**Mohammad Hossein Keshavarz,**
**P.h.D (Linguistics)**
**University for Teacher Education**

# Norms of Correctness in Native Speaker Usage and EFL Situations

**Abstract**

One of the most important issues in teaching English as a second or foreign language is the selection of norms of correctness in the target language. However, the questions as to what should constitute such norms is, indeed, a highly controversial one. If we consider native speaker usage as norms of correctness the question of language variation poses itself. on the other hand, if we base our norms on traditional grammatical rules, issues such as 'prescriptivism' versus 'descriptivism' need to be taken into consideration. In this article, first the notion of correctness will be dealt with in the context of native speaker usage. Then against this background, norms of correnctness for second or foreign learners of English will be discussed and the criteria for choosing a particular variety of English for teaching purposes as well as for the treatment of errors will be examined

Cziko, G. A. (1982). Improving the psychometric, criterion-referenced and practical qualities of integrative language tests. ***TESOL Quarterly, 16***, 367-379.

Davidson, F. (1992). ***The testing of English as a second/ foreign language in the criterion-referenced era.*** Paper presented at the annual meeting of the National Council of teachers of English, Louisville, KY. (ERIC Document Reproduction Service No. ED 361407)

Davidson, F. & Lynch, B. K. (1992). Criterion-referenced language test development: a prolegomenon. In A. Huhta, K. Sajavaara, & S. Takala (eds.), ***Language testing: new openings.*** Jyvaskyla: University of Jyvaskyla, Institute for Educational Research.

Davies, A. (1990). ***Principles of language testing.*** Oxford: Basil Blackwell.

Hambleton, R. K. & Novic, M. R.(1973). Toward an integration of theory and method for criterion-referenced tests. ***Journal of Educational Measureement, 10*** (3), 159-170.

Hatch, E. & Farhady, H. (1994). ***Research design and statistics.*** Tehran: Rahnama Publications.

Hudson, T. D. & Lynch. B. K. (1984). A criterion-referenced measurement approach to ESL achievement testing. ***Language Testing, 1***, 171-201.

Hudson, T. (1991). Relationships among IRT item discrimination and item fit indices in criterion-referenced language testing. ***Language Testing, 8***, (2), 160-181.

Hughes, A. (1989). ***Testing for language teachers.*** Cambridge: Cambridge University Press.

Kunnan, A. J. (1992). An investigation of a criterion-referenced test using G-theory, and factor and cluster sampling. ***Language Testing, 9***, (1), 30-49.

Lynch, B. (1992). Evaluating a program inside and out. In J. C. Aldrson and A. Beretta. ***Evaluating second language Education*** (pp. 61-99). Cambridge: Cambridge University Press.

Lynch, B. K. & Davidson, F. (1994). Criterion-referenced language test development: linking curricula, teachers, and tests. ***TESOL Quarterly, 28*** (4), 727-743.

Pophom, W. J. (1978). ***Criterion-referenced Measurement.*** New Jersy: Prentice-Hall, Inc.

Purves, A., Soter, A., Takala, S., and Vahapassi, A. (1984). Towards a domain-referenced system for classifying composition assignments. ***Research in the Teaching of English, 18*** (4), 385-416.

Skehan, P. (1984). Issues in the testing of English for specific purposes. ***Language Testing, I,*** 202-20.

Takala, S. (1985). Criterion-referenced testing in foreign language teaching. *In* ***Language testing in school.*** A Finland Yearbook 1985, No. 41. (ERIC Document Reproduction Service No. ED 266682)

Wilson, K. M. (1989). ***TOEIC research report.*** New Jersy: Educational Testing Service.

because CR and NR tests merge into one at this level.

The results of this study might also be helpful for program evaluators and decision makers. As it was mentioned, the use of either NR and CR tests in program evaluation is a matter of controversy. On the one hand, some scholars, such as Bachman and Clark (1987) and Bachamn (1988), have criticized the use of NR tests in program evaluation, due to their insensitivity to the impact of instruction. On the other hand, Lynch (1992) has argued for the superiority of NR tests in program evaluation because they allow for references beyond the specific programs. In fact, NR tests provide decision makers with more information regarding the effect of the program because such tests enable them to reference the test scores to the external norms.

Furthermore, some scholars have questioned the content validity of tests such as The Test of English as a Foreign Language (TOEFL) and have emphasized developing a CR measure of language proficiency. Since TOEFL is most commonly used as an entrance requirement for advanced international students wanting to study for a degree at an American college or university, and developing a CR measure of language proficiency is by no means an easy task, it may not make much difference to give the applicants TOEFL or a CR test.

Although the present study might be considered as a step towards the settlement of the controversies surrounding the use of CR and NR tests, repeated research projects are needed to support the findings of this investigation.

**References**


Alderson, J. C. (1991). Language testing in 1990's: How far have we come? How much further have we to go? In Anivan, S. (Ed.) ***Current developments in language testing*** (pp. 1-26). Singapore: Regional Language Center.

Bachman, L. F. (1988). The development and use of criterion- referenced tests of language proficiency in language program evaluation. In R.K. Johnson (ed.). ***The second language curriculum*** (pp. 242-258). Cambridge: Cambridge University Press.

Bachman, L. F. (1990). ***Fundamental concepts in language testing.*** Oxford: Oxford University Press.

Bachman, L. F. and Clark, J.L.D. (1987). The measurement of foreign/second language proficiency. ***Annals of the American Academy of political and Social Science, 490,*** 20-33.

Brindley, G. P. (1991). Defining language ability: the criteria for criteria. In Anivan, S. (Ed.) ***Current developments in language testing*** (pp. 139-164). Singapore: Regional Language Center.

Brown, J. D. (1989). Improving ESL placement tests using two perspectives. ***TESOL Quarterly,*** *7*(1), 239-260.

Brown, J. D. (1990). Short-cut estimators of criterion-referenced test consistency. ***Language Testing,*** *7*(1), 77-97.

tests and NR tests fades away as the subjects' language proficiency improves.

The results of this study might be regarded as a step towards the settlement of the controversies surrounding the use of CR and NR tests. One might conclude from the findings that individuals' performance on CR and NR measures of language abilities depends on and varies with their proficiency levels: if individuals are proficient, there might be no difference between their performance on CR and NR tests. In other words, CR and NR tests merge at the advanced proficiency levels. This does not hold true for the elementary levels, however. In fact, the individuals' proficiency levels is a key factor affecting their performance on CR and NR tests.

The findings of the present study might also shed some light on the possible nature of language proficiency. While some scholars have attempted to demonstrate that one underlying competence accounts for language proficiency, some others have proposed different competencies, each of which is distinct from the others. However, as hypothesized by Alderson (1991), this study suggests that the nature of language proficiency depends on and varies with individuals' proficiency levels. In fact, these findings imply that language proficiency is multifactorial at the elementary proficiency levels but more unitary at the advanced levels.

With extreme caution, therefore, it can be concluded that the present language proficiency models proposed by some language testing researchers need some modifications. For example, some components in Bachman's (1990) model might not be distinct at the advanced proficiency levels. So models of language proficiency should consider the level-dependency of language proficiency as a major factor. It is clear that any change in the present language proficiency models will cause some minor or major changes in the methods of language teaching and testing.

With respect to the applications of the present study, the findings suggest that for students of low proficiency, performing NR language tests substantially underestimates their level of achievement. However, for highly proficient students, performing on either CR or NR tests does not affect their performance.

The findings of this study might also be a remedy for the problems related to domain specifications in CR test development procedure. As Bachman (1989) states, the only way to develop adequate CR procedures for the assessment of language proficiency is the attempt to specify the abilities that make up language proficiency and define levels of language proficiency which are independent of particular contexts. Since such levels do not specify particular contexts, however, raters would not have any concrete criteria to guide them (Brindely, 1991). This study provides further evidence that the problems related to domain specification might have not much to do with the advanced proficiency levels

As Table 4 indicates, the first factor receives the highest loadings from CR, G.CR, R.CR, and V.CR, while the second factor is heavily loaded by CELT, G.CELT, R.CELT, and V.CELT.

The pattern of the factor loadings of the variables makes the assignment of labels to the factors rather easy. Factor 1 can be called "language proficiency measured by CR tests" and factor 2 "language proficiency measured by NR tests."

It was naturally difficult to extract two factors at the advanced level because of high correlations. Nevertheless, for comparison purposes, a two factor solution was preferred. It then was rotated using Varimax Rotation Procedure. The results are presented in Table 5.

**Table 5**

*Varimax Rotated Factor Loadings of the CR and NR Measures at the Advanced Level*

| | Factor 1 | Factor 2 |
|---|---|---|
| CELT | .66 | .72 |
| CR | .70 | .62 |
| G.CELT | .87 | * |
| G.CR | .65 | .64 |
| R.CELT | .82 | * |
| R.CR | .72 | .52 |
| V.CELT | * | .95 |
| V.CR | .70 | .62 |

**Loadings below .30 were discarded.*

As Table 5 illustrates, while factor 1 is loaded by G.CELT and R.CELT, factor 2 is loaded by V.CELT. It is interesting to observe that the variables CELT, CR, G.CR, R.CR, and V.CR have loaded almost equally on the two factors.

The pattern of the factor loadings of the variables at the advanced level of language proficiency makes the assignment of labels to the factors difficult and one can not claim that the two factors refer to two different underlying abilities.

The comparison of the results of the factor analyses demonstrates that the difference between CR tests and NR tests regarding what they measure fades away as the subjects' language proficiency improves. So, the null hypothesis of the study which states, **"there is no relationship between students' performance on CR and NR measures at two different (elementary and advanced) levels of language proficiency"** can be rejected. The results show that students' performance on CR and NR measures of language abilities depend on their level of proficiency.

## Conclusions

The major aim of the present study was to arrive at a solution regarding the differences and similarities between NR and CR tests of language abilities at two elementary and advanced levels of language proficiency. To that end, first the subjects' proficiency levels were determined and then four CR tests were constructed and administered to the subjects. The results of the statistical analyses demonstrated that the difference between CR

The second analysis dealt with determining the correlations among the CR and NR measures and their subtests at each proficiency level. The correlations are presented in Tables 2 and 3.

**Table 2**

*Correlation Coefficients among the CR and NR Measures at the Elementary Level*

| | CELT | CR | G.CELT | G.CR | R.CELT | R.CR | V.CELT | V.CR |
|---|---|---|---|---|---|---|---|---|
| CELT | * | | | | | | | |
| CR | .58 | * | | | | | | |
| G.CELT | .88 | .55 | * | | | | | |
| G.CR | .63 | .93 | .63 | * | | | | |
| R.CELT | .75 | .51 | .51 | .52 | * | | | |
| R.CR | .41 | .88 | .37 | .73 | .36 | * | | |
| V.CELT | .81 | .40 | .54 | .40 | .53 | .27 | * | |
| V.CR | .53 | .89 | .47 | .73 | .48 | .73 | .41 | * |

**Table 3**

*Correlation Coefficients among the CR and NR Measures at the Advanced Level*

| | CELT | CR | G.CELT | G.CR | R.CELT | R.CR | V.CELT | V.CR |
|---|---|---|---|---|---|---|---|---|
| CELT | * | | | | | | | |
| CR | .86 | * | | | | | | |
| G.CELT | .81 | .76 | * | | | | | |
| G.CR | .86 | .90 | .75 | * | | | | |
| R.CELT | .75 | .68 | .66 | .63 | * | | | |
| R.CR | .82 | .83 | .71 | .76 | .74 | * | | |
| V.CELT | .86 | .70 | .45 | .70 | .49 | .65 | * | |
| V.CR | .91 | .88 | .79 | .84 | .73 | .78 | .74 | * |

These tables reveal that the correlation coefficients observed at the advanced proficiency level are relatively high, while this does not hold true for the elementary level. For instance, the average correlations at the advanced and elementary levels, are .75 and .59 respectively.

Correlation coefficient, however, is just a measure which indicates how closely the two variables go together. According to Hatch and Farhady (1994), a more useful way of interpreting a correlation coefficient is to convert it into common variance between the two measures. Regarding the common variance, it seems that at the advanced level the amount of common variance explained by the CR and NR measures is higher than the amount of common variance explained by the CR and NR measures at the elementary level.

Furthermore, a high correlation between two measures does not necessarily mean that they are measuring the same traits. To investigate the nature and the number of the underlying traits measured by the CR and NR measures, therefore, two principle factor analyses were performed at two different proficiency levels.

First a factor analysis was conducted on the subjects' scores on the 8 variables at the elementary level. Two factors were extracted. The two factor solution then was rotated using Varimax Rotation Procedure. The results of the factor analysis at the elementary level are presented in Table 4.

**Table 4**

*Varimax Rotated Factor Loadings of the CR and NR Measures at the Elementary Level*

| | Factor 1 | Factor 2 |
|---|---|---|
| CELT | .31 | .95 |
| CR | .94 | .32 |
| G.CELT | .33 | .79 |
| G.CR | .82 | .41 |
| R.CELT | .30 | .74 |
| R.CR | .91 | * |
| V.CELT | * | .84 |
| V.CR | .85 | .31 |

**Loadings below .30 were discarded*

worth noting that the tests were scored objectively. There was only one correct answer for each item and the scores were not influenced by the judgment of the scorers.

Comprehensive English Language Test (CELT) was administered to the subjects two weeks before the CR tests to determine the subjects' proficiency levels. The CELT scores, then, were transformed to z scores, resulting in scores with a common mean and standard deviation. To keep the appropriate proportion of subjects at each level, the subjects standing 0.5 standard deviation above the mean were regarded as advanced students; those standing 0.5 SD below the mean were considered as elementary students; and those standing between + 0.5 and - 0.5 SD above and below the mean were classified as intermediate students. The rationale behind the choice of ± 0.5 SD was the classification of subjects into different levels in accordance with the proportion of the normal distribution. Since this study was just concerned with the elementary and advanced levels of language proficiency, students of intermediate proficiency were excluded from the study sample.

## Results

To test the hypothesis of the present study certain statistical techniques were conducted. The first analysis was concerned with determining the reliability coefficients of the NR and CR measures. To determine the reliability coefficients, KR-21 reliability formula was used for CELT and the formula proposed by Brown(1990) for the CR measures.

The rationale for the choice of Brown's (1990) formula was that regarding the application of NR tests reliability formulas to CR tests, there are discrepancies among different scholars. Pophom and Husek(1969), cited in Kunnan(1992); Hambleton and Novick(1973); Pophom(1978); Hudson and Lynch(1984); and Bachman(1990) have argued for the inappropriateness of NR reliability estimates for CR tests. However, Brown(1990) has established a relationship between the NR tests coefficients of reliability and the CR test estimates of dependability. Hudson(1991) also found that NR tests internal consistency estimate correlate highly with those of CR tests dependability estimate.

Table 1 presents the reliability coefficients of the above-mentioned tests. Apart from CELT at the elementary level, the other NR and CR measures enjoy a high reliability. The moderate reliability of CELT at the elementary level can be attributed to the homogeneity of the subjects and the low amount of variance at this level.

**Table 1**

*Reliability Coefficients of the NR (CELT) and CR Measures*

| | Elementary | | Advanced | |
|---|---|---|---|---|
| CELT: K - R 21 | .51 | | .81 | |
| | Test No. 1 | Test No. 2 | Test No. 3 | Test No. 4 |
| CRTs: Phi Dependability Index | .77 | .72 | .72 | .70 |

**The Criterion-referenced Instruments**

Four different CR tests, developed exclusively for the purpose of this study, were used:

1. The Introductory Grammar end-of-course exam (30 items)

2. The Introductory Reading end-of-course exam, consisting of both vocabulary and reading comprehension (30 items)

3. The Advanced Grammar end-of-course exam (40 items)

4. The Advanced Reading end-of-course exam, consisting of both vocabulary and reading comprehension (32 items)

## Procedures

In order to develop the CR instruments, the procedures presented in Pophom (1987) were followed. The first step involved determining the objectives to be achieved. To accomplish this, an attempt was made to use the course objectives for Introductory and Advanced Grammar, and Introductory and Advanced Reading as presented in their corresponding syllabuses.

The second step was to prepare a table of specifications on the basis of the objectives of each course. The third step was concerned with choosing the most appropriate test format in terms of testing time and ease of scoring.

With reference to the grammar tests, the number of pages allocated to each grammatical structure in the textbooks was counted and the number of items allocated to that grammatical structure in the test was proportionately determined. For example, if the grammatical structures were present in the textbooks in the proportions 4:3:2:1, the same was observed in the collection of test items.

As for the vocabulary tests, the number of new words residing in the textbooks was counted in terms of their parts of speech. Then the number of words allocated to every part of speech in the text was proportionately taken into account in the test items.

To prepare passages of appropriate level of difficulty for reading comprehension, the readability indexes of the reading textbooks were determined utilizing the Flesch Reading Ease formula. The obtained figures, then, served as indices to select appropriate passages to be included in the tests. Regarding reading item types, course objectives were determinant. For example, one of the reading objectives states "by the end of this course, students should be able to skim an entire passage for the main idea". An item dealing with skimming, therefore, was included in the test. The same held ture for other reading objectives.

When the tests were prepared, they were reviewed by the instructors teaching these courses to make sure that the tests were geared towards the course objectives. After receiving the instructors´ comments, the tests were revised. This process was repeated three times until the tests were assumed to be appropriate.

Then all the CR tests were administered, scored, and the results were analyzed. It is

be involved in defining the construct of proficiency.

Although a clear definition of language proficiency is a prerequisite for both NR and CR tests, language testing researchers have offered definitions of language proficiency that vary to a great extent. However, as hypothesized by Alderson(1991), whatever the structure of language proficiency is, it depends on and varies with the level of proficiency.

Despite the fact that there may be two opposing positions regading NR and CR tests, some scholars believe that the differences between them are not as great as they are conventionally imagined(Davies, 1990; Brindely, 1991). Furthermore, with respect to the measurement properties, it has been found that both NR and CR tests provide similar information regarding language abilities (Hudson, 1989, 1991; Brown, 1989, 1999; Cook, 1992 cited in Davidson and Lynch, 1992).

The controversies surrounding the use of CR and NR tests and the claim made by Alderson(1991) were the main motivation for this study. More specifically, the purpose of the present research is to investigate the potential differences and similarities between CR and NR tests at two different levels of language abilities, i. e., elementary and advanced.

To examine the nature of the relationship between students' performance on CR and NR measures of language abilities across the above-mentioned proficiency levels, therefore, the following research question was posed:

**Is there any relationship between students' performance on CR and NR measures at two elementary and advanced levels of language proficiency?"**

## Method

### Subjects

The subjects were 47 students enrolled in "Introductory Grammar and Introductory Reading" course, and 112 students who had just passed "Advanced Grammar and Advanced Reading" course, at the University for Teacher Education. They were majoring either in TEFL or Translation. The courses entitled "Introductory Grammar and Introductory Reading" are designed for students of low proficiency and the courses entitled "Advanced Grammar and Advanced Reading" for those of high proficiency.

### Instrumentation

**The Norm-referenced Instrument:**

The Comprehensive English Language Test(CELT), which included structure and vocabulary subsections, along with Nelson's Reading Comprehension test was used as the NR test in order to determine the language proficiency of the subjects and to classify them into three categories of elementary, intermediate, and advanced.

## Introduction

During the last few decades, language testing has become a controversial issue in the field of language education. Certainly, developing good language tests has never been easy, but to those who take it seriously the task seems harder as our description of language abilities improves.

One of the main issues that has aroused many arguments in the field of language testing is the notion of CR versus NR testing. Generally speaking, as stated by Pophom (1978), a norm-referenced test is "a test designed to ascertain an examinee's stauts in relation to the performance of a group of other examiness who have completed that test" (p. 24), while a criterion-referenced test is "a test used to ascertain an individual's status with respect to a well-defined behavioral domain" (p. 93).

In recent years, methods of language testing based on the use of NR measurement have come under criticism and the call for CR tests of language abilities has been intensified. It is believed that CR approaches to language testing provide information which is not completely available through NR approaches. CR tests seem to have shown much promise as a means of testing both students' achievement and program evaluation (Pophom, 1978; Czico, 1982; Lynch and Hudson, 1984; Purves, Soter, Takala, and Vahapassi, 1984; Takala, 1985; Bachman and Clark, 1987; Bachman, 1988, 1990; Hughes, 1989; Wilson, 1989; Brown, 1989, 1990; Davidson, 1992; Lynch & Davidson, 1994).

Despite this increased interest, however, CR approaches have not received a deserving attention. One reason could be the disagreements among scholars on the concept of real CR tests. Skehan(1984), for instance, emphasizing CR tests desirability, claimed that the case against NR tests had been overstated, and their strengths had been overlooked. He criticizes CR tests for the requirement of a detailed specification for the skill to be measured. In his view, such a requirement makes the construction of CR language tests almost impossible. He, further added that since human performance is likely to change over time, one can not be quite sure whether a language user who meets or misses a criterion today will do so some other days.

In addition to CR test's unattainability, Skehan also objected to the terms used in the level descriptions in proficiency scales. The terms such as "more fluency" or "fewer errors" relates the levels to each other rather than to an external standard. That is, despite their claim to be CR, the descriptions contain NR terminology.

Another reason may be the distinction between achievement CR tests and proficiency CR tests (Brindely, 1991). In achievement CR tests, the test developer should be involved in specifying the behavioral domains and establishing a set of standards by which student's performance is judged. In proficiency CR tests, however, the test developer should