Developments in Language Testing with the Focus on Ethics

Alma Cenkaj Vladi

Teacher of English, Foreign Language Centre, Academy of Albanian Armed Forces, Rr. 'Haxhi Xhediku' Nr. 32, PO code 1032

Abstract
Language tests have become powerful tools, because they are used to measure the success of individuals in different aspects of life. Despite their influence on the lives of individuals taking them, only in the last decades have language theorists started to raise questions of high sensitivity. Tests were considered as purely linguistic acts, therefore very little attention was paid to the social dimension of language as the most important medium of communication among humans. With the advent of integrative testing, and later communicative language testing, linguists became more conscious of the social impact that such tests have. The following article brings a general overview of the history of language testing, the questions that were raised in each phase of theoretical development, as well as further exploration of ethics in language testing.
Keywords: psychometric-structuralist, gatekeeper, communicative approach, ethical approach, alternative assessment

Introduction
1.1 General overview
People have always been interested in testing their cognitive abilities, from entry tests for university admission, which are important life decisions, to much more mundane everyday tasks. Language testing is no exception. Linguists are now aware that tests are a serious business in the lives of individuals who take them, because institutions, education or other fields, have very clear objectives about using them. Sometimes, language tests have become an important tool, unfortunately not always in the hands of language experts. They are sometimes used as screening devices for reasons other than testing language knowledge. Therefore, the testing system must have such impact in the back of their minds.

Examples of using tests for reasons other than linguistic thrive in history, dating from biblical times with the infamous 'shibboleth' test, as one of the earliest cases recorded in history. (McNamara and Roever, 2006). These authors mention that after a struggle, members of the losing tribe tried to merge with the winning tribe population in order to ensure survival. These two tribes were culturally and linguistically similar. Both spoke the same language varieties with very small differences. One of these differences was used by fighters of the winning tribe to identify and destroy the losers. The word 'shibboleth' was pronounced with [s] by the losers of the war, while winners pronounced it with [ʃ]. All persons who did not pass the test were killed. (McNamara, 2000)

One could say that those times are long gone, but looking not so far back in time, we see use of language tests as a gatekeeper for Hebrew immigrants fleeing atrocities of Second World War only to face refusal in Australia. The immigration officer had a duty to exclude all immigrants who did not belong to the British Isles and it was realized through dictation in different languages. A fatal case, a Hungarian Hebrew who had applied for a visa to Australia to escape the terrible regime of Hitler in Europe and had excellent knowledge in several languages, missed dictation in Gaelic language. Refusal and return to Europe was fatal. (McNamara, 2000)

The two examples given above are not the only ones, but enough to create the landscape of the real and comprehensive testing process, which often becomes a party to important policies and influences the lives of people and such policies are sometimes morally controversial.

The situation today is not as dramatic as it was almost 50 years ago. This is in part thanks to developments in language testing which brought into the picture the social dimension of language testing. It was only then that these issues became apparent, so language testers steered discussions towards ethics and social implications on the lives of people. From behaviourism to structuralism, to psychometrics, to communicative language testing, it became clear that language testing is not like testing other cognitive abilities (McNamara and Roever, 2006).

This article will present a general overview of the main testing theories through a historical timeline, which will provide a better understanding of such issues.

1.2 Historical timeline of language testing theories
The history of language testing is closely related to the historical development of theories of linguistics in general. Bernard Spolsky (1976) distinguishes three historical periods of modern language testing, pre-scientific, psychometric-structuralist, and psycholinguistic-sociolinguistic. In a subsequent article, he says that his view

was too simplistic and lacked some historical milestones, which he tried to improve in his book dedicated to this field. (Spolsky, 2007). Morrow (1979, as cited in Fulcher, 2000) has baptized such periods as the "Garden of Eden", "Lawn of Tears" and "The Promised Land". While Shohamy (1996) has identified five stages of development: the discrete point era, the integrative era, the communicative era, the performance testing era and the alternative assessment era. All phases are a reflection of philosophy, viewpoints and trends of their time.

Below we will take a look on each of these stages to create a clear picture of theoretical foundations of language testing in general, but especially how social context came into the picture with the development of communicative language testing and proposals by alternative assessment.

1.2.1 The First phase

Testing is deeply rooted in the pre-scientific phase, which dates as back in time as 201.bc. with tests of Confucian Doctrine (Spolsky, 2005), until beginning of last century, when some basic concepts started to evolve. However, this first phase was characterised by subjective tests with teachers being the primary (and in most cases the only) assessors of the linguistic competence of their students; clearly, no reliability or validity issues were raised. The experienced teachers would score written essays or examine the students orally (although with great limitations), and their judgment on the competence was accepted and relied on. The overall view of this phase was that '*everything was going well*' (Weir, 2005, p.5).

1.2.2 The Psychometric-Structuralist phase

Although the second phase of testing historical developments took its full concrete shape from the middle of the last century, its roots were laid in the 19th century, with discussions on the measurement of human mental knowledge and the concerns of statistician Francis Y. Edgeworth about fairness of tests (Edgeworth, as cited by Spolsky, p.171,). In other areas of science, measurement of knowledge in a reliable and statistically assessed way was relatively easy. Unlike others, the field of language testing was sceptical about achieving objectivity in assessment. The main drawback to this was the discussion on linguistic knowledge and whether that could be divided into measurable and statistically assessed segments or not. Although these issues have not found complete response to date and continue to be the subject of discussion of the division of opinion, the second stage of the historical development of these discussions took a clearer shape.

Unlike the first phase, in the second phase there was an increased interest in theoretical concepts and a pool of linguists and experts entered the field. This stage is named by R. Morrow as "Lawn of Tears" (1979, as cited in Fulcher, 2000) because of the obsessive efforts to reach objectivity, in macro as well as micro-skills. Psychometrics, which flourished at the time, dealt only with the reliability of tests and was interested in sophisticated formulae to achieve that. Shohamy (2007, p.143) says that, at the time, nothing was mentioned about '*testing as an experience*'. Nobody dealt with this experience, its consequences and attitudes towards it.

However, the contribution of psychometrics has not completely lost the relation to the social side of tests, but this connection is very limited. On the other hand, as tests become more qualitative in measuring knowledge, the conclusions drawn from tests results are more rigorous thereby reducing the undesirable social impact (McNamara and Roever, 2006). For a time, the tests were far from being social, but with the advent of communicative competence paradigm and introduction in the game of the social context, tests have managed to fill the gaps of these aspects.

1.2.3 The integrative phase

The third phase of theoretical developments in testing came as a natural consequence of the new movement in linguistics, communicative approach. Linguistic concepts about how students learn a foreign language began to change radically with Chomsky (1965, 1970) who proposed the system of knowledge based on linguistic rules and the difference between competence and performance. According to him, competence was knowledge of the ideal speaker/listener on rules of language, and the performance was the actual use of language in concrete situations.

These developments in communicative approach, and the need for tests that measure the productive language skills, led to demands for language tests that include performance. Language Professionals began to believe that the language is more than a collection of separate elements that were tested during psychometric - structuralist movement and that this tradition was focused on the formal system of language, rather than in how the knowledge is used to achieve a successful communicative act (McNamara, 2000). This led to the development and design of tests which integrated more of the skills mentioned above.

1.2.4 The communicative phase

With the developments of the previous phase the foundations of communicative approach in language teaching and testing were laid. The issued this approach raised were not new in the realm of language testing, but what it did was putting forward a serious effort into tackling them. Under this new light, tests were not seen simply as linguistic tests *per se,* rather, more factors were taken into consideration, such as the individual taking the test, and the impact such test results had on his or her life.

The recent models raised questions on previous concepts of linguistic knowledge and the way the tests were compiled to assess it. As psycholinguistics, which dealt with the individual cognitive abilities, was turning

into 'the ephemeral fad', more and more linguists started to become interested in testing the social dimension of language and its use. This dimension was highlighted more by the surfacing of communicative approach, which appeared after structuralist linguistics and behaviourist psychology of the previous stage had failed to solve the problems of teaching and testing.

Fulcher (2000) believes that communicative testing initially came as a response to the enormous importance that had the reliability and validity in the previous stage. Morrow (as cited in Fulcher, 2000) stated that reliability was the requirement of objectivity and validity existed depending on the criteria that were based on questionable assumptions. Redefinition of these two concepts became an important task for testers of communicative testing. Shohamy (2001) states that a number of authors have begun to ask questions about the social and ethical side testing in much the same time. Consequently, language testing took another direction, expressed doubts raised ethical issues, and focused on being right, responsibilities, society and 'washback'.

1.3. The alternative assessment era and ethical issues

Although communicative testing era brought a sea change in linguistic circles, its rewards were more theoretical than practical. The next phase of theoretical developments brought testing closer to solving some of the ethical issues.

1.3.1 Some issues on ethics in language testing

McNamara and Roever (2006) consider ethical issues only a '*re-voicing of an old issue*'. (p.3), because they were raised since the beginning of the previous century.

Fulcher (http://taesig.8m.com/news1.html) gives valuable advice to all those who undertake studies in this field. He believes that instead of '*wandering in the postmodern sea of uncertainty*' it would definitely save our time and energy if we rather studied researchers like Edgeworth,(1880) and Mesick, a century later, who have taken into considerations aspects like validity and construct intertwined with ethical issues, which we now know under the term 'washback'.

Unlike Fulcher, Hamp-Lyons (1997, p. 326) believes that these issues have been introduced only in recent years in research of language testers. After a survey of the most important works from the 60s until recently, she discovered that no known author mentions issues of ethics and responsibility. Furthermore, according to this author, they are not observed even in later works of authors such as Henning and Weir, respectively 1987 and 1993. The only ones who have contributed in this field are Spolsky, 1977, 1980, and Stevenson 1981, who mentioned social, political and educational consequences of tests.

Hamp-Lyons, (1997, p. 326) says that ethics in testing were introduced rather late and have become an integral part of discussions about fairness. In correspondence with Alan Davis, she states that the latter was dissatisfied with the term 'being ethical', and that the professional aspects and humanist orientation should be highlighted. The validity of a test that does not necessarily make it ethical, because it is not exactly a feature of the test itself, but a general approach towards students, the learning process, etc.

Therefore, Hamp-Lyons says humanistic and ethical approach are not readily applicable in the testing process. One of the reasons that affect this difficulty is that '*understanding of what is ethical becomes very difficult in these times of moral and cultural relativism*', (p 326.) which (i.e. cultural relativism) Fulcher (http://taesig.8m.com/news1.html) considers related to cultural habits that vary from community to community, from society to society so that the importance of certain aspects under the different perspectives becomes relative.

However, for Hamp-Lyons (1997), although it is difficult, admitting to ourselves that our decisions affect the lives of many people, as students, assessors, teachers, parents, education policymakers, taxpayers, etc., will make us free. Thus, the ultimate goal of fairness is potentially feasible. On the other hand, the very concept of 'fairness' is difficult to define, especially when the same test can be perceived very differently by different interest groups. However, ethical issues serve as fair guidelines for the impact of tests on testees, stakeholders, and society.

Fulcher (2000) holds a more sceptical attitude towards these issues. He said that the moral dilemmas have accompanied humanity since the days of the Sophists in Ancient Greece. According to him, the most complete form of cultural relativism appears in the latest philosophical shift, the question we ask now is not '*what is knowledge*', but '*what it means*', and that, at postmodern times, this knowledge is the product for consumers, who retain the right to choose what they want to learn. He also mentions the fact that, basically, people are linguistic beings; as a result, our understanding of the reason, logic, ethics, are only linguistic games. For this author, ethics in the postmodern world is local, temporary, and without logical basis.

These questions the author raises are part of reflection of many other authors, because in the end, ethical issues of testing by nature are not consolidated, since, according to this author, the only possible answer is: 'for you, it's okay, but not for me, but next week we could change our minds again." Likewise, Hamp-Lyons *(ibid)* states that we have left behind the simplified positivist paradigm and we have entered the world of relativism, in which the reality depends on the questions you ask and the facts cannot be separated from the sets of values.

1.3.2 Tests and social policies

McNamara and Shohamy (2008) mention the power of tests as the main reason for their use in social policies. These authors believe that '*tests motivate students to learn and teachers to become more effective in their instruction'* (p. 90), and this is the factor that contributes to their use as '*devices which are effective in enforcing conformity and in ensuring the continuity of various declared agendas of policy-makers'.* This makes it clear that the need for success among students is a high stakes matter, therefore, the social impact is huge (Shohamy, 2001, as cited in McNamara and Shohamy, 2008).

McNamara (Jana Fox, et al., 2007) raises yet another question in the long array of issues related to social impact, namely, the social context theory. Although attempts have been made to tackle this issue, the author feels that '[w]*hat we do not have here is a theory of the social context in its own right, as it were, that is, a theory that is not primarily concerned with the cognitive demands of the setting on the candidate.*' (p. 133). Many authors have limited their work in discussing '*the intended and unintended consequences of the test, not to the wider social meaning of the test in its context*' (Bachman, 1990 as cited by McNamara in Jana Fox, et al., 2007 p.133.) By having such theory, this author believes language testers will '*avoid being naive... and become aware ot the roles that tests will play in the operation of power and of systems of social control'.* (p. 136), thus recognizing once more the power of tests.

Because the tests '*are impartial, but often represented in political, social, educational, ideological and economic contexts'* (Shohamy, 2001, p. 113) the first Code of Ethics was adapted for the first time at the annual meeting of the International Language Testing Association (ILTA) (Shohamy, 2001, p.125). This Code is the best example of the expression of the latest concerns of researchers on issues of ethics. It is neither the statute nor the regulation, only suitable guide for ethical behaviour testers of the language.

According to the principles of this Code, the test makers, researchers and linguists who are engaged in all phases of the testing process should be responsive and aware of the consequences of testing on the lives of examinees, however minor. There are three aspects mentioned by McNamara (McNamara, 2000, p. 72) as important: *firstly*, the responsibility to those who will undergo the test, while Fulcher (http://taesig.8m.com/news1.html) notes that this group is wider than just direct consumers. It also includes teachers, school administrators, etc.; *secondly,* the 'washback' effect, originally formulated by Wall and Alderson (1993), dealing with the impact of testing on teaching and especially their relationship, which turns out to be much more complicated than thought earlier; and *thirdly* the impact of the test beyond the educational institution that administers designs in society. Considering these aspects, our discussion takes place in the broader prism philosophical, ethical, social and theoretical than just a matter of language and its components.

While Shohamy (2001, p. 114) provides five views related to the social responsibility of the testers: *first*, an ethical perspective, according to Davis (as cited in Shohamy, 2001) serves as protective contract between the profession of testers, the individual and the public; *secondly,* awareness of others, which means that the tester should make known publicly the test impact and consequences; *thirdly*, the consequences, the tester himself must be aware of the consequences that his decisions cause; *fourth*, sanctions on misuse of tests; and *fifth*, shared responsibility, which deals with more active involvement of the examinees in the testing process, not simply as the end user, but as active participants in policy making on the use of tests.

1.3.3 Solutions offered by alternative assessment

While traditional testing could not fill the gap of involving more actors in the process, alternative assessment could do this, because now testers were introduced with the idea of incorporating portfolios, self-assessments, peer assessments, interviews and observations.

Although concerns about reliability and validity for such kind of assessment are voiced by many linguists, this procedure brought testing closer to its sociocultural character and by doing this, '*we offer more choices to examinees, such as choice of topic, or of texts, for example. In other, more traditional testing systems examinees "may be asked to simulate some real language-use situation, such as giving directions about how to drive somewhere, or treating a patient, but since they know they are in a test, this imposes another set of expectations and norms on the communication*" (Luoma, 2004, p.103, as cited in Dubeau, Master thesis, 2006).

References

Alderson, J. C., Clapham, C. and Wall, D (1995) *Language Test Construction and Evaluation*. Cambridge: Cambridge UniversityP ress.

Davies, A. (1997) "Demands of being professional in language testing". *Language Testing* 14 (3):328–339.

Dubeau. J.J, (2006) "Are we all on the same page". Master thesis, School of Linguistics and applied language studies. Carleton University, Ottawa.

Fulcher, G. "*Ethics in Language Testing*" Retrieved from http://taesig.8m.com/news1.html

McNamara, T. (2000) *Language Testing*, Oxford University Press, p. 72

McNamara, T. and C. Roever (2006) *Language Testing: The Social Dimension*. Malden, MA & Oxford: Blackwell.

McNamara, T., Shohamy, E., (2008) "Language tests and human rights". *International Journal of Applied Linguistics ,*Vol. 18, No. 1, Blackwell Publishing Ltd.

McNamara, T. **(**2007) "Language testing: A matter of context". *Language testing reconsidered* / ed. J Fox, J.et al. University of Ottawa Press: 132.

Hamp-Lyons, L. (1997) "Ethics in Language testing". *Encyclopedia of Language and Education, Volume 7: Language testing and Assessment,* eds. Clapham C. and D.Corson et al. 323-333, Kluwer Academic Publishers, p. 326

Hamp-Lyons, L. (1997) "Washback, impact and validity: Ethical concerns". *Language Testing* 14 (3):295–303.

Shohamy, E. (2001) *The Power of Tests: A Critical Perspective on the Uses of Language*

Tests . London: Pearson.

Shohamy, E. (2007) "Tests as power tools: Looking back, looking forward". *Language testing reconsidered*, ed. J Fox, J.et al. University of Ottawa Press p.143

Shohamy, E., (2001) 'The social responsibility of the Language testers'. *New Perspectives and Issues in Educational Language Policy*, John Benjamins Publishing Company, 2001, p. 113-115

Spolsky, B. (1995) *Measured Words: The Development of Objective Language Testing*. Oxford: Oxford UniversityP ress.

Spolsky, B. (1997) "The ethics of gatekeeping tests:What have we learned in a hundred years?" *Language Testing* 14 (3):242–247.

Spolsky, B. (1995) *Measured Words* . Oxford: Oxford University Press.