

Alternatives in Assessment or Alternatives to Assessment: A Solution or a Quandary

Ali Derakhshan

Allameh Tabataba'i University, Tehran, Iran

E-mail: aderakhshan@gmail.com

Saeed Rezaei

Allameh Tabataba'i University, Tehran, Iran

E-mail: Rezaei63@yahoo.com

Minoo Alemi

Sharif University of Technology, Tehran, Iran

E-mail: alemi@sharif.ir

Abstract

The last few years have witnessed the introduction of assessment terminologies into the language evaluation research discussion, elucidating not merely a semantic change but a profound conceptual one, with 'assessment' construed to be an overarching term used to refer to 'all methods and approaches to testing and evaluation whether in research studies or educational contexts' (Kunnan, 2004, p. 1). Assessment fulfils the criteria for being the cutting-edge topic with which comes a recognition of the need for multiple forms of assessment for gathering information for various objectives in diverse contexts (Huerta-Macias, 1995). However, this paradigm shift exceeds far beyond notions of alternative assessment, alternative assessments and alternatives in assessment (Brown & Hudson, 1998), culminating in alternatives to assessment which regard the language evaluation process as a socially constructed activity embedded in the local context with teachers, students and other community members recognized as meaningful assessment partners (Lynch, 2001; Leung, 2004; Lynch & Shaw, 2005; McNamara & Roever, 2006). This paper tries to elaborate comprehensively on the consensuses and controversies of alternative assessments, alternatives in assessment, and alternative approach to assessment.

Keywords: Alternative assessments, Alternatives in assessment, Alternatives to assessment

1. Introduction

Assessment has undergone a paradigm shift, from psychometrics to a broader model of educational assessment, from a testing and examination culture to an assessment culture (Gipps, 1994; Lynch, 2001). Consequently, quite a number of alternative assessments have become widespread. Assessment is a crucial activity in any instructional operation. As a school of thought, which is increasingly gaining acceptance, it argues that it is essential for both learners and teachers to be involved in and have control over the assessment methods, procedures and outcomes, as well as their underlying rationale. Why is alternative assessment accentuated in the 2000's? What does this emphasis on assessment mean for researchers, teachers, and learners?

2. Alternative Assessments

Language testing literature is undoubtedly replete with a variety of alternative assessments which would, in turn, lead to a bone of contention among language testing experts about the benchmarks differentiating alternative assessments from traditional assessments. Put it another way, what common characteristics would distinguish alternative assessments from traditional assessments? What are the implications of introducing alternative assessments? What purposes and functions would alternative assessments serve?

Huerta-Macias (1995) refers to alternative assessment as "an alternative to standardized testing" (p.8). He then enumerates some alternative assessment procedures as checklists, journals, logs, videotapes and audiotapes,

self-evaluation, and teacher observations. He also puts forward some of the characteristics of alternative assessments which are as follows:

- a. alternative assessments are multiculturally sensitive when properly administered;
- b. alternative assessments go beyond the day-to-day classroom activities already in place in a curriculum;
- c. alternative assessments allow students to be assessed on what they normally do in class every day; and
- d. provide information about both the strengths and the weaknesses of students.

In a similar line of inquiry, Aschbacher (1991) enumerates some other common characteristics of alternative assessments, pointing out that they

- a. require problem solving and higher level thinking;
- b. involve tasks that are worthwhile as instructional activities;
- c. focus on processes as well as products;
- d. encourage public disclosure of standards and criteria; and
- e. use real-world contexts or simulations.

A somewhat different set of characteristics were further proposed by Herman, Aschbacher, and Winters (1992, p. 6). They state that alternative assessments

- a. require students to perform, create, produce, or do something;
 - b. tap into higher level thinking and problem-solving skills;
 - c. approximate real-world applications;
 - d. use tasks that represent meaningful instructional activities;
 - e. ensure that people, not machines, do the scoring, using human judgment; and
6. call upon teachers to perform new instructional and assessment roles.

Hargreaves *et. al* (2002) also pinpoint that alternative assessments are often intended to motivate students to take more responsibility for their own learning, to make assessment an integral part of the learning experience, and to embed it in authentic activities that recognize and stimulate students' abilities to create and apply a wide range of knowledge, rather than simply engaging in acts of memorization and basic skill development (Wolf *et al*, 1991; Earl & Cousins, 1995; Stiggins, 1997). Alternative classroom assessments are not ends in themselves but they are designed to foster powerful, productive learning for students (Hargreaves *et. al*, 2002).

As it is cited in Hargreaves *et. al* (2002), House (1981) views alternative assessments from three different perspectives, namely, technological, cultural, political, and Hargreaves *et. al* (2002), propose postmodern perspective. As far as the technological perspective is concerned,

House (1981) believes that teaching and innovation are technologies with predictable solutions that are transferable across different contexts. This perspective focuses on issues of organization, structure, strategy, and skill in developing new assessment techniques. This perspective views alternative assessment as a complex technology that requires sophisticated expertise in, for example, devising valid and reliable measures for performance-based assessments in classrooms, which will capture the complexities of student performance (Torrance, 1995). The challenge of alternative assessment, in this view, is not only to develop defensible technologies that are meaningful and fair but for teachers to develop the understandings and skills necessary to integrate assessment techniques, such as performance-based assessment, portfolios, self-assessment, video journals, and exhibitions, into their practice.

The cultural perspective of classroom assessment puts emphasis on the interplay among points of view, values, and beliefs. From this standpoint, the task of developing alternative assessment moves far beyond technological matters of measurement, skill, coordination, and existing relationships into the area of establishing communication and building understanding among all those involved in the assessment exercise. The political perspective on educational innovation, in House's (1981) view, encompasses the exercise and negotiation of power, authority, and competing interests among groups. A political perspective on alternative assessment recognizes that all assessments involve acts of power, and it identifies the problems of implementing alternative classroom assessment as moving beyond issues of technical coordination and human communication to include the power struggles among ideologies and interest groups in schools and societies. A postmodern perspective on alternative assessment, proposed by Hargreaves *et. al* (2002), is based on the view that in today's complex and uncertain world, human beings are not completely knowable. No assessment process or system can therefore be

fully comprehensive. Alternatively, in positive terms, a postmodern assessment practice can offer multiple representations of students' learning in ways that give maximum voice and visibility to their diverse activities and accomplishments. In this sense, a postmodern system of alternative assessment comprises multiple forms of representation of students' achievement through written, numerical, oral, visual, technological, or dramatic media that are collected in a diverse portfolio of activity and achievement. This allows students' work to be seen through multiple perspectives and allows the complexity of their abilities and identities to be acknowledged more readily.

Recapitulating the perspectives enumerated in Hargreaves *et. al* (2002), the present authors believe that the distinctions between alternative assessments, alternatives in assessments, and alternatives to assessment are blurred. In other words, these terminologies have been interchangeably used by Hargreaves *et. al* while in reality they could be construed differently.

Lynch (2001) elucidates this concept by stating that in order to understand the meaning of alternative assessment and its potential to contribute, along with testing, to our ability to make informed decisions and judgments about individual language ability, we are in need of different research paradigms. In this sense, alternative assessment is meant to describe something more than just procedures and methods. It acknowledges needs for assessment that fall outside of the traditional testing approach and its research paradigm. Alternative assessment points out a research paradigm, a 'culture', which differs from traditional, testing culture.

Wolf *et al.* (1991) and Birenbaum (1996) enumerate some of the characteristics of assessment culture which are as follows:

- a. Teaching and assessment practices should be inextricably bound and integral;
- b. Students should have a voice in the process of developing assessment procedures, including the criteria and standards by which performances are judged.
- c. Both the process and the product of the assessment tasks should be evaluated.
- d. Reporting of assessment results should usually be in the form of a qualitative profile rather than a single score or other quantification.

Lynch (2001) argues that ostensibly these stipulations are amenable to traditional testing culture as well. Yet, within the testing perspective they would be regarded as desirable features if, and only if, the postpositivist, psychometric requirements with respect to reliability and validity which are defined by traditional psychometric characteristics such as inter-rater reliability, objectivity and construct generalizability can be met. He goes on to contend that when attempting to make use of the alternative assessment paradigm, we are not seeking measurement and tests as the means for providing the evidence we need to make decisions. And since, within this paradigm, we are not conceiving of that which we want to assess as an independently existing entity to be measured, the traditional criteria of reliability, generalizability and objectivity will not be the ones we need to make judgments about validity. It can be inferred that Lynch does not try to replace alternative assessment with testing, but rather he believes that we need to judge its validity with criteria that are appropriate to its underlying paradigmatic.

3. Validity and Reliability Revisited from Alternative Assessments' Viewpoint

There is a bone of contention about the validity and reliability of alternative assessments. Huerta-Macías (1995), an advocate of alternative assessments, argues cogently that

trustworthiness of a measure consists of its credibility and auditability. Alternative assessments are in and of themselves valid, due to the direct nature of the assessment. Consistency is ensured by the auditability of the procedure (leaving evidence of decision making processes), by using multiple tasks, by training judges to use clear criteria, and by triangulating any decision making process with varied sources of data (for example, students, families, and teachers). Alternative assessment consists of valid and reliable procedures that avoid many of the problems inherent in traditional testing including norming, linguistic, and cultural biases. (p. 10)

Brown & Hudson (1998) concede that credibility, auditability, multiple tasks, rater training, clear criteria, and triangulation of any decision-making procedures along with varied sources of data are important ways to improve the reliability and validity of any assessment procedures used in any educational institution. In fact, these ideas are not new at all. What is new is the notion that doing these things is enough, that doing these things obviates the necessity of demonstrating the reliability and validity of the assessment procedures involved.

However, they are adamant that alternative assessment should be judged by the same token as conventional tests:

As in all other forms assessment, the designers and users of alternative assessments must make every effort to structure the ways they design, pilot, analyze, and revise the procedures so the reliability and validity of the procedures can be studied, demonstrated, and improved. The resulting decision-making process should also take into account what testers know about the standard error of measurement and standards setting. Precedents exist for clearly demonstrating the reliability and validity of such procedures in the long-extant performance assessment branch of the educational testing literature, and the field of language testing should adapt those procedures to the purposes of developing sound alternative assessments. (p. 656)

Norris *et. al* (1998) also contend that “the issues of reliability and validity must be dealt with for alternative assessments just as they are for any other type of assessment—in an open, honest, clear, demonstrable, and convincing way” (p. 5).

In a similar vein, Lynch (2001) objects to this view on philosophical grounds, contending that alternative assessment represents a different paradigm (an ‘assessment culture’) and therefore cannot be evaluated from within the traditional positivist framework of educational measurement (a ‘testing culture’). He goes on to state that if all forms of assessment were united by the same requirements for reliability and validity, then he would agree with Brown and Hudson (1998) who prefer the term ‘alternatives in assessment’.

4. Alternatives in Assessment

Brown & Hudson (1998) criticize alternative assessments on three grounds stating that the phrase alternative assessments may itself be somehow destructive because it implies three things: (a) that these assessment procedures (like alternative music and the alternative press) are somehow a completely new way of doing things, (b) that they are somehow completely separate and different, and (c) that they are somehow exempt from the requirements of responsible test construction and decision making. They further view procedures like portfolios, conferences, diaries, self-assessments, and peer assessments not as *alternative assessments* but rather as *alternatives in assessment*. Language teachers have always done assessment in one form or another, and these new procedures are just new developments in that long tradition.

They also classified the various kinds of language assessments into three broad categories: (a) selected-response assessments (including true-false, matching, and multiple-choice assessments); (b) constructed response assessments (including fill-in, short-answer, and performance assessments); and (c) personal-response assessments (including conference, portfolio, and self- or peer assessments). The third category is what they call alternatives in assessment.

The present researchers, however, believe that the counterarguments leveled against alternative assessments by Brown & Hudson (1998) do not hold water simply because of the fact that even within the same category they could have included their third category. Moreover, they argue that the phrase alternative assessments is destructive because it requires somehow a completely new way of doing things, but the present authors believe that Brown & Hudson must have probably misconstrued the phrase *alternatives to assessment* instead of alternative assessments. Logically speaking, it is an *alternative to something* which tries to bring about some major modifications and renaissances to its precursors. Here, alternative assessment itself is comprehensive enough which does not necessitate subsuming it under alternatives in assessment. The present authors go on to explicate that even the categories given by Brown & Hudson (1998) are conventional and rather haphazard.

Moreover, the second bone of contention is attributed to the second criticism leveled against the phrase alternative assessments. The second pitfall is almost the same as the first one. When something requires "somehow a completely new way of doing things", it, in turn, necessitates a separate and different way of doing. We assume that since the phrase "alternatives in assessment" is more eye-catching, Brown & Hudson (1998) have fully endorsed it in their paper.

5. Alternative Approach to Assessment

The phrase alternative approach to assessment is proposed by McNamara (2001). He states that making the needs of learners a priority would represent an alternative approach to assessment, not ‘alternative assessment’, which is compatible at face value with a scales and- frameworks approach. He further suggests the rationales underlying his stipulation. First, such a focus would lead to a greater research emphasis on the implementation of assessment schemes, including an analysis of the impact of assessment reforms and a critique of their consequences. Secondly, assessment specialists can help more adequately to theorize and conceptualize alternative, more facilitative functions of assessment in classrooms. These would not replace but supplement the functions of placing students in appropriate groupings for learning, certifying achievement and the like. Such a

step involves expanding our notion of assessment to include a range of activities that are informed by assessment concepts and that are targeted directly at the learning process.

McNamara (2001) defines assessment as “any deliberate, sustained and explicit reflection by teachers (and by learners) on the qualities of a learner’s work can be thought of as a kind of assessment” (p.343). He further demonstrates that while most performance assessment procedures require such reflection as a key component, it should not be confined to those contexts in which formal reports or whole-class comparisons (class tests) are involved. Instead, teachers and learners can engage in systematic reflection on the characteristics of an individual performance as an aid to the formulation of learning goals in a variety of contexts.

From this standpoint, teachers are not engaged in comparison of performances of different individuals, except in order to sharpen awareness of the characteristics or features of difference. McNamara clarifies this viewpoint by stating that we are not interested in who is relatively better or worse. We are not even involved in thinking of performances against a particular benchmark. Even where performances of the same individual at different points are compared, it should be sought largely descriptively and qualitatively and should not result in questions of score comparison.

Having been enlightened by Mesick’s validity framework (1982), McNamara (2001) recapitulates that the kinds of difficulties with subjective assessment that are exposed through careful validation research are not a major concern with this approach.

He goes on to justify his standpoint by elaborating that from a certain perspective, each instance of this kind of assessment is unique; it does not always have to be fitted into a larger framework of comparison across individuals or across occasions (although such comparisons, particularly of progress over time, can facilitate consciousness of the nature of development) Nor does this kind of assessment activity necessarily involve record keeping and reporting to fulfill managerialist agendas. Teachers may opt to record some details of their reflections to help them see the bigger picture of development over time, but this need not be formalized into a report.

Notwithstanding the fact that Lynch (2001) does not propose alternative approach to assessment as an overarching framework where the critical perspective would be embedded, the researchers have a predilection to argue that if McNamara’s stipulations (2001) are theoretically and pedagogically plausible, Lynch’s critical perspective (2001) can be regarded as alternative approach to assessment whereby language ability and use can best be understood as realms of social life that do not exist independently of our attempts to know them. Judgments or decisions about language ability and use cannot, therefore, be accomplished as a measurement task.

Shohamy’s (2001) democratic perspective can also be construed as an alternative approach to assessment by which she means that there is evidence that tests are often introduced by those in authority as disciplinary tools, often in covert ways for the purpose of manipulating educational systems and for imposing the agendas of those in authority. However, such uses of tests as instruments of power violate fundamental values and principles of democratic practices. It can be concluded that while McNamara (2001) proposes alternative approach to assessment embedding social dimension to assessment, Lynch (2001) and Shohamy (2001) neither implicitly nor explicitly mention *alternative approach to assessment*.

6. Conclusion

The researchers conclude that alternative assessment and all other derived concepts (alternative assessments, alternatives in assessment, and alternative approach to assessment) are trendy buzzwords which can be placed along the same continuum with little or no major pedagogical and practical differences. Although they may differ in respect to reliability and validity issues, in reality, they are operationalized the same. Tests are still the same tests with the same purposes that they used to serve; they are still used by those in power without giving individuals any right to voice their ideas; and the critical aspects of tests and evaluative measurements have not yet been fully embraced.

References

- Aschbacher, P. A. (1991). Performance assessment: State activity, interest, and concerns. *Applied Measurement in Education*, 4: 275–288.
- Birenbaum, M. (1996). Assessment 2000: towards a pluralistic approach to assessment. In Birenbaum, M. and Dochy, F.J.R.C., editors, *Alternatives in assessment of achievements, learning processes, and prior knowledge*. Dordrecht, Netherlands: Kluwer, 3–29.

- Brindley, G. (2001). Outcomes-based assessment in practice: some examples and emerging insights. *Language Testing*, 18(4): 393–408.
- Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4): 653–675.
- Earl, L., & Cousins, J. B. (1995). *Classroom assessment: Changing the face, facing the change*. Toronto, Canada: Ontario Public School Teachers Federation.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Hargreaves, A., Earl, L. & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1): 69–95.
- Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5(1): 8–11.
- Kunnan, A. J. (2004). Regarding language assessment. *Language Assessment Quarterly*, 1(1): 1.
- Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice and change. *Language Assessment Quarterly*, 1(1): 19–41.
- Lynch B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4): 351–372.
- Lynch, B. & Shaw, P. (2005). Portfolios, power and ethics. *TESOL Quarterly*, 39(2): 263–297.
- McNamara, T. & Roever, C. (2006). *Language Testing: The social dimension*. Oxford: Blackwell Publishing.
- Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments*. Honolulu: University of Hawai'i Press.
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18 (4): 373-391.
- Stiggins, R. (1997). *Student-centered classroom assessment*. Englewood Cliffs, NJ: Prentice-Hall.
- Torrance, H. (1995). *Evaluating authentic assessment: Problems and possibilities in new approaches to assessment*. Philadelphia: Open University Press.
- Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of Research in Education*, 17.