
Persistent Issues in Assessment of English as a Foreign Language

Kassim Shaaban

American University of Beirut, Lebanon

The procedures and techniques used in assessment of language abilities have continuously come under scrutiny by educators and researchers whose main concerns have always been to ensure the suitability, practicality, and fairness of the tools of assessment and to have the assessment process generate the right impact on the various aspects of the teaching/learning process. The need for such scrutiny is certainly greater in foreign language contexts where the English classroom is normally the only place for exposure to English in the school or outside it and the teachers themselves are mostly non-native speakers of the language. Needless to say, assessment is perceived as a trying experience in first language situations, and it is more so in second and foreign language contexts where there is an ever-increasing recognition among researchers and educators alike of the role of the affective factors in second language learning (Stevick, 1990; Krashen, 1982; Lozanov, 1979).

The findings of research studies in second and foreign language acquisition (Klein, 1986; Gass & Schachter, 1989; Cook, 1993) have brought about a major shift in teaching approaches, methods, and techniques from structural, teacher-centered, audio lingual, discrete-point teaching methodologies to global, learner centered, affective-humanistic, communicative methodologies (Krashen, 1985, Savignon, 1983; Nunan, 1988; Oller, 1993; Brown, 1994). This shift was accompanied by an acute awareness among researchers and educators of the need for new assessment procedures that would be compatible with the new trends in teaching methodology. This awareness has resulted in a clear shift in testing since the 80's in the direction of making evaluation a teaching tool (Carroll, 1980, Harris & McCann, 1994; Weir, 1993; Brown, 1996). In fact, the learner-centered approaches to language teaching, which have generated "humanistic" methods of teaching, have also called for humanizing evaluation through introducing innovative assessment principles, procedures, and techniques including continuous evaluation, making assessment an integral part of instruction, diversifying alternatives in assessment, and using informal evaluation measures such as conferences, diaries, logs, portfolios, and peer and self-assessment (Oller, 1987; Burton, 1992; Pierce & O'Malley, 1992; O'Neil, 1992; Arter, Spandel, & Culham, 1995; Katz, 1997). Underlying all these new trends is the principle of the inter-relatedness of assessment and instruction. Oller (1987) states that "within such a practical and comprehensive philosophy of language instruction and

testing, every test becomes a natural rung in the ladder toward the instructional goal . . . and every instructional activity in which students participate becomes a language-testing activity” (p. 45).

The present study will discuss some of the issues that are, after years of experience with new methods of teaching and assessment, still stirring controversy and disagreement among testing specialists. Analysis and discussion of these issues will help throw light on the place of assessment in classroom teaching and in the curriculum and will help “. . . teachers and administrators be aware of the issues involved and make informed choices regarding language tests and how they are used” (Savignon, 1987, p. 20).

The Teacher as Tester

One of the immediate consequences of introducing the new alternative procedures and techniques of assessment has been that “control over the collection and interpretation of assessment information has shifted from centralized authority towards the classrooms where assessment occurs on a regular basis” (Fradd & Hudelson, 1995, p. 5). These new developments have made the role of the classroom teacher in the evaluation process a vital and decisive one. In fact, educators agree that classroom teachers are the best judges of their students’ linguistic and communicative abilities. Their judgement would be the most representative and accurate, for only they are aware of all the parameters of the teaching context in which they operate. Harrison (1983) considers that the teacher has the ultimate responsibility for the content and format of the assessment tool as well as for the decision on whether to assess or not, for “he is influenced by the aims and needs of the students he is teaching, the course book he is using, the demands of the school and the system and so on, and must therefore devise tests to fit these conditions” (p. 134).

But assessment is a skill that does not come easy to the practitioner; classroom teachers usually have problems in constructing, administering, and using assessment procedures and techniques. Alderson and Clapham (1995) recognize the problems teachers have with assessing students’ abilities and attitudes and attribute these problems to the association of assessment with “an arcane terminology, a heavy emphasis on numbers and statistics, and an aura of objectivity and rigor which makes people feel that testing is too difficult and that it needs to be left to experts” (p. 184). Traditionally, EFL teachers have tried to develop assessment procedures modeled after the examples given in the textbook series being used or in assessment handbooks. These procedures and techniques need to be looked at by teachers as demonstrations of the various assessment types and formats and of the content, but they should not be taken and modeled after blindly. Teachers of English as a foreign language, working with alternatives in assessment, need training in how to assess students’ performance through

conferencing, reading and learning logs, performance tasks, and portfolios; all these forms require in-depth knowledge of and adequate training in the dynamics of holistic scoring, cooperative learning, team evaluation, inter-rater reliability, and design of the evaluation profile—skills that most non-native English teachers lack. Alderson and Clapham (1995) suggest that these obvious deficiencies could be remedied through training. They believe that training EFL/ESL teachers in constructing achievement and progress tests is not unattainable, for “the construction of class-based tests requires less specialist knowledge and is related far more closely to the devising of class exercises” (p. 185). They further call on teachers to develop explicit marking criteria, which would improve the construction of the tests, their scoring, and the teaching/learning process itself.

Harrison (1983) suggests seeking the help of colleagues in the design and revision of assessment procedures and techniques, for he believes that “no one person can write a test by himself . . . [there is] no substitute for the comments of an interested colleague, who will see the test from a different view point and will point out ambiguities and possibilities for error which the test writer cannot see” (p. 134-35). Heaton (1990) goes a step further and recommends teamwork in the development of tests. “Getting together with other teachers can make the writing of good achievement tests a lot easier. Indeed, such team work will improve all the various kinds of tests you may want to write” (p. 14). Brown (1998) includes in his book samples of the various types of formal and informal assessment procedures and activities prepared by classroom teachers; these activities could be used in a training course or for individual study by teachers.

Test Anxiety

Traditionally, the mere mention of tests brings to mind images of anxiety and fear in most teachers and students. Teachers are usually worried about the fairness, well-constructedness, suitability, and practicality of their tests. Students, on the other hand, are fearful of the unknown and of failure to do well on the tests. The anxiety generated by assessment could be detrimental to a student’s future; it therefore becomes the responsibility of the teacher to provide assessment measures that allow students to show their best performance and ability.

Testing research has shown that some tests cause more anxiety than others. Cohen (1984), for example, reports that in tests of literature, “Open-ended questions are preferred to multiple choice and that the cloze was perceived as a high-anxiety proficiency test” (p. 71). Bradshaw (1990) studied the reactions of Spanish and Italian students to types of tests and the effects of these types on their test scores and motivation. His findings confirm Cohen’s conclusions; the C-test, a variation on the cloze, was more difficult and more anxiety-generating than the multiple choice and the open-ended questions. “The C-test was, overall, the most negatively rated by all groups,

regardless of English proficiency, first language, or gender” (Bradshaw, 1990, p. 25). These findings point to the conclusion that, in order to avoid causing unneeded anxiety, these types of tests should not be used to evaluate students.

Another way of dealing with the anxiety issue is to evaluate students through using the types of assessment they are comfortable with such as open-ended and essay questions. One other suggestion is to allow students to have take-home exams or open-book exams whereby the emphasis will be on the identification, analysis, and synthesis of the needed information. This will help teachers develop and assess their students’ critical thinking skills. More recently, practitioners and researchers have been calling for giving students a say not only in the format of the test but in its content as well. Thus Mayerhof (1992) calls for allowing students to discuss questions during the test quietly as long as each writes his own answers; of course, she is referring to subjective types of questions. Murphey (1994/95) goes beyond the concept of involving students in suggesting topics or points for the test, or generating some questions as suggested by Friel (1989), to having students prepare their own tests. Students choose the items to go into the test; the teacher then identifies the types of questions with the help of the students; the students, a few days later, give each other tests orally in pairs, during class while sitting at their desks, or outside while walking around the premises of the school. Later on, the test is repeated with a new partner to reinforce what is being learned, giving students the feeling that their learning is not for the test only. Students are graded by their partners for the correctness of their answers and for using English.

Finally, it is important to mention that practice with taking tests helps decrease tension normally associated with test-taking. This is especially true with objective tests where familiarity with the format, the instructions, and the question types could help decrease the anxiety of the test-taker when he or she faces the real testing tasks.

Authenticity

A major problem in classroom testing is that which relates to “the relationship between the language use required by tasks on language tests and that which is part of our every day communicative use of the language” (Bachman, 1990, p. 356). This relation is often referred to as “test authenticity.” Modern approaches to language teaching, especially the communicative approach, insist that the language used in testing tasks need to be related to the language used in real life in order for these testing tasks to have credibility. One way of achieving this is through identifying examples of real-life language use required in the assessment of communicative competence and, then, designing test tasks that mirror these examples. But real-life language use is complex and varies among language users. Heaton (1990) cites the following performance objectives that could be included, rather indirectly, in a task-based test: “Can students

understand and deal with messages in English over the telephone? Can they complete an application form for a visa? Can they persuade someone in English to buy a second hand car?" (p. 29). He suggests for the first question a task involving recording a telephone conversation and using it as basis for a listening comprehension test. As for the second question, an authentic application form may be given to be filled out following the instructions usually supplied on the form. For the third question, he suggests using role play as part of an oral test.

Most testing specialists feel that these testing tasks should be evaluated through criterion-referenced scales which usually range from a nil to a perfect level of performance (Bachman, 1990; Clark & Clifford, 1988). Research relating to the psychometric aspects of criterion-referenced tests, including inter-and intra-rater reliability and concurrent validity, is being conducted. As for rating scales, though the idea is attractive, no clear definitions of such scales have emerged. In fact, the "proliferation of rating scales seems counterproductive" (De Jong, 1992, p. 43). Once such scales have been validated and the mechanism of establishing reliability and validity of these performance-based, criterion-referenced, authentic tests has been established, the classroom teacher will feel more comfortable with such tests.

In brief, the development of proficiency-oriented tests for classroom use is recommended by testing specialists who espouse the use of authentic language and authentic, performance-based language tasks; however, the reliability, validity, and practicality (for the classroom teacher) of such tests need to be established. Furthermore, teachers need to be educated and trained in the procedures of communicative test construction; otherwise, they will spend long hours trying procedures they are not sure of. This does not exclude the need for incorporating within testing procedures some performance-based tasks, especially in testing oral fluency. The Oral Proficiency Interview (OPI) is a good example of communicative testing as it provides clear guidelines for the teacher in terms of what to assess and how to score (NCTFL, 1999).

The Place of Standardized Tests

Although standardized tests are not recommended for use as classroom tests, except perhaps for diagnostic, placement, or exit purposes, familiarity with their formats, content, and scoring is a necessity for students who will, in the future, have to be evaluated on the basis of such tests in vocational, professional, or academic contexts. Therefore, training students in how to take standardized tests to help develop the much needed familiarity with their types, formats, and content should be one of the objectives of teaching English in English as a foreign language contexts, especially in the last two or three years of high school.

It is important to point out that testing specialists are moving standardized tests away from multiple-choice, discrete-point testing in the direction of performance based, criterion-referenced, communicative testing to keep them in harmony with classroom methodologies. Thus we have recently witnessed the launch of a new project, TOEFL 2000, which "is charged with leading the way for improved English proficiency assessment in the future" (*TOEFL Update*, 1994, p. 1). The TOEFL Policy Council at Educational Testing Services (ETS) has justified its decision in the following manner.

The impetus for TOEFL 2000 comes from TOEFL users who have called for a test that is more reflective of the current understanding of communicative competence and performance-based language assessment and provides more information about international students' ability to use English for Academic Purposes (EAP). These TOEFL scores users include representatives of the college and university admissions community...and applied linguists, language testers, and second language teachers. (*TOEFL Update*, 1994, p. 2)

Similarly, the Council of Europe calls for criterion-referenced, performance-based assessment on the understanding that "learning in our modern society must be viewed as a continuous and life-long task; such a framework should make it possible to assess progress independent of any curriculum or the time spent on learning" (De Jong, 1992, p. 42).

Another emerging trend in relation to standardized testing is the development of internationally comparable standardized language tests. Thus the Association of Language Testers in Europe has as one of its aims promoting "the establishment of common standards for all stages of the language testing process" (De Jong, 1992, p. 43). Similarly, the International Language Testing Association, established in 1991, has among its goals the development of international standards for assessment which focus on requirements relating to testing procedures and test quality. Research into the comparability of the TOEFL and the Cambridge Test has already been conducted, and more research is being done in this area (Davidson & Bachman, 1990).

A quick glance at standardized tests used nowadays in the United States and Great Britain will show a major tendency towards using authentic texts and performance-based testing (Alderson & North, 1995). However, the procedures involved are too complicated for the untrained classroom teacher to use in his class as models. Hence, giving teachers practice in the use of the different forms of assessment is crucial to the success of classroom applications.

Error and Error Correction

For a long time, learners' errors have been considered by teachers and students alike as evidence of lack of learning. The prevailing practice was to correct errors on the spot following the principle of the preference of immediate, rather than delayed, feedback. However, second/foreign language acquisition research has developed a radically-different view from the above; errors are seen as steps on the way to mastery. They are viewed as developmental in nature, reflecting the stage at which the learner's "interlanguage" is. Therefore, these errors, which represent learning opportunities, should be tolerated unless they interfere with the message (Heaton, 1988).

But learners usually expect and look "for feedback on both their spoken and written output. They must be able to recognize oral feedback when it comes, to interpret written feedback, and above all to know what to do as a result" (Bolitho, 1995, p. 47). Actually, in most EFL classes, failure of teachers to correct their students' errors is viewed by these students as irresponsible behavior on the part of the teachers. Many teachers respond to such feeling by over-correction which could engender in students a sense of total incompetence.

While it is important to stress fluency over accuracy, we should recognize that the teacher in EFL contexts provides the only model and the only source of feedback, unlike the case in ESL contexts where classroom work is reinforced outside the classroom. Therefore, EFL teachers should develop awareness of the seriousness of the error. If it interferes with the message conveyed, it should be corrected; if it is a recurring error despite sustained instruction, it should also be corrected, lest it turns into a fossilized form in the learner's linguistic repertoire.

One way of correcting errors without creating anxiety in learners is to develop among students the concepts of self-correction and peer-correction. With proper training, in the context of learner-centered education, students may gradually take more responsibility for recognizing and correcting their own errors. There are certain learning tasks, like process writing, which by their very nature reinforce the role of the learner in recognizing and correcting the error. Cooperative learning could serve as a good model of classroom interaction, especially in group writing where students are responsible for working together on producing a good manuscript. When a good student corrects the rhetorical or grammatical mistakes of another student in his group for the benefit of the group standing in the class, feelings of anxiety will be much less acute than feelings generated by teacher correction. The teacher takes on a new role in the classroom; he or she becomes the facilitator and the authority on the suitability of corrections made by students.

Conclusions

This paper has examined some of the issues in the assessment of language abilities in the English as a second/foreign language classroom that remain controversial and without clear answers. School administrators, program coordinators, teachers, parents, and students need to be aware of the nature of these thorny issues and of possible solutions so that they may help make assessment a positive contributor to the teaching/learning process. The following points raised in this paper need to be kept in mind whenever assessment of the achievement and abilities of EFL students is considered.

First, assessment procedures can not be and should not be separated from the learning/teaching process. In fact, the principle of the inter-relatedness of teaching and assessment is a major characteristic of all new EFL/ESL teaching methodologies (Oller, 1987; Krashen, 1982; Nunan, 1988). Therefore, classroom tests should reflect teaching objectives, teaching methods, teaching activities and exercises, and teaching materials. Thus, students learning through a communicative ESL/EFL program should not be tested by means of discrete-point tests.

Secondly, evaluation of ESL/EFL assessment procedures and techniques for instruction, administration, content, length, and format must be done in light of student performance and level of anxiety generated. Tests should be looked at as indicators of ability; when they fail to perform their function, they should be abandoned in favor of other more reliable forms of assessment.

Thirdly, the teacher/assessor should not look at himself as the only authority on evaluating his students. He may benefit a great deal from involving his colleagues and the students themselves in the assessment process in general, and test construction in particular.

Fourthly, practice in taking tests could be very useful to students because it could motivate them to study. Actually, examinations are, for many students, the only source of motivation; thus we can clearly see that, in most EFL contexts, both teachers and students stress reading and writing and ignore oral fluency because it is not usually tested. However, training students in test-taking should not lead us to teach for tests only because tests represent, at best, a very minor part of the learners' linguistic and communicative competence.

Finally, as questions remain about whether any form of formal assessment is a fair reflection of the linguistic behavior of the test-taker and about the fairness of tests in terms of their content, discourse structures, and linguistic complexity, teachers should be careful before using formal assessment procedures as the only basis for deciding on students' performance. Preferably, teachers should assess students frequently, using a

multiplicity of tools of evaluation, both formal and informal. Needless to say, no teacher is born with the ability to construct reliable, valid assessment tools. Test construction and test use are skills that need to be acquired through training and experience. The lack of such training and experience could lead teachers to develop or use the wrong tests for their students, to misinterpret the test results, or to misapply these results.

References

- Alderson, J. C., & North, B. (1995). *Language Testing in the 1990's*. Hertfordshire: Phoenix ELT.
- Alderson, J. C., & Clapham, C. (1995). Assessing student performance in the ESL classroom. *TESOL Quarterly*, 29(1), 184-187.
- Arter, J. A., Spandel, V., & Culham, R. (1995). Portfolios for assessment and instruction. Eric Document Reproduction Service (No. ED 388 890).
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bolitho, R. (1995). Current issues: Error correction. *Modern English Teacher*, 4(1), 47-48.
- Bradshaw, J. (1990). Test-takers' reaction to a placement test. *Language Testing*, 7(1), 13-30.
- Brown, H. D. (1994). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- Brown, J. D. (Ed.). (1998). *New Ways of Classroom Assessment*. Alexandria, VA: TESOL.
- Burton, L. (1992). Who assess whom, and to what purpose? In M. Stephens and J. Izzard (Eds.), *Reshaping Assessment Practices: Assessment in the Mathematical Sciences under Challenge* (pp. 1-18). Hawthorn, Victoria: Australian Council for Educational Research.
- Carroll, B. J. (1980). *Testing Communicative Performance*. Oxford: Pergamon.
- Clark, J. L. D., & Clifford, R. J. (1988). The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status, and needed research. In A. Valdman (Ed.), *The Assessment of Foreign Language Oral Proficiency. Studies in Second Language Acquisition*, 10(2), 16-21.
- Cohen, A. D. (1984). On taking language tests. *Language Testing*, 1(1), 70-80.
- Cook, V. (1993). *Linguistics and Second Language Acquisition*. London: Macmillan.

- Davidson, F., & Bachman, L. (1990). The Cambridge-TOEFL comparability study: An example of the cross-national comparison of language tests. In J. H. A. L. De Jong (Ed.), *Standardization in Language Testing*. *AILA Review*, 7, 24-25.
- De Jong, J. H. A. L. (1992). Assessment of language proficiency in the perspective of the 21st century. In J. Matter (Ed.), *Language Teaching in the Twenty First Century: Problems and Prospects*. *AILA Review*, 9, 39-45.
- Educational Testing Services. (1994). *TOEFL Update*. Princeton, NJ: ETS.
- Fradd, S., & Hudelson, S. (1995). Alternative assessment: A process that promotes collaboration and reflection. *TESOL Journal*, 5(1), 5.
- Friel, M. (1989). Reading technical texts: A class test. *English Teaching Forum*, 27(1), 32-33.
- Gass, S. M., & Schachter, J. (1989). *Linguistic Perspectives on Second Language Acquisition*. Cambridge: Cambridge University Press.
- Harris, M., & McCann, P. (1994). *Assessment*. Oxford: Heinemann.
- Harrison, A. (1983). *A Language Testing Handbook*. London: Macmillan.
- Heaton, J. B. (1988). *Writing English Language Tests*. London: Longman.
- Heaton, J. B. (1990). *Classroom Testing*. London: Longman.
- Katz, L. (1997). A developmental approach to assessment of young children. ERIC Document Reproduction Service (No. ED 407 172).
- Klein, W. (1986). *Second Language Acquisition*. Cambridge: Cambridge University Press.
- Krashen, S. (1982). *Principles and Practices in Second Language Acquisition*. Oxford: Pergamon.
- Krashen, S. D. (1985). *Input in Second Language Acquisition*. Oxford: Pergamon.
- Lozanov, G. (1979). *Suggestology and Outlines of Suggestopedy*. New York: Gordon & Breach.
- Mayerhof, E. (1992). Communication dynamics as test anxiety therapy. *English Teaching Forum*, 30(1), 45-47.
- Murphey, T. (1994/1995). Tests: Learning through negotiated interaction. *TESOL Journal*, 4(2), 12-16.
- National Council on the Teaching of Foreign Languages. (1982). *ACTFL Proficiency Guidelines*.
- Nunan, D. (1988). *The Learner-Centered Curriculum*. Cambridge: Cambridge University Press.
- Oller, J. W., Jr. (1987). Practical ideas for language teachers from a quarter century of language testing. *English Teaching Forum*, 25(4), 42-46 & 55.

- Oller, J. W., Jr. (1993). *Methods that Work: Ideas for Literacy and Language Teachers*. (2nd edition). Boston: Heinle & Heinle.
- O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, 49(8): 14-19.
- Pierce, L. V., & O'Malley, J. M. (1992). *Performance and Portfolio Assessment for Language Minority Students*. Washington, DC: National Clearinghouse for Bilingual Education. [Http://www.ncbe.gwu.edu/ncbepubs/pigs/pig9.html](http://www.ncbe.gwu.edu/ncbepubs/pigs/pig9.html)
- Savignon, S. (1983). *Communicative Competence: Theory and Classroom Practice*. Reading, MA: Addison-Wesley.
- Savignon, S. (1987). What's what in communicative language teaching. *English Teaching Forum*, 25(4), 16-21.
- Stevick, E. (1990). *Humanism in Language Teaching: A Critical Perspective*. Oxford: Oxford University Press.
- Weir, C. (1993). *Understanding and Developing Language Tests*. New York: Prentice Hall.

About the Author

Kassim Shaaban is Associate Professor of Applied Linguistics in the English Department at the American University of Beirut. His main research interests are motivation and attitudes, assessment and evaluation, language-in-education policies, cooperative language learning, and teaching effectiveness.
