

Ethics and Validity Stance in Educational Assessment

Masoomeh Estaji

Ph. D, Allameh Tabataba'i University, Iran

E-mail: mestaji74@gmail.com

Received: June 27, 2011

Accepted: August 1, 2011

Published: December 1, 2011

doi:10.5539/ells.v1n2p89

URL: <http://dx.doi.org/10.5539/ells.v1n2p89>

Abstract

Assessment is a powerful tool which can shape curriculum, teaching and learning. One of the major topics which has been the focus of so much debate among the scholars is the concept of ethics in educational assessment systems. The purpose of this paper is to touch upon the issue of ethics in language assessment from validity perspectives. To ensure effective planning, implementation, and evaluation of learning, certain principles and guidelines for beliefs and behaviors are adopted. These principles and guidelines are identified as ethical standards. The central question to be addressed here is whether any test can be defended as ethical, or moral. This study also examines professional ethics, specifically ethical principles as they relate to educational assessment. And finally, some suggestions and guidelines will be offered for applying ethics in educational assessment to maximize validity and fair application of language test scores.

Keywords: Ethics, Equity, Consequential validity, Fairness, Power relations

1. Introduction

Language assessment is a complex and rapidly evolving field which underwent significant change in the 1990. From a theoretical perspective, considerable progress has been made which can model the multiple factors involved in formulating ethical standards (Bachman and Eignor, 1997). A sound assessment system provides information about a full range of knowledge and abilities considered valuable and important for students to learn, and, therefore, requires a variety of assessment methods. Multiple measures of assessment should be used in evaluating students as opposed to a single standardized test. Alternative assessment represents a different research paradigm, a different cultural view about assessment. It requires a framework for making the case for validity that differs from the traditional testing approach. The framework required for alternative assessment must include the relations of power, fairness, equity & ethical aspects that exist in the assessment process.

All purposes and uses of assessment should be beneficial to students. The results should be used to overcome systemic inequalities. Assessment results should be reported in the context of other relevant information. In addition, assessment tasks and procedures must take into account cultural, racial, class, and gender differences. Therefore, the assessment system should be subject to continuous review and improvement.

In the last few decades, appreciable attention has been directed toward a fair implementation of test results under the topic of ethics in language testing (Kunnan, 2000, Lynch, 1997 and Shohamy, 1997). The issue is touched upon here from different perspectives. More specifically, this study intends to answer the following research questions.

1. Do the assessment and reporting principles represent ethicality in educational assessment?
2. Can the ethicality of alternative forms of assessment be judged, via applying Lynch and Shaw's framework of examining educational assessment?

To this end, the researcher presents the key issues in assessment; ethics and validity including considerations of fairness, equity, and consequential validity which make an ethical validity framework. The conclusion highlights the need for further investigation & thought. Therefore, promoting ethical practices will be an important goal in educational assessment.

2. Assessment (A Paradigm Shift)

Assessment is a popular and sometimes misunderstood term in current educational practice. The term assessment refers to a variety of ways of collecting information on a learner's language ability or achievement. Although testing and assessment are often used interchangeably, the latter is an umbrella term encompassing measurement instruments, as well as qualitative methods of monitoring and recoding student learning such as observation,

simulations or project work. Assessment is undergoing a paradigm shift, from psychometrics to a broader model of educational assessment, from a testing and examination culture to an assessment culture. There is a wide range of assessment in use now than there was twenty five years ago: teacher assessment, standard tasks, coursework, records of achievement as well as practical and oral assessment, written examinations and standardized tests. There is criterion-referenced assessment, formative assessment, written examinations, standardized tests, performance-based assessment, as well as norm-referenced assessment. In addition, assessment has taken on a high profile and is required to achieve a wide range of purposes: selection, certification, accountability, diagnosis, instructional decision-making, motivation, and drive curriculum and teaching.

These new forms and range of purposes for assessment mean that the major traditional model underpinning assessment theory, the psychometric model, is no longer adequate, hence the paradigm shift. A paradigm shift or scientific revolution occurs when the old paradigm is unable to deal with an outstanding problem (Kuhn, 1970) i.e. to develop the theory of educational assessment.

Therefore, we need to consider ethical issues in the framework which will guide our development and use of assessment, bearing in mind the enormous influence that assessment has on pupils' lives.

3. Alternative Assessment & Its Perspective on Validity

In its simplest form, alternative assessment is a turn to something new, a search for a wider palette of choices. This vision of alternative assessment appears in such characterizations as Shohamy's (1996) "Portfolios, observations, poor assessment interviews, simulations ... self assessment" (pp. 153) or Darling-Hammond's (1994) "oral presentations debates ... exhibition ... videotapes of performances and other learning occasions, constructions & models ... experiments" (pp. 5-6). However such listings of various learning activities and products converted into components of assessment will not, in the end, be satisfactory. Indeed Brown and Hudson (1998) & Norris, Brown, Hudson, Yoshioka (1998, pp. 3) argue that these are not new forms of assessment at all, that they do not represent an alternative in the sense of being completely different from the range of assessment practices that have gone before, and they suggest using instead the term alternatives in assessment.

Perhaps, then, the true soul of assessment lies not in the components or tools, but in the perspective or set of assumptions motivating their use. Wolf et al (1991) have characterized this notion as assessment culture. This culture assumes that assessment involves an investigation of developmental sequences in student learning, a sampling of genuine performances that reveal the underlying thinking processes, and the provision of directions and opportunities for further learning. Assessment culture also assumes that teaching, learning, and assessment practices are inseparable & integral; students should be active participants in the process of developing assessment procedures, including the criteria and standards by which performances are judged; both the process and product of assessment tasks should be evaluated & the assessment results should be reported as a qualitative profile rather than a single score or other quantification (Birenbaum, 1996).

Traditional testing establishes validity by adhering to certain principles (e.g. reliability i.e. it must provide scores that are consistent, usually estimated by determining how consistent the answers on the individual test items are across a representative group of test takers) before it can claim to produce valid inferences or conclusions. Moss (1994, 1996) challenges this tenet. In articulating an approach to validity from the alternative assessment perspective, Moss (1994) argues for a hermeneutic approach. This approach represents an interpretive research perspective and responds to some of the essential qualities of alternative assessment. It acknowledges the importance of the context of assessment and begins to formulate validity as a consensus reached through dialogue which happens between stakeholders in the assessment - teachers, students, parents - not between disinterested external experts.

Moss (1996) also challenges the principle in traditional testing and assessment that the goal is generalization. This goal means that educators want to be able to generalize from one performance, or from one set of test answers, to all such performances, contexts, or areas of knowledge. Validity, from this perspective, is the degree to which the specific performance or test score matches constructs or meanings that educators have predetermined as the object of assessment. Comparing current generalized understandings with new, contextualized interpretations is done in a reciprocal fashion rather than in the scientific hypothesis-testing mode; that is, there is a give and take between one interpretation and the next that results in a reshaping or reconceptualization of both (Hamp-Lyons & Lynch, 1998).

The important contrasts with traditional validity raised by alternative assessment, then, include challenges to the notions that reliability is prerequisite for validity and that generalizability is the ultimate goal. Further, the discussion of the differences in research paradigms leads to a different conceptualization of validity and its criteria, and these criteria can be examined from various ethical perspectives.

4. Ethics

Ethics is what scholars such as Shohamy, Spolsky and Hamp-Lyons (2001) call critical pedagogy or critical language testing. One of the by-products of a rapidly growing testing industry is the danger of an abuse of power. A potentially reasonable way to avoid power misuse in testing is to inform those in power through giving them information on the nature and consequences of the tests. Therefore, the orientation should be directed toward the people in the power position in order to help them make fair decisions. These who are affected by test scores also need orientation to understand how they have been judged and what the test scores could mean to them. Nevertheless, decision-makers are usually administrators, bureaucrats, and politicians who are not well aware of the problems involved in testing. That is why language testers have always complained about the total authoritarian attitude of the decision-makers. Language testers claim that the decisions should be made with taking the potential flaws of test scores into considerations. Language testers also claim that the decisions should be fair, not harming, and just regarding the test-takers abilities. In recent years, these issues in language testing have been receiving an increasing attention under the cover term of "Ethics" (i.e. the moral and social considerations involved in the construction and use of tests).

No one would deny the fact that ethical considerations are important, not only in language testing, but also in any other academic activity. Nor would anyone disagree that language testers should assume responsibility toward social and individual aspects of the test-takers' lives. What is difficult to operationalize, however, is the definition, extent, and limits of ethics.

Regarding ethical issues, then, some confusions exist. Concepts such as fairness, bias, morality, and the context in which these concepts have been used are not at all clearly identified in the field & the effects of other parameters of ethicality such as deception, privacy, equity, and confidentiality seems to be, at best, unclear. Furthermore, the importance of the consequences of the test results, what Messick (1989) calls consequential validity, seems to have been overemphasized in recent years. It can be argued that in some sociopolitical contexts, a tester is not even allowed to assume responsibility towards the consequences of the decisions made on the test scores. In such cases, Language testers are assigned to develop such tests and deliver the tests to certain governmental agencies. What kind of use the agencies would make of the test scores is often a matter of sociopolitical concerns rather than an academic endeavor. Thus, ethical considerations should be discussed in different contexts each with different parameters because what may be ethical in one context, may not be so in another. Some of these contexts are educational contexts (research VS. decision making), sociopolitical contexts (public VS. government), and moral contexts (fairness VS. bias). In addition, the application of ethical concerns in terms of validity, access & justice should be taken into account in all the stages of test development (Kunnan, 2000).

In an assessment which looks for best rather than typical performance the context of the item should be the one which allows the pupil to perform well but this suggests different tasks for different groups which is in itself hugely problematic. However, what we can seek is the use, within any assessment program, of a range of assessment tasks involving a variety of contexts, a range of modes within the assessment, and a range of response format and style. This broadening of approach is most likely to offer pupils alternative opportunities to demonstrate achievement if they are disadvantaged by any one particular assessment in the program (Gipps, 1994).

5. Codes of Ethics and Ethical Principles

In a research study on five major ethical codes in the field of language testing, Jia (2009) examines the different perspectives of professional codes of ethics. What constitutes an ethical code (or code of ethics)? Davies (2005, p. 46) states that an ethical code is "a public written statement by an organization of its principled approach to its professional and/or business activities". In other words, ethical codes refer to "a set of social practices that has a purpose, namely, the promotion of the common welfare" (Davies, 2004, pp. 98).

According to Fortin (1996, pp. 1720) successful codes of ethics are not idealized descriptions of behavior – they are evolving codifications of baseline standards in a given field. Fortin also points out that codes of ethics, by themselves, do not create virtue: what they can do is to increase the likelihood of responsible behavior mainly by "appealing to the moral sense of those whose activities they govern" (1996, pp. 1720). For this to occur, there needs to be an ongoing process of engagement with a given code and a basic acceptance of the ideas behind it. Longstaff (1994) echoes this point by stating codes of ethics must be internalized by their practitioners to be effective. For that reason, successful codes also need to be simple enough so that they can actually be remembered and universal enough so that different persons from varied cultures in a given profession can agree with them. Codes have a valuable function in shaping behaviors, yet they should also be flexible enough to recognize that many teachers and test developers operate under diverse (and often less than optimal) conditions.

Similarly, ethical principles should guide the behaviors of professionals in everyday practice. Principles are assumed

to be constant and, therefore, provide consistent guidelines for decision-making. In addition, professionals should strive to develop the virtues, or habits of behavior, that are characteristic of people in helping professions. Contextual issues must also be taken into account. Such issues include, but are not limited to, culture, temporality (issues bound by time), and phenomenology (individual perspective) and community norms.

6. The Relationship of Ethics to Validity

As with validity, ethical determinations can be made from obviously different perspectives. Ethical practice should be viewed as protecting research participants' basic right not to be harmed (socially, psychologically, emotionally, physically) and not to be coerced or manipulated. Hamp-Lyons (1989) points out that ethics is properly seen as an essential aspect of validity. This interrelationship can be thought of in terms of the following questions:

What should be the relationship between assessor and the person being assessed?

What should the person being assessed be asked to do to demonstrate his or her ability?

What does the assessor believe the nature of that ability to be?

How does the assessor decide or, perhaps, who decides what counts as evidence?

The way these questions are answered will differ, depending on the assessor's perspective (i.e., conventional testing or alternative assessment). It is arguable that the basic ethical questions should remain the same, regardless of the approach to assessment, that is, whether conventional or alternative assessment is being carried out, test-takers' rights not to be harmed or coerced remain constant. However, depending on the validity perspective, how educators define harm, coercion and power relations and the degree of ethical responsibility that they attribute to those in the assessment context may differ.

Raising the stakes enhances the ethical dimension in assessment. The higher the stakes, the greater the likelihood of cheating and unethical practices and the more important it is that assessment practice be fair and equitable. The ethics of assessment demand that the constructs and assessment criteria are made available to pupils and teachers, and that a range of tasks and assessment be included in an assessment program. These requirements are consonant with enhancing construct validity in any case.

We return to these ideas concerning alternative assessment and synthesize them into a framework for addressing ethics and validity in educational assessment.

7. Examining Validity and Ethics of Educational Assessment

Considering Validity from the alternative assessment perspective and many scholar's thought (e.g. Bachman, Brown & Hudson, , chapel, Foucault, Guba & Lincoln, Ham-Lyons, Linn, Messick, Kunnan, Moss, (Norris, Brown & Hudson), shohamy & others) in relation to ethics, then, has provided a framework for examining educational assessment in the admission programs (Lynch & Shaw, 2005).

As displayed in table 1 the validity framework developed integrates validity with ethical considerations, especially in terms of consciously addressing the power relations that are at play in the assessment context. By examining the multiple perspectives of this framework and a consideration of alternative assessment outcomes, most of the concerns of an ethical approach can be addressed.

8. Methodology

The qualitative research elicits phenomenological data that represents the worldview of the participants being investigated. A primary goal of this qualitative study is to develop a better understanding of the principles involved in ethical assessment. This study intends to explore the principles exploited by the teachers in the process of language assessment. It investigates these questions through the approach of naturalistic inquiry. The researcher gathered data through the use of in-depth interviews guided by a list of principles for assessment and reporting (see Appendix A and B) reflecting both the research literature and the research questions underpinning the study. This semi-structured interview approach was chosen because it ensured a degree of consistency in the data collected while also allowing for the use of probes to invite participants to elaborate on their attitudes and experiences (Patton, 2002).

8.1 Participants

The sample of the study included teachers at KEI (Kish English Institute). They were 15 English instructors of (Upper-Intermediate and Advance levels). The participants involved in the ethical assessment project were all Iranian. Advanced, rather than teachers from other proficiency levels, were selected for the study on the grounds that they would have been more concerned with and have a better knowledge of the philosophy of the test, assessment, and ethicality issues. Because this study focused on the ethical assessment perspective of teachers, it was this cohort

of teachers that formed the pool of potential participants for our investigation. At the time the study was undertaken, all potential participants had had at least three years of teaching and as a consequence had made assessment-related decisions for their students that ranged from making tests, piloting tests, grading individual assignments to marking final exams, and making end-of-year pass/fail decisions. They were all female instructors (mostly aged 21-39). The teachers were interviewed in order to garner their views of the assessment and reporting principles. The key questions for these interviews were whether the proposed principles represent ethicality and to what extent these principles signify ethicality in educational assessment.

8.2 Data collection and analysis

Semi-structured interviews were held with Upper-Intermediate and advanced English teachers. These include questions regarding tests and ethical concerns, test decisions, equity, power relation, consequences and fairness. Teachers were asked to respond to their perceptions and beliefs about the tests and their attitude toward the proposed principles for assessment and reporting. Teachers were allowed the freedom to express their personal reactions to the principles and to any changes that they expected in educational assessment. All interviews were recorded on a Digital Voice Recorder, which were transcribed and reviewed later. The interviews were conducted by the researcher who transcribed, summarized, and compared the participants' responses after the interviews.

The interviews were administered to the teachers of KEI. The interviews with the teachers were held on a one to one basis. Each interview lasted for about one hour. All interviews were recorded and then transcribed. Finally, the responses were summarized and compared with responses of the other teachers. Data analysis involved an iterative process of examining the collated data segments, field notes, and interview transcripts; grouping topics; and identifying themes in the participants' comments. These themes were used to build a portrait of ethicality in educational assessment based on what the participants reported. Recurrent themes regarding the ethical principles were identified and the information has proved to be very valuable in understanding the issues under study.

9. Results and Discussion

Interviews on the issues of ethicality and the concern for the relationship of assessment to educational opportunity were carried out with 15 teachers and the following themes emerged from the teachers' interviews. The question to be addressed was whether the assessment and reporting principles represent ethicality in educational assessment. To this end, semi-structured interviews, with focus on the principles given in the questionnaire and ethicality issues, were conducted. As with most questions of ethicality, the responses varied and depended on who was asked. The findings can be summarized under two main categories i.e. current perspective of teachers as to the assessment principles and attitudes toward the implementation of the new assessment system. Below is the presentation of each of these themes.

9.1 Teachers' perspective of the assessment principles

The themes indicated that the perspectives of teachers concerning the assessment principles were different. In particular, the ethicality of the tests mainly achievement tests was called into questions. A majority of the teachers expressed the perception that the assessment principles bring about a great change in educational assessment. Specific comments that revealed a sense that tests were questionable from an ethical point of view included: "it is not an accurate reflection of the learners' capabilities". Responses from interviews with teachers helped to clarify these perceptions. One teacher said that she preferred the assessment done by the teachers while including the assessment principles since the learners will not be judged as better or worse than others but rather on their self-improvement and progress. One of the teachers interviewed gave other insights into the question of why the tests might not demonstrate the learners' competence or capabilities over different skills: the format of the test might be unfamiliar or an unclear instruction would not allow them to demonstrate what they know. Evaluation based on a single test is not relevant, in terms of how the learners approach the tasks; not valid to assess progress. For that you need conferencing, observation, formative assessment, not a "one-off" like an achievement test. Since the learners may even understand the process, but get the wrong answer. This is an inappropriate form of assessment or evaluation. Bennett (1995) surmises that culturally responsive teacher education programs must include opportunities for pre-service teachers to work in schools where all students learn and develop to their highest potential; where teachers and students understand and begin to develop multiple ways of perceiving, believing, behaving, and evaluating.

This study also found that institutional requirements are a significant component of ethical dilemmas that teachers face with assessment. Given the strong impact policy and politics have in shaping and constraining classroom assessment practices and principles across nations (Black & Wiliam, 2005), these findings may not be surprising. They also reinforce the work of Colnerud (1997) who found a similar prevalence of institutional effects on general ethical dilemmas faced by teachers in Sweden. As evidenced in one teacher's comments, contrary to the other

teachers' comments that these principles can be used as a basis for competition between institutions or schools in order to promote their schools over the others or may compete for attracting more students or raising their fees or their funding. More importantly, these principles can not be used to identify areas of need for different individuals. Hence, the needs of different learners go largely undetected. In contrast, one teacher who was also a test developer and in favor of these principles commented that these principles can provide a lot of evaluative information on the curriculum as well as giving teachers a better idea of whether their students are making good progress.

As a result, the ethicality of these principles is a complicated or complex issue. Different values have been placed, with some teachers feeling that it leads to further competition and competitive testing and others feeling that it results in important information for themselves, their learners, and the others. However, the teachers unanimously expressed the idea that assessment plays a central part in education. It appears that identifying a limited number of principles to reveal the nature of the classroom and educational assessment is what makes it morally questionable. This leads to the question of whether there are more ethical ways of assessing than are represented by these limited principles. At the heart of these principles is the notion that, in order for assessment to be ethical or moral, teachers and students should be actively involved and the assessment data should not be used to make competitive judgments about the systems or schools across the country. In fact, good assessment should be "integral to the curriculum". This relates to the notion of washback that tests do best to "reveal their influence back into the curriculum".

9.2 Attitude toward the implementation of the new assessment system

Promoting ethical practices in assessment is considered to be a very important goal of the teachers involved in assessment. Principles and codes which can increase the awareness of ethical practices among teachers and students and promote ethical uses of assessment in various contexts: teaching, counseling, evaluation, research, among others. Strike (1990) argues that explicit instruction in ethical concepts ought to be part of teacher preparation. These concepts should be directly related to the work in which teachers engage.

A set of principles for assessing language, in particular, was developed by Wilson (see Appendix B). These are the principles that are more in line with alternative forms of assessment, and achievement tests are not always designed to respond to these sorts of concerns. This is not surprising, given that Wilson's perspective is clearly in favor of alternative assessment over conventional testing, and the given assessment principles are not designed and can not respond to these sorts of concerns. According to teachers and test developers, the search for the ethical test has uncovered difficulties for traditional, standardized forms of testing. With concerns for the consequences of tests forming part of more recent approaches to validity, ethical concerns and fairness issues need to be examined with more than test internal examinations of reliability. In addition, educating others to understand and to engage in ethical practices is a critical goal. Illustrations of good and bad practice within realistic assessment contexts and discussions of ethical dilemmas are excellent ways of promoting ethically practice in assessment.

If alternative forms of assessment are to be pursued as our best hope of improving the morality of the procedures and the decisions that result from them, they will need to be validated with different procedures from those currently employed for traditional tests (Lynch, 1997). However, in the words of Kenneth Howe (1994, p. 31), "to reject it for these reasons places the quest for accurate measurement-and control- above the quest or educationally and morally defensible policies". This search requires an ongoing ethical assessment using multiple assessment instruments and procedures.

10. Conclusions and Suggestions (A Cycle within a Circle of Assessment)

Assessment plays a central role in teaching and learning. It should give an accurate picture of students' knowledge and skills in the subject area or domain being tested. Accurate achievement data are very important for planning curriculum and instruction and for program evaluation. Test scores that overestimate or underestimate students' actual knowledge and skills can not serve these important purposes.

Alternative assessment seems to offer an excellent means of creating ethical and valid assessment practices, analogous to the Greek books of life. Through their work on alternative assessment instruments, students address ontological authenticity, that is, they answer the questions of academic and professional identity. Not only do they create their professional identities, but they also consolidate their academic identities (e.g. portfolio assessment).

However, teachers and assessors do not have a completely passive role, either in the teaching or the assessing process. In order to put the sociopolitical aspects of ethics in language testing in a proper context, there should be a three way conversation between the test-taker, the test developer, and the decision maker both in language curriculum development and assessment. The interaction among language testing, applied linguistics, and psychometrics, has helped the field of language testing to mature in recent decades. However, great care should be exercised in the application of the principles of the other fields to language testing in order to avoid misleading

conclusion, unfair & unethical aspects in educational assessment.

Politicians, educators, bureaucrats, and so many other sectors in the society are in the position to make decisions. Language testers' responsibility in this case is limited to only giving guidelines to the decision makers to assure quality in educational assessment.

Examining impact and evolved power relation, in particular, will show whether the teaching and learning have proceeded ethically and validity. This thinking suggests that power relations are not inherently evil and need not be escaped. They are, rather, part of the human condition, and educators must work to develop practices i.e. practices of assessment- that maximize the mobility, the reciprocity and hence validity of the test. There is also work relevant to assessment designed to make assessment, more open to the active participation of stakeholders other than testers. Therefore, we will come up to a diagram of interactive relationship (see Figure.1) among tests, assessment and language curriculum components i.e. A cycle within a circle of assessment.

Examining educational assessment based on the validity framework developed will show whether the teaching and learning have proceeded ethically and validly. This thinking suggests educators must work to develop practices of assessment to maximize the ethics and validity of assessment inferences. According to McNamara and Deane (1995) by using multiple assessment instruments, traditional measures of alternative assessments, traditional measures & alternative assessment measures, rational decisions can be made about students' lives & curriculum revision policies for revising the program views of needs, goals, objectives, materials, teaching strategies, overall course evaluation & ethical consideration which can be a great source of important and valid information.

References

- Bachman, L. F., & Eignor, D. R. (1997). Recent Advances in Quantitative Test Analysis, in Clapham, C. M., & Corson, D. (eds.), *Language Testing and Assessment. Encyclopedia of Language and Education*, Vol. 7
- Bennett, C. I. (1995). Preparing teachers for cultural diversity and national standards of academic excellence. *Journal of Teacher Education*, 46(4), 259-266. <http://dx.doi.org/10.1177/0022487195046004004>
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum, & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston: Kluwer Academic Publishers. pp. 3-29
- Black, P., & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal*, 16, 249-261. <http://dx.doi.org/10.1080/09585170500136218>
- Brown, J. D., & Hudson, T. (1998). Deciding Which Language Test to Use. *Kanda Institute of Foreign Languages Academic Journal*, 7, 1-12
- Brown, J. D., & Hudson, T. (1998). The Alternatives in Language Testing: Advantages and Disadvantages. *University of Hawaii Working Papers in ESL*, 16 (2), 79-103
- Colnerud, G. (1997). Ethical conflicts in teaching. *Teaching and Teacher Education*, 13, 627-635. [http://dx.doi.org/10.1016/S0742-051X\(97\)80005-4](http://dx.doi.org/10.1016/S0742-051X(97)80005-4)
- Darling-Hammond, L. (1994). Performance-Based Assessment and Educational Equity. *Harvard Educational Review*, 64 (1), 5-30
- Davies, A. (2004). Introduction: Language testing and the golden rule. *Language Assessment Quarterly*, 1 (2&3), 97-107. http://dx.doi.org/10.1207/s15434311laq12&3_2
- Davies, A. (2005). *A glossary of applied linguistics*. Mahwah, N. J.: Lawrence Erlbaum Associates.
- Fortin, P. (1996). Address to a Special Joint Committee on a Code of Conduct of the Canadian Parliament. [Online] Available: http://www.parl.gc.ca/35/Archives/committees352/sjcc/evidence/03_96-05-08/sjcc03_blk101.html (March 30, 2009)
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational Assessment*. Washington. D. C: The Falmer Press.
- Gipps, C. V. (1995). *Beyond Testing: Towards a Theory of Educational Assessment*. London: The Falmer Press.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Hamp-Lyons, L. (1989). Language testing and ethics. *Prospect*, 5 (1), 7-15
- Hamp-Lyons, L. (2001). Ethics, fairness(es), and developments in language testing. In Catherine Elder (Ed.), *Experimenting with uncertainty: Essays in honour of Alan Davies*. Cambridge: Cambridge University Press. pp. 222-227

- Hamp-Lyons, L., & Lynch, B. K. (1998). Perspectives on Validity: a Historical Analysis of Language Testing Conference Abstracts. In Kunnan, A.J. (ed.), *Validation in Language Assessment*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Howe, K. R. (1994). Standards, assessment, and equality of educational opportunity. *Educational Researcher*, 23, 27-33. <http://dx.doi.org/10.2307/1176860>
- Jia, Y. (2009). Ethical standards for language testing professionals: An introduction to five major codes. *JALT Testing & Evaluation SIG Newsletter*, 13 (2), 2-8
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kunnan, A. J. (2000). *Studies in language testing 9: Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida*. Cambridge: Cambridge University Press.
- Longstaff, S. (1994). Developing Codes of Ethics: Why codes of ethics fail. [Online] Available: <http://www.ethics.org.au/about-ethics/ethics-centre-articles/ethics-subjects/codes-ofethics/article0124a.html> (April 30, 2009)
- Lynch, B. K. (1997). In Search of the Ethical Test. *Language Testing*, 14 (3), 315-527
- Lynch, B., & Shaw, P. (2005). Portfolios, Power, and Ethics. *TESOL Quarterly*, 39, 263 – 297. <http://dx.doi.org/10.2307/3588311>
- McNamara, T., & Deane, D. (1995). Self-assessment activities: Toward autonomy in language learning. *TESOL Journal*, 5(1), 17-21
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement*. New York: Macmillan. pp. 13-103
- Moss, P. A. (1994). Can There Be Validity without Reliability? *Educational Researcher*, 23(2), 5-12. <http://dx.doi.org/10.2307/1176218>
- Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25 (1), 20-28. <http://dx.doi.org/10.2307/1176724>
- Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing Second Language Performance Assessments*. University of Hawai'i, Honolulu.
- Patton, M. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks. CA: Sage.
- Shohamy, E. (1991). Connecting testing and learning in the classroom and on the program level. In J. K. Phillips (Ed.), *Building bridges and making connections*. Northeast Conference Reports So Burlington. VT: Northeast Conference. pp. 154-178
- Shohamy, E. (1997). Second Language Assessment. In Tucker, G. R., & Corson, D. (Ed), *Second Language Education. Encyclopedia of Language and Education*, Vol. 4. Kluwer, Dordrecht.
- Shohamy, E. (1997). Testing Methods, Testing Consequences: Are They Ethical? Are They Fair? *Language Testing*. Dordrecht, Netherlands: Kluwer Academic.
- Spolsky, B. (1997). The Ethics of Gatekeeping Tests: What Have We Learned in a Hundred Years? *Language Testing*, 14 (3), 242-247
- Strike, K. (1990). Teaching ethics to teachers: what the curriculum should be about. *Teaching and Teacher Education*, 6 (1), 42-53. [http://dx.doi.org/10.1016/0742-051X\(90\)90006-Q](http://dx.doi.org/10.1016/0742-051X(90)90006-Q)
- Wilson, L. (1995). Principles for evaluating language development. *NLLIA Network Notes*, 8, 3
- Wolf, D. et al. (1991). To Use Their Minds Well: Investigating New Forms of Student Assessment. In G. Grant (ed.), *Review of Research in Education*, 17. Washington, DC: American Educational Research Association.

Table 1. A Framework for Analyzing Ethical Assessment

Ethics/Validity Category	Definition	Extension to alternative assessment
Fairness (Kunnan, 2000)	Everyone receives the same assessment procedure, and the results of that assessment are interpreted in the same way for all candidates; all participants in a particular assessment setting have been taken into account.	Are the perspectives of all affected participants in the educational assessment process being taken into account?
Equity (Gipps, 1995)	Equal access to curriculum and learning _ Having an explicit account of the constructs being assessed and criteria for assessment.	What knowledge is assessed and equated with achievement?
Ontological Authenticity_ Authenticity Criteria (Guba & Lincoln, 1989)	Participants gain and use information as a result of the assessment process.	Do the participants in the assessment process establish a meaningful identity, a sense of who they are?
Cross referential Authenticity_ Educative Authenticity (Guba & Lincoln, 1989)	Stakeholders and participants gain an understanding of the perspectives and meaning constructions of those outside their own group.	Are the participants in the assessment process able to gain an improved understanding of the perspectives outside their own group?
Impact_ Consequential Validity (Messick, 1989)	A consideration of the values and goals that inform assessment interpretations, and the consequences of interpreting and using assessment information; the degree to which something happens as a result of the assessment process.	What is actually done as a result of assessment?
Evolved power relations_ Tactical Authenticity (Guba & Lincoln, 1989; Shohamy, 1997, and Spolsky, 1997)	The degree to which participants are empowered to carry out the changes that are made possible through the assessment process.	Do the participants change the way they relate to each other and to themselves?

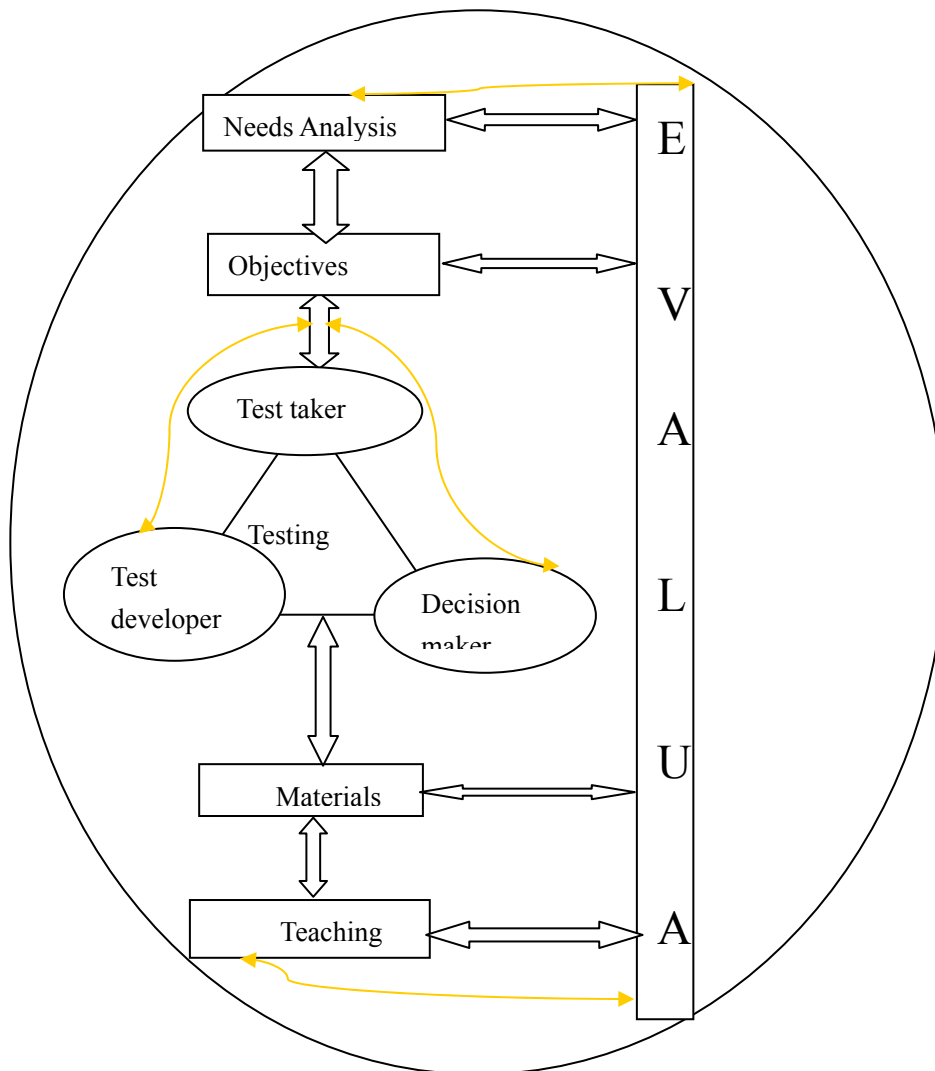


Figure 1. A Cycle within a Circle of Assessment

Appendix A (Teacher Interviews)

Principles for Assessment and Reporting

Full Name: _____

Nationality: _____

Age: _____

Gender: _____

University Degree _____

Teaching/Test Development Experience _____ Ys

- 1) Parents are entitled to continuing, quality information regarding their children's education through a variety of mechanisms.
- 2) Any form of assessment should be integral to the curriculum and designed to inform, support and improve learning outcomes.
- 3) Assessment and reporting processes should make provision for parent and student input about teaching and learning.
- 4) Parents and their organizations must have an active role in developing and implementing assessment and reporting policies and processes at the school, the system, the state and the nation levels.
- 5) Schools, systems and Governments must make explicit and public the purposes for which they wish to collect assessment data.
- 6) Assessment data must not be used for the purpose of establishing and publishing competitive judgments about

schools, systems, etc.

- 7) Parents must be informed by all those who seek such data about student performance of the uses to which such information will be put.
- 8) Data collected from students in schools should be used in accordance with its stated purposes. Any other subsequent uses should be specifically negotiated.
- 9) Individual student assessments are confidential to the student, his/her parents and appropriate school staff.
- 10) Parents have the right to withdraw their children from specific system, statewide and national testing.
- 11) Assessment data for statewide or national purposes should be collected by statistically valid, light sampling procedures only.
- 12) Appropriate appeal mechanisms should be established and made public to protect the rights of students and parents in matters of student assessment and reporting at the school, state and national level.

Appendix B (Teacher Interviews)

Principles for Evaluating Language Development (Wilson, 1995: 3)

- 1) What and how we evaluate must be Consistent with what we value about language and language development.
- 2) Evaluation strategies must evaluate what they set out to evaluate (e.g., to check spelling by having the student select the correct spelling of a word from a list of alternative spellings tells nothing about how that student would spell that same word in a writing situation.
- 3) The purpose of evaluation is to inform the learners, teachers and parents. It thus must be descriptive.
- 4) Learning is ongoing, therefore evaluation must be ongoing.
- 5) Language is learned in use, language use is context related. Language evaluation therefore must occur in authentic contexts.
- 6) As language use relates to language experience, an externally administered evaluation procedure may evaluate the teaching program more than it evaluates the student's language competence.
- 7) Language learning is developmental. It involves experimentation and approximation. Language assessment must reveal the student's developing understandings.
- 8) Learning to control the surface features of a language does not necessarily occur at the same rate as learning to control the functions of a language.
- 9) Student self evaluation is an important part of the evaluation process.