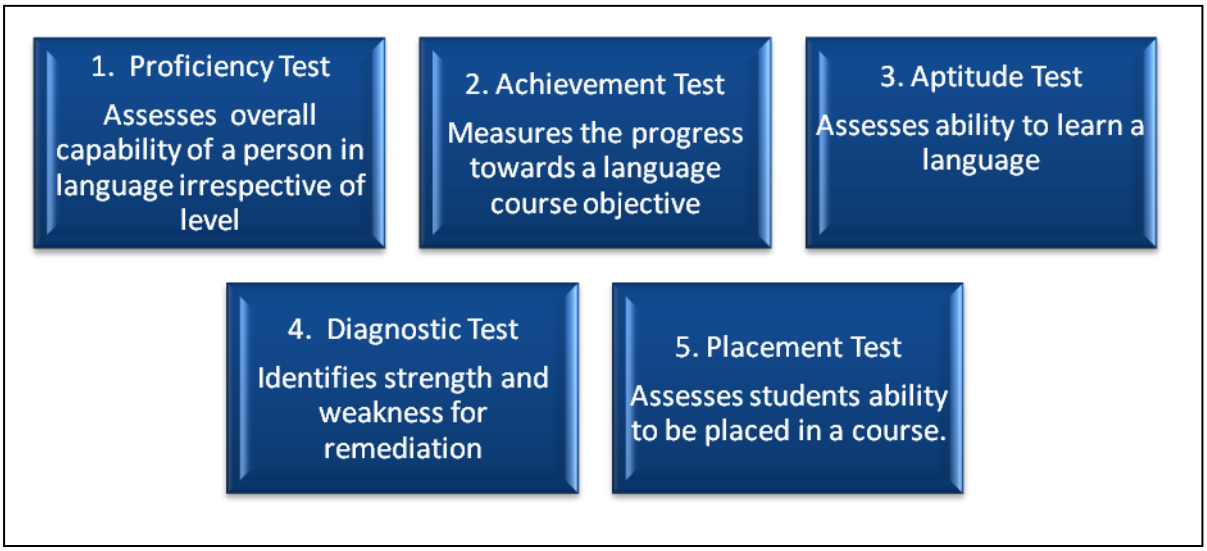


Figure 2: Kinds of tests adapted from Bachman (1995) & Hughes (2003)

## II. CLEARING THE BOUNDARY

It was Robert Glaser, an American educational psychologist who termed ***Norm-referenced Test*** and ***Criterion-referenced Test*** for the first time in 1963 (Montgomery & Connolly 1987).

***2.1 Norm-referenced Test (NRT)***

In testing, the scores or a performance of a particular group (the "norm" group) as measured in some way is known as 'norm'. Norms may be used to compare the performance of an individual or group with the norm group. Norms may be expressed by reference to such factors as age, grade, region and special need group on a test (Brown 1976; Noll, Scannell & Craig, 1979). NRT is a test that measures how the performance of a particular test taker or group of test takers compares with the performance of another test taker or group of test takers whose scores are given as the norm. Norm-referenced standardized tests can use local, state, or national norms as a base. A test taker's score is, therefore, interpreted with reference to the scores of other test takers or groups of test takers, rather than to an agreed criterion (Richard & Schmidt 2002). Hence, NRT is an approach of evaluation through which a learner's individual relative rank is compared to other students in the classroom (Brown 1976; Mrunalini 2013; Salvia & Ysseldik 2007). For example, if a student receives a percentile rank score of 34, this means that he or she performed better than 34% of the students in the norm group (Bond 1996). Hence, if we conclude the test performance that a particular student achieved in the classroom as 'better than 34 percent of the other students', it is an approach of evaluating through NRT. NRT tells us where a student stands compared to other students in her performance and it is only useful to take certain types of decisions (Bachman 1995; Kubiszyn & Borich 2007). The examples of NRTs include IQ tests, developmental-screening tests (used to identify learning disabilities in young children or determine eligibility for special educational services), cognitive ability tests, readiness tests etc. SAT (Stanford Achievement Test), CAT (California Achievement Test), MAT (Metropolitan Achievement Test), TOEFL, IELTS etc. are best practices of NRT. To be simpler, theater auditions, course placement, program eligibility, or school admissions and job interviews are NRTs because their goal is to identify the best candidate compared to the other candidates, not to determine how many of the candidates meet a fixed list of standards. Educators use NRTs to evaluate the effectiveness of teaching programs, to help determine students' preparedness for programs, and to determine diagnosis of disabilities for eligibility.

***2.2 Criterion-referenced Test (CRT)***

The word "criterion" in CRT has been referred in two ways in the literature. One, criterion refers to the material being taught in the course. Here, CRT would assess the particular learning points of a particular course or program. Second, criterion is the standard of performance or cut-point for decision making that is expected for passing the test/course. Here, CRT would be used to assess whether students pass or fail at a certain criterion level or cut-point (Bond 1996). CRT is a test that measures a test taker's performance according to a particular standard or criterion that has been agreed upon. The test taker must reach this level of performance to pass the test, and a test taker's score is interpreted with reference to the criterion score, rather than to the scores of other test takers (Richard & Schmidt 2002). Hence, CRT is an approach of evaluation through which a learner's performance is measured with respect to the same criterion in the classroom (Brown 1976; Mrunalini 2013; Salvia & Ysseldik 2007). For instance, if we conclude the test performance that a particular student achieved in the classroom as '90 percent', it is an approach of evaluating through CRT. The popular way to show CRT is percentage (Mrunalini 2013; Salvia & Ysseldik 2007). CRT tells us about a student's level of proficiency in or mastery over a set of skills and help us decide whether a student needs more or less work over a set of skills saying nothing about the student's place compared to other students (Bachman 1995; Kubiszyn & Borich 2007). For instance, if a test is designed to evaluate how well students demonstrate mastery of the specified content (e.g. types of tense) it is CRT. Most everyday tests, quizzes and final exams conducted in the classroom teaching can be taken as CRT. A 'Basic Writing' CRT would include questions based on what was supposed to be taught in writing classes. It would not include 'speaking' or 'advanced writing' questions. Students who took 'Basic Writing' course could pass this test if they were taught well and they studied enough and the test was well-prepared.

***2.3 Criterion Related Validity***

Criterion-related validity refers to the correspondence between the results of the test in question and the results obtained from an outside criterion. The outside criterion is usually a measurement device for which the validity is already established. Criterion-related validity is determined by correlating the scores on a newly developed test with scores on an already-established test. When a newly developed language proficiency test (ROSHD) is tested against an established language proficiency test (TOEFL) and both the scores are correlated, the degree of correlation is the validity index of the 'ROSHD' test validated against 'TOEFL.' It means that to the extent that the two tests correlate, they provide the same information on examinees' language proficiency (Brown 2005; Farhady 1986).

## III. TEST ITEMS IN NRT AND CRT

Constructing items do not differ much from one another in NRT and CRT. Mostly NRTs are multiple choice and rarely open and short answer questions. The items are constructed with moderate difficulty, with some items easier and some more difficult. On the other hand, CRT items are curriculum based which includes the local textbook. The items are constructed on objectives in depth without considering the difficulty level. NRT items are generally built with items that range in difficulty from very easy to very hard. The important characteristic for items is their ability to discriminate among students. NRTs frequently have items that cover a broad range of content but only have a few items that measure each content area (Cronbach 1970; Kubiszyn & Borich 2007) CRT items generally range from moderately difficult to easy. Overall, these tests tend to be easier than an NRT. CRT items are more likely to measure a very limited range of content but will include many items for each content area. As a result, they provide more detailed information about what the student knows. Let us take an example given below to measure a student's knowledge about the Gulf War, 2003:

Table 1: Test Items of NRT and CRT Adapted from Kubiszyn & Borich (2007)

| Item 1 | Item 2 |
|---|---|
| A. During the 2003 Gulf war, more Iraqi tanks were destroyed by this aircraft than by all other aircraft combined.<br>A. F14 Tomcat<br>B. F16 Hornet<br>C. A6 Intruder<br>D. A10 Thunderbolt | B. During the 2003 Gulf war against Iraq which of the following were employed against American soldiers?<br>A. Biological weapons<br>B. Nuclear weapons<br>C. Roadside bombs<br>D. Naval bombardment |

For a high school student, due to the specificity and difficulty level question A is an NRT item since it has more subtle distinction among the fact. But question B is general and it may be answered easily by high school history students, therefore, it is a CRT item. For a strategic science class, question A may be considered in CRT item because they are expected to be familiar with the specificities and minute details of weaponry used during the Gulf War (Kubiszyn & Borich 2007). Hence, it goes without saying that test items are not norm-referenced and criterion-referenced by nature. They need to be considered in terms of the test takers for whom they are prepared. The same test can be used in both ways (Cronbach 1970) but mixing the two without considering the test takers is invalid.

## IV. SCORING METHODS IN NRT AND CRT

On an NRT, the score reflects how many *more* or *fewer* correct answers a student gives in comparison to other students. In NRT, scores are generally reported in percentile ranks, linear standard score, normalized standard score and developmental scales (Linn & Gronlund 2000). On the other hand, CRT results are often based on the number of correct answers provided by students, and scores might be expressed as a percentage, checklists, rating scales, grades, and rubrics. The figure 3 clearly shows how students' performances are reported:

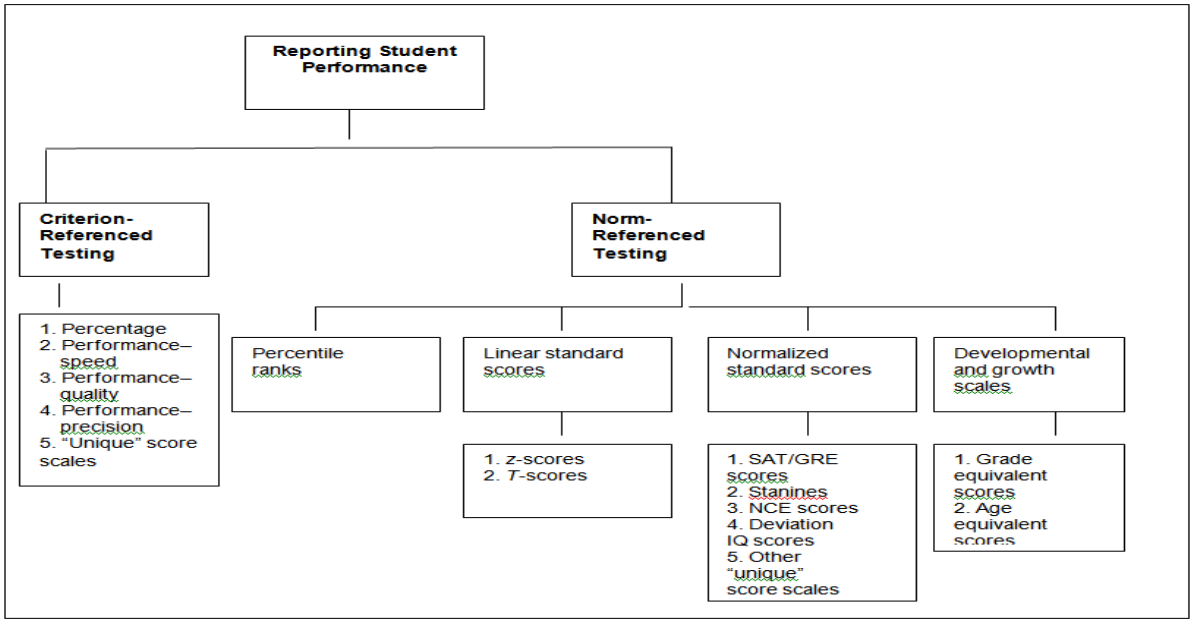


Figure 3: Scoring method adapted from Salvia & Ysseldik (2007)

## V. STRENGTHS OF NRT AND CRT

The proponents of **NRT** have proposed many representative arguments on its strengths. For instance, NRT is simple to administer and easy to score (Kubiszyn & Borich 2007; Sanders & Horn 1995). It is a good tool to measures variety of skills and ability at once (Bond 1996; Linn 2000; Swanson & Watson 1982). It helps identify those who may require specialized assistance for example mental retardation, learning disabilities (autism, dyslexia), attention-deficit/hyperactivity syndrome (ADHD), and conduct disorder (Stiggins 1994). This test discourages biasness and favoritism from among students. The quality of NRT is usually high because they are developed by testing experts, piloted, and revised before they are used with students, and they are dependable and stable for what they are designed to measure. NRT is also good for ranking and sorting students for administrative purposes (Anastasi 1988). It is intended to judge the school performance and schools accountability of providing learning standards and maintaining quality of education. The test is also used to determine a young child's readiness for preschool or kindergarten. These tests may be designed to measure oral-language ability, visual-motor skills, and cognitive and social development. NRTs are administered for admission decisions at entry level and promotional decision at exit level. At policy level, NRTs are very useful because based on NRT data, programs are selected and evaluated; remedial and gifted strategies are developed; funding is appropriated for teachers' professional development; and textbooks are prepared.

On the contrast, the representative arguments on strengths typically made by proponents of **CRT** are also bulky. CRT is good to measure specific skills, objectives or domain (Bond 1996; Kubiszyn & Borich 2007; Linn 2000; Sanders & Horn 1995; Swanson & Watson 1982). It gives direction to learning how well students are learning. It is good to determine learning progress if students have learning gaps or academic deficits that need to be addressed (Bond 1996). CRT gives direction to teaching and re-teaching. Instructors can use the test results to determine how well they are teaching the curriculum and where they are lagging behind (Bond 1996). CRT helps measure the academic achievement of students usually for the purposes of comparing academic performance among schools, districts and states. The results provide a basis for determining how much is being learned by students and how well the educational system is producing desired results (Cohen, Manion & Morrison 2004). Through open ended questions this test promotes higher-level cognitive skills such as critical thinking, problem solving, reasoning, analysis, or interpretation (Sanders & Horn 1995). CRT can give parents more information about what exactly their children have learned, what competencies still need to be mastered and which ones have been already mastered (Bond, 1996). It is well suited for training programs to assess learning of trainees. It is used to determine that a person is qualified to receive a certificate or not (Swanson & Watson 1982).

## VI. DRAWBACKS OF NRT AND CRT

Although the NRT and CRT have many advantages, they do have some drawbacks. The kinds of arguments typically made by the critics of **NRT** are eyebrow rising. Although testing experts and test developers warn that major educational decisions should not be made on the basis of a single test score, NRT scores are often misused in schools when making critical educational decisions, such as grade promotion or retention, which can have potentially harmful consequences for some students (Huitt 1996). Multiple-choice tests—the dominant NRT format, promote rote learning and memorization in schools over more sophisticated cognitive skills, such as argumentation, conceptualization, decision making, writing, critical reading, analytical thinking, problem solving and creativity (Corbett & Wilson, 1991). Overreliance on NRT results can lead to inadvertent discrimination against minority groups and low-income student populations, both of which tend to face more educational obstacles than non-minority students and higher-income households (Bond 1996). The test is considered biased by the experts because those questions are eliminated which low scorers might get right (Huitt 1996). The results of NRT are shown in score band and not on true score; therefore, it has measurement error. NRTs encourage teachers to view students in terms of a bell curve, which can lead them to lower academic expectations for certain groups of students, particularly special-needs students, English-language learners, or minority groups. And when academic expectations are consistently lowered year after year, students in these groups may never catch up to their peers, creating a self-fulfilling prophecy (The Glossary of education …2014). NRT has to be finished in a time limit which may favor or disfavor an individual student.

Although the **CRT** has many advantages, it does have some bottlenecks. The kinds of arguments typically made by the critics of NRT are worth considering. CRT does not allow for comparing the performance of students in a particular location with national norms. For example, a school would be unable to compare 5th grade achievement levels in a district, and therefore be unable to measure how a school is performing against other schools (Huitt 1996). If not institutional, It costs a lot of money, time and effort. Creating a specific curriculum takes time and money to hire more staffs; and most likely the staff should be professionally competent (The Glossary of education …2014). CRT needs efficient leadership and collaboration, and lack of these can cause problems. For instance, if a school is creating assessments for special education students with no well-trained professionals, they might not be able to create assessments that are learner-centered (Bond 1996). It is difficult

for curriculum developers to know what is working and what is not working because tests tend to be different from one school to another. It would require years of collecting data to know what is lacking and what is not. It may slow the process of curriculum change if tests are constantly changed (Corbett & Wilson, 1991). The process of determining proficiency levels and passing scores on CRT can be highly subjective or misleading—and the potential consequences can be significant, particularly if the tests are used to make high-stakes decisions about students, teachers, and schools. Because reported "proficiency" rises and falls in direct relation to the standards or cut-off scores used to make a proficiency determination, it's possible to manipulate the perception and interpretation of test results by elevating or lowering either standards and passing scores. And when educators are evaluated based on test scores, their job security may rest on potentially misleading or flawed results. Even the reputations of national education systems can be negatively affected when a large percentage of students fail to achieve "proficiency" on international assessments (Huitt 1996). The subjective nature of proficiency levels allows the CRT to be exploited for political purposes to make it appear that schools are either doing better or worse than they actually are. For example, some states have been accused of lowering proficiency standards of standardized tests to increase the number of students achieving "proficiency," and thereby avoid the consequences—negative press, public criticism, large numbers of students being held back or denied diploma that may result from large numbers of students failing to achieve expected or required proficiency levels (The Glossary of education …2014).

## VII. COMPARING NRT AND CRT

In NRT, difficulty of items vary from those that no one answers correctly to those that everyone answers correctly whereas in CRT, the difficulty of items are equivalent to each other (Bond 1996; Kubiszyn & Borich 2007; Linn 2000; Sanders & Horn 1995). NRT covers many objectives at a time while CRT covers a few objectives to be achieved that is instructed (Bond 1996; Kubiszyn & Borich 2007; Linn 2000; Montgomery & Connolly 1987; Sanders & Horn 1995). In a NRT 'distracters' are constructed whereas in CRT item responses are 'relevant' among each other (Bond 1996; Kubiszyn & Borich 2007; Linn 2000; Sanders & Horn 1995). The purposes of NRT are screening, diagnosis, classification and placement whereas the purposes of CRT are skill or objective achievement (Bond 1996; Linn 2000; Montgomery & Connolly 1987; Sanders & Horn 1995). In NRT, item construction usually does not develop from task analysis; test items may or may not be related to the objectives of instruction (intervention). In CRT, items developed from task analysis; test items are related to the objectives of instruction (Bond 1996; Linn 2000; Montgomery & Connolly 1987; Sanders & Horn 1995). Scoring of NRT is based on standards relative to a group; variability of scores (ie, means and standard deviations) which is desired with normal distribution. Scoring of CRT is based on absolute standards; variability of scores is not obtained because perfect or near-perfect scores are desired (Bond 1996; Linn 2000; Montgomery & Connolly 1987; Sanders & Horn 1995). In a NRT percentile rank is used for relative ranking whereas in a CRT percent is used for performance (Bond 1996; Kubiszyn & Borich 2007; Linn 2000; Montgomery & Connolly 1987; Sanders & Horn 1995). NRTs are breadth but not depth in content specification whereas CRTs are depth but not breadth in content specification (Bond 1996; Linn 2000; Sanders & Horn 1995).

Table 2: Difference in NRT and CRT adapted from Brown (1992)

| Differences | CRT | NRT |
|---|---|---|
| **Test Characteristics** | | |
| Underlying Purposes | Foster learning | Classify/group students |
| Types of Decisions | Diagnosis/progress/achievement | Aptitude/proficiency/placement |
| Levels of Generality | Classroom Specific | Overall global |
| Students Expectations | Know content | Don't know content |
| Score Interpretations | Percent | Percentile |
| Score Report Strategy | Score with Answer | Score without Answer |
| **Logistic Dimensions** | | |
| Group size | Small | Large |
| Range of Abilities | Homogenous | Heterogeneous |
| Test Length | Smaller | Larger |
| Time Allocated | Shorter | Longer |
| Cost Involved | Teacher constructed | Fee involved |

## VIII. CONCLUSION

This term paper dealt with the types and approaches of testing; the concept of NRT and CRT; their item construction and scoring method, their strengths and weaknesses, their comparative analysis and how they differ from each other. Which of these methods is preferable? It is a question of concern of this paper. In literature, a mix voice comes from different parts of the world and there is no prescriptive point of view available. For some scholars, students' grades in language tests should be decided based on a mix of both NRT and CRT scores because they are interdependent and complementary to each other. Some scholars emphasize that there should be a striking balance between the two and this balance should be strongly oriented towards CRT as the primary and dominant principle. The third group of scholars retains that both NRT and CRT are poles apart in an education system because the former serves the function of benchmarking, vision and policy decisions and the latter gives an improvement in classroom instruction. Hence, NRT would be misleading in the classroom evaluations and CRT would be misleading outside the classroom evaluations. Both have their own importance at their respective places. Thus, debating which one has upper hand is useless exercise because their purpose is different and they are complementary to each other in an education system. Before a state can choose what type of NRT to use, the state education officials will have to consider if that test meets three standards -- whether the assessment strategies of a particular test matches the state's educational goals; addresses the content the state wishes to assess; and allows the kinds of interpretations state education officials wish to make about student performance. Likewise, before a teacher chooses what type of CRT to use, they have to consider whether the test is effective in taking different academic decisions regarding language attainment or not.

## REFERENCES


[1] Alderson, JC, Clapham, D & Wall, D. 1995. *Language Test Construction and Evaluation.* New York: Cambridge University Press

[2] Anastasi, A. 1988. *Psychological Testing.* New York: MacMillan.

[3] Bachman, LF. 1995. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

[4] Bond, L. 1996. *Norm and Criterion-referenced Testing: Practical Assessment: Research & Evaluation* (O). http://www.ericae.net/pare/getvn.asp?v=5&n=2 (Accessed 10 November 2014).

[5] Brown, DH. 2003. *Language Assessment: Principles and Classroom Practices*. San Francisco: Longman.

[6] Brown, F. 1976. *Principles of Educational and Psychological Testing.* 2nd edition. New York: Holt, Rinehart and Winston.

[7] Cohen, L, Manion, L & Morrison, K. 2004. *A Guide to Teaching Practice.* London: Routledge Falmer.

[8] Corbett, HD & Wilson, BL. 1991. *Testing, Reform and Rebellion*. New Jersey: Ablex Publishing.

[9] Cronbach, LJ. 1970. *Essentials of Psychological Testing*. 3rd edition. New York: Harper & Row.

[10] Farhady, H. 1986. Roshd. *Roshd Foreign Language Teaching Journal.* 12(2): 12-13.

[11] Henning, G. 1987. A Guide to Language Testing, Development, Evaluation and Research. Boston: Heinle & Heinle.

[12] Hughes, A. 2003. *Testing for Language Teachers.* 2nd edition. Cambridge: Cambridge University Press.

[13] Huitt,W.1996. *Criterion versus Norm-referenced Testing.* Valdosta: Valdosta State University. http://chiron.valdosta.edu/whuitt/col/measeval/crnmref.html (Accessed 10 November 2014).

[14] Kubszyn, T & Borich, G. 2007. *Educational Testing and Measurement: Classroom Application and Practice.* 8th edition. New Jersey: Wiley.

[15] Linn, R. 2000. Assessments and Accountability. *ER Online 29*(2): 4-14.

[16] Linn, RL & Gronlund, NE. 2000. *Measurement and Assessment in Teaching.* 8th edition. New Jersey: Prentice Hall.

[17] Madsen, HS.1983. *Techniques in Testing.* Oxford: Oxford University Press.

[18] Montgomery, PC & Connolly, BH. 1987. Norm-Referenced and Criterion-Referenced Test: Use in Pediatrics and Application to Task Analysis of Motor Skill. *Physical Therapy* 67(12): 1873-1876.

[19] Mrunalini, T. 2013. *Educational Evaluation*. 5th edition. New Delhi: Neelkamal Publications.

[20] Noll, V, Scannell, D & Craig, R. 1979. *Introduction to Educational Measurement*. 4th edition. Boston: Houghton Mifflin.

[21] Richards, JC & Rodgers, TS. 2001. *Approaches and Methods in Language Teaching*. New York:Cambridge University Press.

[22] Richard, JC & Schmidt, R. 2002. *Longman Dictionary of Language Teaching and Applied Linguistics.* London: Pearson Education.

[23] Salvia, J & Ysseldyke, JE. 2007. *Assessment*. 10th edition. Boston: Houghton Mifflin.

[24] Sanders, W & Horn, S. 1995. Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for Assessment of Educational Outcomes. *Education Policy Analysis Archives 3*(6): 14-23.

[25] Stiggins, RJ. 1994. *Student-centered Classroom Assessment.* New York: Merrill.

[26] Swanson, HL & Watson, BL. 1982. *Educational and Psychological Assessment of Exceptional Children: Theories, Strategies, and Applications*. St Louis: Mosby Company.

[27] The Glossary of Educational Reform by Great Schools Partnership.2014. http://www.creativecommons.org/licenced/by/-n-c-sa/4.0/ (Accessed 10 November 2014).

[28] Wojtczak, A. 2002. *Glossary of Medical Education Terms*. http://www.iime.org/glossary.htm, (Accessed 10 November 2014).