

Language Testing

<http://ltj.sagepub.com>

In search of the ethical test

Brian K. Lynch

Language Testing 1997; 14; 315

DOI: 10.1177/026553229701400308

The online version of this article can be found at:
<http://ltj.sagepub.com/cgi/content/abstract/14/3/315>

Published by:

 SAGE Publications

<http://www.sagepublications.com>

Additional services and information for *Language Testing* can be found at:

Email Alerts: <http://ltj.sagepub.com/cgi/alerts>

Subscriptions: <http://ltj.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations (this article cites 13 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
<http://ltj.sagepub.com/cgi/content/refs/14/3/315>

In search of the ethical test

Brian K. Lynch *University of Melbourne*

The central question to be addressed here is whether any test can be defended as ethical, or moral. Ethicality is defined in terms of issues such as harm, consent, confidentiality of data and fairness. Frameworks for determining equity of educational opportunity are presented and discussed. A statewide assessment project in Victoria, Australia (the Learning Assessment Project) is then examined in relation to these concerns, and the possibility of more ethical approaches to testing is considered.

I Definition of test

Davies (1990) has defined a test quite clearly as having the central function of discrimination, a word that on the surface at least connotes a lack of morality. However, for language testers, discrimination can be a positive concept, something we strive for, at least in certain types of testing. As Hamp-Lyons (1989:13) has phrased it, we try to be 'discriminating in our discriminations'. While Hamp-Lyons would include discrimination as one of the things tests are meant to do, Davies (1990: 17) uses it as the defining characteristic for all tests:

Here we are indicating what is, in our view, one of the fundamental purposes of a test, that a test is intended above all to clarify the differences in the matter under test, in what is being tested (proficiency, aptitude, achievement) among the candidates.

Davies holds firm to this definition when discussing 'criterion-referenced testing' (CRT), maintaining that a CRT is either a special use of a 'norm-referenced test' (NRT) or should be considered as an 'exercise', not a 'test'.

II Morality of discrimination

Whether or not Davies intends this as the limiting definition for all tests, it raises an important issue for the topic being addressed: that is, the morality of tests that *are* used to discriminate between test-takers. Is test discrimination an essentially immoral function? Wood (1991: 83), in his survey of testing and assessment research, references the claim by certain educational philosophers that such (norm-referenced) testing was 'a violation of students' civil rights, inasmuch

as students were ranked publicly and might therefore feel humiliated' and reminds us that, in the late 1970s and early 1980s, there were court cases in the USA where test-takers did, in fact, claim such damage as a result of taking minimal competency tests. Wood (1991: 83) notes, ironically, that these tests had been 'built on the most impeccable criterion-referenced principles'. It would seem that even tests that are not designed to rank order individuals can result in perceived harm to the test-taker.

A discussion 'thread' on an electronic listserv (vista@ets.org) has debated the question of inherent moral problems for tests, with one participant declaring: 'I think that it is quite obvious that testing is linked to inequality, whether conceived of as "equity", that is "fairness" or "equal opportunity" or as referring to human rights, such as health care . . .' and that 'It is very possible that the function of testing is to legitimate inequality' and, more specifically, that 'we might conclude that built into the test itself . . . is inequality; i.e. the bell curve assumption makes the test necessarily represent people as unequal' (Garrison, 1995). There were various replies to this provocative position; in particular, a counterclaim that the assumption of inequality in the normal distribution was irrelevant – all score distributions assume such variability. This latter argument assumes the 'measurement' perspective that whatever is being tested does 'vary' from individual to individual; otherwise there would be no purpose in trying to test.

I would like to refocus this debate on the moral question: does the test, or the act of testing, create the inequality? If yes, as Garrison argues, then tests, at least the traditional, NRT variety, are inherently immoral. If no, what is the relationship of the test to that inequality? If a test merely describes or identifies differences in ability that already exist, is it in any way morally responsible for those differences, for that inequality? Garrison's claim that tests can serve to 'legitimate inequality' seems important to address in this light. For some, if a test is free of obvious technical flaws and, perhaps, less obvious characteristics such as cultural bias, then it is a 'good' test. These characteristics have been articulated as part of various sets of standards and codes of practice for testing (see Stansfield, 1993; Alderson, Clapham and Wall, 1995; TFTS, 1995). Some of these standards also include attention to the broad range of social and legal concerns that accompany test use which others (e.g., Hamp-Lyons, 1989; Shohamy, 1993) have argued for in relation to language testing. This can also be seen as a part of the larger discussion of 'washback' (Alderson and Wall, 1993; Alderson and Hamp-Lyons, 1996; Messick, 1996; Shohamy, Donitsa-Schmidt and Ferman, 1996; Wall, 1996; Watanabe, 1996), as well as 'consequential' (Messick, 1989;

1994) and 'systemic' validity (Fredericksen and Collins, 1989). From this perspective, the responsibility of the test developer and, presumably, the testing researcher, does not necessarily end with the verification of high reliability coefficients and a statistically confirmed model of the trait being assessed, and Garrison's position that the normal distribution inherently dooms tests as immoral is less easily cast aside. Part of the goodness of a test, part of its validity, would be a consideration for the consequences of its use. If those consequences entail unethical or immoral outcomes, then the test and those associated with it are accountable, albeit to greater or lesser degrees.

III Morality of test use: ethical issues

But what would constitute unethical or immoral outcomes for a language test? The ethical issues that have been identified for the conduct of research, in general, may be instructive here. For example, Punch (1994) mentions the issues of consent, deception, privacy and confidentiality.

1 Consent

In terms of 'consent' it could be argued that individuals who are being tested rarely have a choice in the matter. However, formulated as 'informed consent', the issue becomes whether they have been clearly told that they are being tested, what they are being tested for and what will be done with the results of the testing. This may not be a matter for much concern in the context of language testing, except where language test data are used for research purposes that are not explicitly articulated at the time of testing.

2 Deception

The issue of deception is problematic when test-takers are engaged in indirect measures of the ability being tested. Here the notion of face validity can be seen as a moral issue, in part. Is it morally deceptive to have an individual engaged in an activity that he or she cannot clearly see as being directly related to the ability supposedly being tested? For example, if the individual has never seen a cloze test or C-test, much less read the literature on the construct validity of these procedures, is it moral to make decisions about his or her language ability based on how well these tasks have been performed? Have the test makers and test givers (as well as test interpreters) been deceptive, or is it simply a matter of telling the test-takers at some point that these tasks are known to be valid measures of their ability? Internal

to some tests there may be deception. In the case of the ubiquitous multiple-choice test format, the 'distractors' (the wrong answers) can be thought of as deceptive by definition. In more general research contexts, the deception issue is portrayed as one of degrees of acceptability; for example, failing to be 'brutally' honest and direct with people is seen as more acceptable than stealing documents or deliberately lying to them (Punch, 1994).

3 *Privacy and confidentiality*

In terms of the issues of privacy and confidentiality, we return to the concern for public humiliation as a result of being tested. Hamp-Lyons (1989: 13) identified the core principle of ethical language testing as 'no test-taker shall be harmed by the test'. Have we harmed a test-taker by reporting results from a test that place him or her at the bottom of the pile: that deny him or her entrance to university; that prevent him or her from accessing certain social and economic resources? In that strict, perhaps naive, sense of harm, it may be that no test is moral; that is, that tests can always be used to disadvantage, to cause harm to some. We return to the argument that it is not the act of testing that is to blame. Tests are seen as efficient means for identifying differences in ability that exist. They do not create those differences. If there are scarce resources to be distributed, then what better way, what fairer, more moral way than by merit? To those with the demonstrated ability (via tests) go the spoils.

4 *Tests and fairness*

In that sense, perhaps the potential for moral tests lies in giving everyone a fair chance to demonstrate his or her ability. There has been a great deal of research on *item bias* and *differential item functioning* in the educational literature, as well as a certain amount of attention in the language testing field. What results from a consideration of this research is a respect for the complexities of the issue, and this topic is taken up in greater detail elsewhere (see Elder, this issue).

5 *Test decisions/consequences and fairness*

However, if we are concerned about the broader sense of test validity and, therefore, test ethicality, we cannot limit our attention to assuring that tests give everyone a fair chance to demonstrate their abilities. How, then, do we decide what constitutes fairness in the consequences of language testing decisions? In the general educational context, Howe (1994: 27, emphasis added) maintains that

There may be ways of justifying the decisions made on the basis of differential test performance that are consistent with the requirements of equality . . . so long as individuals are afforded equal opportunities to obtain an education, inequalities in educational results are *morally permissible*.

IV Equal opportunity to learn: Howe's frameworks

There are different conceptions of how we arrive at equal opportunity to learn. Howe (1994) outlines three frameworks: the formal, the compensatory and the democratic.

1 Formal

The formal framework requires 'the absence of barriers to access based on "morally irrelevant" characteristics such as race or gender' (Howe, 1994: 27). Characteristics such as academic performance, from this perspective, can be used to discriminate among individuals for the distribution of educational opportunities.

2 Compensatory

The compensatory framework also requires the absence of morally irrelevant barriers to access, but further maintains that 'it is unjust to hold students who have not had adequate educational opportunities responsible for the same level of educational performance as students who have had adequate opportunities' (Howe, 1994: 28). There are two varieties of the compensatory framework: the utilitarian, where the principle of maximizing benefits (e.g., economic benefits) is central; and the egalitarian, epitomized by the philosophy of John Rawls and adding the notion that inequality should be minimized by redistribution from the more advantaged to the less advantaged.

3 Democratic

The third framework, which Howe himself identifies as holding the most moral promise, is the democratic. From this perspective, most, if not all, educational standards and practices are currently found to be lacking from a moral point of view; that is, the other two frameworks simply do not work. The democratic model provides a viable alternative by requiring the inclusion of 'voices that have historically been excluded in negotiating educational goods *worth wanting*. It also requires changing schooling accordingly' (Howe, 1994: 30, emphasis in original). This answers the criticism that the compensatory/egalitarian model assumes a white male ideal. The

democratic framework also replaces the goals of economic efficiency and mastery of the traditional domains of knowledge with 'fostering general habits of mind that render individuals capable of and disposed to gathering and evaluating information' (Howe, 1994: 30); in other words, to ensure that each individual can participate effectively in the political processes of society and, on the personal level, can achieve a sense of self-esteem.

Howe ends his discussion by first claiming that assessment, even large-scale testing to monitor national educational outcomes, is possible under the democratic framework. However, he cautions that the attempt to use assessment and standards as the solution to 'opportunity to learn' problems is like 'suggesting the way to end world hunger is to first develop more rigorous standards of nutrition and then provide physicians with more precise means of measuring ratios of muscle-to-fat' (Howe, 1994: 31).

V The LAP as an example of ethical issues for testing

With the issues of ethicality and the concern for the relationship of testing to educational opportunity as the background, I will now turn to an example of a testing project that has recently been implemented in the state of Victoria, Australia. In 1995 the state Ministry of Education initiated the Learning Assessment Project (LAP), a statewide testing of all Grade 3 and Grade 5 students in English language, mathematics and one other area from the recently established Curriculum and Standards Framework (CSF). The question to be addressed here is the morality of the LAP. As with most questions of morality, the answer depends on who is asked. The government agency that sanctioned the LAP (the Board of Studies) claims that it is designed to provide information for parents and teachers that will complement the existing school assessment procedures. In the 1996 brochure for parents, the Board of Studies says: 'It [the LAP] assesses your child's learning progress and provides information on areas where he or she could benefit from greater challenges or extra help'. It goes on to say the LAP will be used to develop support for children in the identified areas of need and makes certain strong claims about the ethical issues of confidentiality and potential for harm:

Your child's results are not made public. They are confidential to you and the school. Results are not used to rank your child or to make comparisons between teachers or schools ... Teachers will get confidential information about achievement in their classes and Principals will benefit from information which will help overall planning of curriculum programs. For example, after the results of the LAP last year, some schools started new programs and others shifted resources to match newly identified needs (Board of Studies, 1996, emphasis in original).

In an effort to obtain a fuller picture of this assessment project, I conducted a small study of the 1996 administration of the LAP at one Victorian school. The results indicated that the perspectives of parents, teachers and the school principal concerning the LAP were significantly different from those of the Board of Studies. In particular, the morality of the test was called into question. It should be pointed out that the school which informed the study proved to be one of those where a majority of the parents and teachers were (and continue to be) opposed to the LAP. In a survey conducted with the parents of Grades 3 and 5 children before its 1996 administration, 81% of those who responded (admittedly, less than half the total number of relevant families) said that they were primarily opposed to the LAP. Only one family indicated they were in favour of the test. The teachers at this school were unanimously opposed to the LAP and were prepared to boycott its administration. As it turned out, only two families requested that their children sit the LAP (there were 40 Grade 5 children and 53 Grade 3 children at this school), and these were given a special administration that did not involve school staff members.

Why were the parents and teachers at this school opposed to this assessment? A majority of the families who responded to the questionnaire and with whom I conducted interviews expressed the perception that the LAP was a 'waste of time and money'. Specific comments that revealed a sense that the test was questionable from a moral point of view included: 'It doesn't treat children as individuals'; 'The LAP is degrading to the children'. Responses from interviews with parents helped to clarify these perceptions. One parent said that she preferred the assessment done by the school's teachers to that of the LAP because the former 'is particularized to the child's progress . . . If you have external assessment, grading children against the State, it's a "comparative thinking" – the child will be judged as better or less than others rather than the child improving upon themselves . . .' Another parent, who had seen example problems from the LAP, commented that one of the problems on the test was too complicated and confusing for even the parents to answer. One of the teachers interviewed gave other insights into the question of why the LAP might be 'degrading' to children – the format of the test was unfamiliar and would not allow them to demonstrate what they know:

It seems very alien to other assessments the children have had; they would find it difficult to figure the tasks out . . . This type of evaluation is not relevant, in terms of how the children approach tasks; not valid to assess progress. For that you need observation, continuous assessment, not a 'one-off' like the LAP. For example, . . . the children may understand the process, but get the wrong answer.

Both parents and teachers also expressed the feeling that this form of testing was inappropriate for children as young as those in Grade 3.

As evidenced in one parent's comments, contrary to the Board of Studies' contention that the LAP would not be used to rank or compare children, teachers or schools, many parents felt that it would be used in this way. At a school committee meeting of parents, several voiced the opinion that the LAP was already being used as the basis for competition between schools – competition for funding and for attracting students. As one parent said, 'Coming from a government that you know is closing down schools, you can't help but be opposed'. In an interview, another parent said: 'There are fears of schools being closed – the LAP could be used to close schools. Schools are now competing for enrolments, and may use the LAP results to promote one school over another'.

Another issue that was raised concerned the morality of the LAP in relation to children from non-English-speaking backgrounds. In an interview, one parent commented that the LAP seems to be

basically a test designed for white Anglo-Saxon children, but will be given to others, for example non-English-speaking background children, and this will disadvantage them. [Our school] is a diverse school – the LAP is not appropriate; maybe it would be in the suburbs [wealthier, more homogeneously white Anglo-Saxon].

In fact, the Board of Studies was aware of this potential criticism from the start of the project, and allows for the exemption of children who are from non-English-speaking backgrounds (NESB). The problem with such exemptions seems to be that, if the LAP is used to identify areas of need for individuals and schools, then the needs of NESB children could go largely undetected.

I also attempted to gain an understanding of the perspective of the parents who might be in favour of the LAP. In an interview with a parent from the one 'pro LAP' family who responded to the questionnaire, the purpose of the LAP was described as

[providing a] benchmark for government to assess priorities . . . low scores would send a signal back to the government that the school needs review . . . to either close it or put in additional resources . . . This makes teachers and schools accountable to the State and its set curriculum . . . From my point of view, we're all measured against performance standards; it's not a bad idea.

At a public lecture held at another school in the area, a spokesperson for the LAP (a university educator and school parent) argued that parents want and need more information about the academic progress of their children than teachers can give them (although the importance of teacher assessment information was acknowledged). The LAP was characterized as providing evaluative information on the state's set curriculum, as well as giving parents a better idea of whether their

child's progress is 'good enough' or not. This spokesperson exemplified this need by challenging the idea that reports from the teacher such as 'the child can count to 100' were sufficient. As parents, it was argued, we 'don't know if this is a great achievement or not' – parents need 'benchmarked reporting for true accountability'.

The morality of the LAP, then, is a tricky issue. The results from a study of one school indicate that the issue of competition for resources and the potential harm from ranking children are at the heart of the question. Different values have surfaced, with some parents feeling that such competitive testing is 'degrading' to the children and others feeling that it results in important information for the schools and themselves as parents concerned about their child's progress. Most importantly, however, none of the parents or teachers in this study expressed the idea that all assessment was a bad thing. In the case of the LAP, it appears that the 'standardized', 'external', 'imposed-by-the government' nature of the assessment is what makes it morally questionable. A further study of the consequential validity of the LAP has been planned (Elder and Lynch, 1996).

VI More ethical ways of assessing?

This leads to the question of whether there are more ethical ways of testing or assessing than are represented by standardized tests such as the LAP. How can the ethicality of these alternative forms of assessment be judged, beyond applying Howe's 'democratic' criterion? The Australian Council of State School Organizations (ACSSO) and the Australian Parents' Council (APC) recently conducted an investigation of this question, resulting in a set of 12 principles for assessment and reporting (Victorian Council of School Organizations, 1996 – see Appendix 1). At the heart of these principles is the notion that, in order for assessment to be moral, parents and students should be actively involved and assessment data should not be used to make competitive judgements about schools or systems across the states and territories. Also mentioned is the idea that good assessment should be 'integral to the curriculum'. This relates to the previously mentioned notion of washback or 'backwash', which, as Hamp-Lyons (1989: 13) has pointed out, is the other thing (besides discrimination) that tests do best – 'reveal their influence back into the curriculum'.

A parallel set of principles for assessing language, in particular, was developed by educator Lorraine Wilson and published in the newsletter of the National Languages and Literacy Institute of Australia's Adult and Child Literacy and ESL Research Networks (see Appendix 2). When we apply Wilson's principles to LAP, it does not fare well. This is not surprising, given that Wilson's perspective is

clearly in favour of alternative assessment over traditional testing. Although there would be disagreement over whether the LAP conforms to Wilson's principles 1, 2 and 3, it is clear that even those in favour of the LAP would not claim that it is a form of 'ongoing evaluation' (principle 4), or that it assesses language use 'in authentic context' (principle 5). It clearly does not include student self-assessment (principle 9), is 'an externally administered evaluation procedure' (principle 6), and it has been criticized for focusing too much on 'the surface features of a language' (principle 8), as well as being potentially insensitive to 'the student's developing understandings' (principle 7). These are principles that are more in line with alternative forms of assessment, and tests like the LAP are not usually designed to respond to these sorts of concerns.

VII Conclusion

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 the more recent approaches to validity, ethical concerns of potential harm and fairness need to be examined with more than test internal estimations of reliability and bias. Alternative forms of assessment may indeed help us to address the goals of ethicality.

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). These improvements may be more time-consuming and less efficient and may involve goals that are difficult to capture with precise instruments of measurement. However, in the words of Kenneth Howe (1994: 31), 'to reject it for these reasons places the quest for accurate measurement – and control – above the quest for educationally and morally defensible policies'. The search for ethical assessment continues.

VIII References

- Alderson, J.C., Clapham, C. and Wall, D. 1995: *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alderson, J.C. and Hamp-Lyons, L. 1996: TOEFL preparation courses: a study of washback. *Language Testing* 13, 280–97.
- Alderson, J.C. and Wall, D. 1993: Does washback exist? *Applied Linguistics* 14, 115–29.
- Board of Studies (Victoria, Australia) 1996: *Learning Assessment Project*

- 1996: *information for parents of children in years 3 and 5*. Carlton: Board of Studies.
- Davies, A.** 1990: *Principles of language testing*. Oxford: Blackwell.
- Elder, C. and Lynch, B.K.** 1996: Investigating test impact: public perceptions of the LAP tests in Victoria. *Language Testing Update* 20, 53–54.
- Frederiksen, J.R. and Collins, A.** 1989: A systems approach to educational testing. *Educational Researcher* 18, 27–32.
- Garrison, M.** 1995: [V056NS88@ubvms.cc.buffalo.edu]. BIO. In VISTA (discussion list) [VISTA@ets.org], 6 November.
- Hamp-Lyons, L.** 1989: Language testing and ethics. *Prospect* 5, 7–15.
- Howe, K.R.** 1994: Standards, assessment, and equality of educational opportunity. *Educational Researcher* 23, 27–33.
- Lynch, B.K.** 1997: The ethical potential of alternative assessment. Paper presented at the 31st annual convention of TESOL, Orlando, FL.
- Messick, S.** 1989: Validity. In Linn, R.L., editor, *Educational measurement* (3rd edn), Washington, DC: The American Council on Education and the National Council on Measurement in Education, 13–103.
- 1994: The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23, 13–23.
- 1996: Validity and washback in language testing. *Language Testing* 13, 241–56.
- Punch, M.** 1994: Politics and ethics in qualitative research. In Denzin, N.K. and Lincoln, Y.S., editors, *Handbook of qualitative research*, Thousand Oaks, CA: Sage, 83–97.
- Shohamy, E.** 1993: The power of tests: the impact of language tests on teaching and learning. *NFLC Occasional Papers* June, 1–19.
- Shohamy, E., Donitsa-Schmidt, S. and Ferman, I.** 1996: Test impact revisited: washback effect over time. *Language Testing* 13, 298–317.
- Stansfield, C.W.** 1993: Ethics, standards and professionalism in language testing. *Issues in Applied Linguistics* 4, 189–205.
- Task Force on Testing Standards (TFTS)** 1995: *Report of the Task Force on Testing Standards (TFTS) to the International Language Testing Association*. Melbourne: ILTA (also available in the ERIC Microfiche Document Reproduction System, file no. ED390268).
- Victorian Council of School Organizations** 1996: Australian parents agree on principles for assessment and reporting. *School Bell* 50, 2.
- Wall, D.** 1996: Introducing new tests into traditional systems: insights from general education and from innovation theory. *Language Testing* 13, 334–54.
- Watanabe, Y.** 1996: Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing* 13, 318–33.
- Wilson, L.** 1995: Principles for evaluating language development. *NLLIA Network Notes* 8, 3.
- Wood, R.** 1991: *Assessment and testing: a survey of research*. Cambridge: University of Cambridge Local Examinations Syndicate, University of Cambridge Press.

Appendix 1 Principles for assessment and reporting

(Formulated by the Australian Council of State School Organizations and the Australian Parents' Council)

- 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, State and Federal, 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 judgements about schools, systems, states or territories.
- 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 2 Principles for evaluating language development

- 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.

(Wilson, 1995: 3)