COMPARING CRITERION AND NORM REFERENCED ASSESSMENTS OF LANGAUGE SKILLS IN THE SECOND LANGUAGE

Iroda I. Mirmakhmudova*

*Department of Intercultural Communication and Tourism,
Tashkent State University of Uzbek Language and Literature,
Tashkent, UZBEKISTAN
Email id: i_rosh@rambler.ru

ABSTRACT

This article addresses different approaches towards assessing students' language skills. The increased interest in new assessment is based on an issue: traditional assessment does not provide full description of students' outcomes which is important for the teachers to monitor learners' progress and to plan for instructions. The test score mainly shows that a student has succeeded or failed, but it gives the teacher an incomplete picture of student needs and strengths. The concept of criterion-referenced assessment is to assess language as communicative competence. This article gives clear idea about these assessments and how useful it is to use criterion-referenced assessment.

KEYWORDS: *Assessment, Norm Referenced, Criterion Referenced, Feedback, Scoring Process, Rating Scales, Criteria, Validity*

INTRODUCTION

The assessment of students' language ability is crucial for both teachers and students. It makes teachers responsible to create accurate and reliable assessment criteria on the basis of which decisions and inferences about students should be made. The increased interest in new assessment is based on an issue: traditional assessment does not provide full description of students' outcomes which is important for the teachers to monitor learners' progress and to plan for instructions. The test score mainly shows that a student has succeeded or failed, but it gives the teacher an incomplete picture of student needs and strengths. Over the past years the need to achieve more reliable and more transparent test results caused the development of criterion-referenced assessment. Hudson (2005) notes that the development relates to issues surrounding characteristics of proficiency or ability scales and how these scales are conceptualized in criterion-referenced performance assessment. In most assessment projects criterion-referenced assessment has been accepted as alternative to norm-referenced, traditional assessment to grade and report students' achievements [1].

The concept of these assessment projects is to assess language as communicative competence. Language takes place in a social context as a social act, and this frequently needs to be recognized in language assessment [4]. Indeed, I want to add that there are many current test projects that are criterion-referenced in their constructions: the Canadian Language Benchmarks, the Common European Framework of References for Languages (Council of Europe, 2001), and Education Week (2003) which reports most US states use criterion-referenced assessment in their English/language arts assessments. Several reasons make it interesting to apply the Canadian Language Benchmarks project into practice. First, the format of 12 benchmarks in four language skill areas is easily understandable by both teachers and students. Second, the detailed descriptions of each benchmark help the teachers to trace individual progress and let students get full idea about their strengths and weaknesses. We know sometimes a teacher cannot provide one-to-one

feedback, the reason may be different. In this case, I think students can use 12 benchmarks as feedback on his/her grade. Third, it can be adapted into any language program or curriculum. Fourth, in my opinion it is also important, very helpful, and useful for young teachers.

At the beginning of teaching experience pre-service teachers often find it difficult to assess students' language ability and to write feedback on students' achievements in different skills, especially in writing which. Language assessment needs some experience and knowledge. But established framework would assist inexperienced teachers in tracking the development of students' learning process. Finally, it provides teaching staff with a common discourse to discuss student growth and ultimately to have a positive washback on a program or curriculum as a whole. Adopting the Canadian Language Benchmarks would "establish a frame of reference that can describe achievement in a complex system in terms meaningful to all the different partners in or users of that system". [3]

Criterion-referenced assessment and Norm-referenced assessment

Criterion-referenced and norm-referenced assessments are quite different methods of assessment according to their purposes, the way in which content is selected, and the scoring process. Testing intended to get information about students' general language ability are called norm-referenced assessment and it is widely used. Stiggins (1994) defines the usage of norm-referenced assessment as highlighting achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers. Norm-referenced assessment is appropriate for getting information about a test-takers or student's knowledge in order to compare the result with others. The tests of English as a foreign language, known as the TOEFL and IELTS, and all language proficiency tests are good examples. Scores on norm-referenced assessment focus on how a test-taker or an individual has scored in relation to the scores of other persons. So the interpretation given on a test-taker's or student's score is called a relative decision.[2] For instance, if a student scored 87, we can say he/she showed better result than those with 86 or lower and he/she was worse than those with score 88 or above. The major reason for using norm-referenced assessment is to produce a rank order [6], it is very useful in selecting relatively high and low achievers among students. However, an obvious disadvantage of norm-referenced assessment is that it gives little information about what a test-taker actually knows or can do and cannot measure students' progress or learning outcomes and determine the effectiveness of certain curriculum. Only by comparing with other result we can see how well a learner has succeeded or failed. In contrary, criterion-referenced assessment determines "what a student can actually do in the language". [6, p18] It shows the progress of students and curriculum.

Criterion-referenced assessment works along with goals of curriculum or language program and gives detailed information about how well a student has performed on each of educational goals. Accordingly, choosing criterion-referenced assessment requires teachers, educators as well as administrators and curricula developers to specify what they are trying to teach or what students might be taught. Usually, in criterion-referenced assessment concrete criteria are established and a test taker or student is challenged to meet them. The interpretation of criterion-referenced assessment scores is called an absolute decision, as each test-taker or student' score is meaningful without references to the scores of the others. [2] In contrast to norm-referenced assessment, in criterion-referenced assessment all test-takers can pass or get 100 score, if they do well in exams. In recent years, criterion-referenced assessment is widely adopted by different projects as it gives more descriptive and more transparent results. As a teacher I can say, in language classes, assessment is aimed to find out how much material a learner has learned or whether a learner can cope with a certain task in real language situations successfully. For example, does a learner have enough command of English to pass entrance exams of universities or to do well in a job? Can a learner speak fluently or make him/her understood where only English is spoken? We do not

Asian Journal of Research in Social Sciences and Humanities
ISSN: 2249-7315   Vol. 11, Issue 11, November 2021   SJIF 2021 = 8.037
A peer reviewed journal

evaluate the language ability of the learners to compare the results with others, as in the case of norm-referenced assessment. As mentioned above, criterion-referenced assessment is very useful for the teachers and curriculum developers, because it can be used to diagnose the weaknesses and strengths of a particular course, to improve the materials, instructions, and teaching **[5].**

The differences and similarities between Criterion-referenced assessment and Norm-referenced assessment

While doing literature review I read several authors discussing the differences and similarities between criterion-referenced assessment and norm-referenced assessment. For the first time Hudson and Lynch **[2]** refer the differences between criterion-referenced and norm-referenced assessment. Here I have given several tables discussing the differences and similarities between criterion-referenced assessment and norm-referenced assessment. I found Brown's one more relevant as he provides differences with more details in terms of six characteristics. **[1]** Criterion-referenced assessment and Norm-referenced assessment contrast in:

1. The way that scores are interpreted

2. The kinds of things that they are used to measure

3. The purposes for testing

4. The ways that scores are distributed

5. The structures of tests

6. The students' knowledge of test questions content

TABLE 1. DIFFERENCES BETWEEN CRITERION-REFERENCED ASSESSMENT AND NORM-REFERENCED ASSESSMENT THE FOLLOWING IS ADAPTED FROM BROWN J.D. (1996)

Characteristics	Norm-referenced	Criterion referenced
1.Type of interpretation	Relative (a student's performance is compared to that of all other students)	Absolute (a student's performance is compared only to the amount, or percent, of material known)
2.Type of measurement	To measure general language abilities or proficiencies	To measure a specific domain or objective-based language points
3.Purpose of testing	Spread students out along a continuum of general abilities or proficiencies	To assess the amount of material known, or learned, by each
4. distribution of scores	Normal distribution of scores around a mean	Varies, often non-normal, students who know all of the material should all score 100%
5. test structure	A few relatively long sub-tests with heterogeneous item content	A series of short, well-defined sub-tests with homogeneous items content
6.knowledge of questions	Students have little or no idea what content to expect in questions	Students know exactly what content to expect in text questions

Asian Journal of Research in Social Sciences and Humanities
ISSN: 2249-7315   Vol. 11, Issue 11, November 2021   SJIF 2021 = 8.037
A peer reviewed journal

TABLE 2. DIFFERENCES BETWEEN CRITERION-REFERENCED ASSESSMENT AND NORM-REFERENCED ASSESSMENT THE FOLLOWING IS ADAPTED FROM POPHAM, J. W. (1975). EDUCATIONAL EVALUATION. ENGLEWOOD CLIFFS, NEW JERSEY: PRENTICE-HALL, INC

Dimension	Criterion-Referenced Tests	Norm-Referenced Tests
Purpose	• 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.	• To rank each student with respect to the achievement of others in broad areas of knowledge. • To discriminate between high and low achievers.
Content	• 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.	• Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts.
Item Characteristics	• Each skill is tested by at least four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing. • The items which test any given skill are parallel in difficulty.	• Each skill is usually tested by less than four items. • Items vary in difficulty. • Items are selected that discriminate between high and low achievers.
Score Interpretation	• 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.	• 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.

TABLE 3. COMPARISON OF CRITERION-REFERENCED ASSESSMENT AND NORM-REFERENCED ASSESSMENT THE FOLLOWING IS ADAPTED FROM GRONLUND (1988)

Common Characteristics of NRT and CRT:

1. Both require specification of the achievement domain to be measured.

2. Both require a relevant and representative sample of test items.

3. Both use the same types of test items.

4. Both use the same rules for item writing (except for item difficulty).

5. Both are judged by the same qualities of goodness (validity and reliability).

6. Both are useful in educational measurements

Asian Journal of Research in Social Sciences and Humanities
ISSN: 2249-7315   Vol. 11, Issue 11, November 2021   SJIF 2021 = 8.037
A peer reviewed journal

Differences between CRT and NRT (but it is only a matter of emphasis):

1. NRT – Typically covers a *large* domain of learning tasks, with just a few items measuring each specific task.

   CRT – Typically focuses on a *delimited* domain of learning tasks with a relatively large number of items measuring each specific task

2. NRT – Emphasizes *discrimination* among individuals in term of relative level of learning

   CRT – Emphasizes *description* of what learning tasks individuals can do and cannot perform

3. NRT – Favors items of average difficulty and typically omits easy items.

   CRT – Matches item difficulty to learning tasks, without altering item difficulty or omitting easy items.

4. NRT – Used primarily (but not exclusively) for *survey* testing.

   CRT – Used primarily (but not exclusively) for *mastery* testing.

5. NRT – Interpretation requires a clearly defined group.

   CRT – Interpretation requires a clearly defined and delimited achievement domain.

FINDINGS AND DISCUSSION

Why was Criterion-referenced assessment developed?

According to Brown and Hudson (2002) criterion-referenced assessment has been developed in response to problems or weaknesses of norm-referenced assessment and they are as follows:

1. Teaching/testing mismatches

2. Lack of instructional sensitivity

3. Lack of curricular relevance

4. Restriction to the normal distribution

5. Restriction to the items that discriminates

Teaching/testing mismatches: Most cases norm-referenced assessment based examinations use very broad materials to check students' knowledge. As example I would describe course examinations at most schools in Uzbekistan. From the beginning of English language course the students study the books which encompass different skills, but at the entrance exams to University they have to take only grammar oriented examinations. **[6].**

Lack of instructional sensitivity: Because of general and abstractive nature and putative global applicability across a variety of instructional settings, norm-referenced assessment is not suitable for measuring the amount of knowledge or skill developed in a particular course or program. **[2]** Internationally known proficiency test TOEFL is not quite convenient to use for particular program examinations. **[7].**

Lack of curricular relevance: Because of teaching/testing mismatches and lack of instructional sensitivity norm-referenced assessment is not believed to be effective for assessing the weaknesses and strengths of a particular program or comparing the relative weaknesses and strengths of different language programs. **[2]**

Restriction to the normal distribution: Norm-referenced assessment is designed to create a normal distribution of scores. But Brown and Hudson (2002) disputes that restriction to the normal distribution occurs in norm-referenced assessment. For example, when a group of students has just

Asian Journal of Research in Social Sciences and Humanities
ISSN: 2249-7315   Vol. 11, Issue 11, November 2021   SJIF 2021 = 8.037
A peer reviewed journal

completed a 900-hour intensive language program, a normal distribution may not be desirable. The assessment results should show that learning materials happened mostly equally. **[8].**

Restriction to the items that discriminates: If we go back to the situation discussed above with a group of students who have just completed a 900-hour intensive language program, selecting items that discriminate well between students would not be appropriate **[9].**

CONCLUSION:

Teachers should have appropriate scoring procedures to know much more about students' prior language knowledge and abilities, to evaluate their progress, to be able to give more effective feedback. In language classes a new approach to assessment that helps to achieve these goals is very important. Although the criterion-referenced assessment has been developed recently, it is widely used by different assessment projects **[10].** The Canadian Language Benchmarks is one of the assessments based on the criterion-referenced assessments. This kind of assessment in language classes enables teachers not only to grade students' knowledge but also to monitor the learning progress. It tries to give a new concept to assessment. Moreover, it evaluates students' ability to apply the language in different contexts. Especially it can be of a great use for teachers who just have started their career in education **[11].** Having clearly developed assessment criteria, teachers come to be able to assess ESL/EFL learners' language proficiently more appropriately and multifaceted.

REFERENCES:

1. Dean BJ. Testing in Language Programs. A Comprehensive Guide to English Language Assessment, International Edition 2005, McGraw-Hill

2. Dean BJ. Hudson Thom, Criterion-referenced Language Testing, Cambridge University Press, 2002

3. North B. The development of descriptors on scales of language proficiency. Washington DC: National Foreign Language Center. 1993.

4. Alan D. Principles of Language Testing, Applied language studies, Oxford, 1990

5. Thom H. Trends in Assessment Scales and Criterion-referenced Language Assessment, Annual Review of Applied Linguistics, Cambridge University Press. 2005

6. Arthur H. Testing for Language Teachers, Cambridge University Press. 1989

7. Michael OJ, Valdez PL. Authentic Assessment for English Language Learners. Practical Approaches for Teachers, Addison-Wesley Publishing Company, Inc. U.S.A. 1996

8. Smith P, Grazyna. Canadian Language Benchmarks 2000: Theoretical Framework, Centre for Canadian Language Benchmarks, Ottawa (Ontario). 2002.

9. Dunn PL, Sharon, and Morgan Chris M. Seeking quality in criterion referenced assessment, the Learning Communities and Assessment Cultures Conference organized by the EARLI Special Interest Group on Assessment and Evaluation, University of Northumbria, August 2002. 28-30.

10. Timothy S, Sally R, and Bill P. Adapting the Canadian Language Benchmarks for Writing Assessment, TESL Canada Journal, 2001;18(2):48-64.

11. Liz HL. Second Language writing: assessment issues, Second language writing: research insights for the classroom, Barbara Kroll, Cambridge University Press, 1994.