

Language tests as language policy tools

Elana Shohamy*

Tel Aviv University, Israel

This paper contextualizes language tests in relation to educational and national language policies by demonstrating how these language measures may be used as mechanisms for affecting *de facto* language policies. This phenomenon is of special relevance given current controversies in nation states between multilingual and multicultural realities and government policies that perpetuate homogenous policies with regard to national languages. The introduction of language tests in certain languages delivers messages and ideologies about the prestige, priorities and hierarchies of certain language(s), and not others, leading to policies of suppression of diversity. Tests also influence language policy with regard to the nature of language, as derived from the criteria used for judging language quality via rating scales, guidelines and frameworks, thus leading to a view of language as standardized and homogenous. Yet language tests, when they incorporate research findings about language learning and language use, can serve as tools for creating more valid and real language policies that mediate and negotiate between ideology and practice. The discussion of these issues is supported by examples from a range of international contexts in this paper.

Tests in context

The past decade has witnessed a major shift in the understanding of the functions, status and roles of language tests. From tools used to measure language knowledge, they are viewed today as instruments connected and embedded in political, social and educational contexts. Accordingly, the quality of tests is not judged merely by their psychometric traits but rather in relation to their impact, ethicality, fairness, values and consequences. The term ‘use oriented testing’ refers to this very phenomenon when tests are connected to psychological, social and political variables that have an effect on curriculum, ethicality, social classes, bureaucracy, politics and knowledge (Messick, 1981, 1994, 1996, 1998). The topic of ‘test use’ poses questions about the roles that tests play in education and society, referred to also as *critical language testing* (Shohamy, 2001). Specifically, it asks questions about tests and does not take them for granted; examples of such questions are: what happens to test-takers taking tests, to the knowledge created by tests, to teachers preparing their students for tests, to materials and methods used in preparation for tests, to decisions made based on test

*School of Education, Tel Aviv University, Israelm69978. Email: elana@post.tau.ac.il

results, to motivation and intentions of policy-makers in introducing tests and to the effect of tests on language knowledge.

This view of tests is derived mostly from the power of tests and its manifestations with regards to high-stakes decisions based on test results for individuals, educational systems and society as a whole. After all, tests are capable of affecting the behaviours of teachers, students, parents and institutions as well as national educational policies. Such behaviours include teachers and students making special preparations for tests by covering specific contents and materials included on tests, principals who put pressure on teachers to succeed on tests to benefit the status and financing of the school and governments allocating special resources to educational systems based on results obtained on national and international tests. It is also claimed that the awareness of the power of tests motivates those in authority to *introduce* tests as ways of controlling and imposing specific knowledge of students, teachers, principles and educational systems alongside certain agendas and educational ideologies (Broadfoot, 1996; Shohamy, 2001).

A number of empirical studies were conducted in the past decade to examine the various consequences of tests in terms of teaching and learning (Alderson & Wall, 1993; Watanabe, 1996; Hamp Lyons, 1997; Cheng, 1998, 2005; Rea-Dickins & Gardner, 2000; Shohamy, 2001; Andrews *et al.*, 2002; Stobart, 2003; Cheng *et al.*, 2004; Little, 2005; Qi, 2005; Hawkey, 2006; Saif, 2006); on construct validity (Bachman, 2005); on policy towards immigrants and asylum seekers (Eades *et al.*, 2003; McNamara & Roever, 2006); on various aspects of multiculturalism (Shohamy, 2004; Evans & Hornberger, 2005; Menken, 2005, 2006); on ethicality and professionalism (Davies, 1997); and on fairness (Kunnan, 2000).

In Shohamy (2001), there is a detailed discussion of the power of tests, its sources, features, temptations and manifestations, along with examples showing how tests can manipulate behaviours, language knowledge, teaching methods, pedagogy and teaching contents. It presents a number of cases in which it is demonstrated how the introduction of a new national reading comprehension test affected teachers so that they felt forced to engage in teaching 'test-like' materials such as texts and multiple choice questions in the classroom rather than teaching reading in more integrated ways. Similarly, the introduction of an English oral test in 12th grade resulted in teaching test-like tasks in the classroom so that teaching became limited to these very tasks which were included on the test. In the case of learning Arabic as a second language, a new test was introduced for the sole purpose of affecting the methods and pace of teaching the Arabic letters in the classroom. This test, it was shown, affected the classroom pedagogy of teaching Arabic, but the effect was limited to the period *before* the test had been administered while the pedagogical methods were changed drastically *after* the test had been administered.

Language policy

While most of the work that examined the impact and washback of tests focused primarily on educational and social dimensions, recent studies have begun to examine

the effects that tests have on additional dimensions such as the direct and indirect effects of language tests on language policies which is the focus of this paper. Of the many definitions of language policy in this paper, *Language policy* (LP) refers to and is concerned with decisions made about languages and their uses in society; *Language education policy* (LEP) refers to such decisions in the specific contexts of schools and universities in relation to home languages, foreign, and second languages. These may include decisions about which language(s) should be taught, when (at what age), for how long (number of years and hours of study), by whom (who is qualified to teach), for whom (who is entitled and/or obligated to learn), and how (which teaching methods, curriculum, materials, tests to be used).

Language policies are often stated explicitly through official documents, such as national laws, declarations of certain languages as 'official' or national, language standards, curricula and tests. At times, though, language policies are not stated explicitly but can rather be derived and deduced *implicitly* by examining a variety of *de facto* practices; in these situations policies are 'hidden' from the public eye. In countries with centralized educational systems decisions regarding language policies are made by central authorities, such as government agencies, parliaments, Ministries of Education, and/or regional and local education boards. In most situations the policies serve as the arm for carrying out national language policy agendas. Thus, when certain entities (i.e., neighbourhoods, communities, cities, nations, global regions), grant languages special priority and status in the society, this policy is manifested in the educational system. This often means using the language(s) as medium(s) of instruction, often the language(s) declared as 'official' and/or teaching the language(s) as foreign/second languages. Such preferred languages may include heritage, community, immigrant, indigenous, foreign or global languages but especially the preferred national language(s), an issue that has special relevance when the official and/or national language(s) are different from the home language(s) for some learners. In the current political environment where states are becoming more multilingual, multi-national and global, students are required to learn language(s) that reflect and affect the interests of different groups in quite different ways.

The inherent complexity of current language use is manifested particularly in the changing nation state, in developing regional and global entities (e.g., the European Union) and in vast migration worldwide, leading to strong demands by multilingual and multiethnic groups for civil rights. In many entities conflicts take place between and within groups and with central authorities, due to the demands for recognition and acknowledgment of difference and special linguistic rights. Many of these battles are manifested through language policies, as the control of languages and linguistic rights facilitates or hinders access to resources in various societal domains such as the workplace, education, or government, and enhances or denies status in society as language is associated with power. In response, state authorities adopt a variety of approaches, from repression of differences to providing solutions that reflect democratic pluralism, although dominant groups are rarely inclined to give up their advantage and accept pluralist policies as these may lead to redistribution of wealth and realignment of political power. Thus, the new composition of the nation

state, with its different ideologies and rules of representation (e.g., common history, kinship) and its connection to the global world, stands in strong tension with the traditional nation state and even threatens it because of the many 'others'. As a result authorities often use propaganda and ideologies about language loyalty, patriotism, collective identity and the need for 'correct', 'pure' or 'native' languages as strategies for maintaining their control and holding back the demands of these 'others'. Language policies therefore provide an interesting platform for observing the dynamic whereby minorities have begun to demand rights while established groups fight to retain their privileged status; on the surface, language policies appear to follow the rules of pluralist democratic societies, including advocating that all citizens should have the opportunity to use a variety of languages.

Language policies and language testing

A number of strategies and mechanisms are used by central authorities to create, perpetuate and manipulate language policies. Even while policies which are expressed in official documents provide relatively transparent information about specific decisions regarding languages, much of language policy is realized through a variety of *indirect* actions and practices that serve as *de facto* policies that can override and contradict existing policies and create alternative policy realities. Language tests, it is argued (Shohamy, 2003, 2006a) represent such a covert mechanism. This results in a situation whereby language policies become no more than declarations of intentions that can easily be manipulated, often in ways that contradict official policies. This can be demonstrated when a language policy declares a specific language as significant and of priority for the educational system. Yet, by establishing entrance criteria that include a test of another language, new *de facto* policy is created, the implication of which is that the 'tested language' becomes the most important language to acquire and master. Indeed, since tests are often more powerful than any written policy document, they lead to the elimination and suppression of certain languages in societies (Shohamy, 2004; Evans & Hornberger, 2005; Menken, 2006). Tests can also be used as tools to privilege certain forms and levels of language knowledge. The LEP may state that correct grammar or 'native-like' accents are not essential for acceptable proficiency; yet language tests do set correct grammar and native-like accents as part of the criteria and these can become barriers for keeping unwanted groups such as immigrants and indigenous groups from entering educational institutions and/or the workplace. Such indirect and covert agendas also stand behind language policies in transnational and global domains in cases when, through the use of English as the language of instruction and/or as a requirement for acceptance to institutions of higher education, the power of English and its speakers is further perpetuated.

Thus, language tests, given their power and influence in societies, play a major role in the implementation and introduction of language policies. They act as mediators between ideologies about language and *de facto* language use in complex political controversies. Schiffman (1996) differentiates between overt and covert policies. Overt policies are explicit, formalized, *de jure*, codified and manifest. Covert policies,

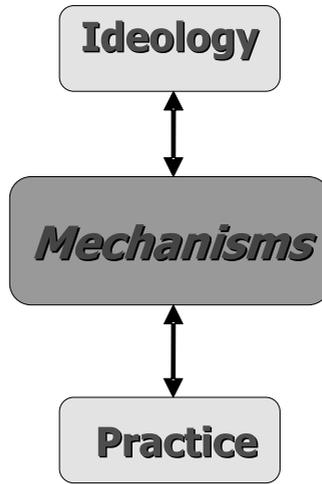


Figure 1. Mechanisms within ideology and practice

on the other hand, refer to language policies that are implicit informal, unstated, *de facto*, grass-roots and latent. He further claims that what are usually ignored are the covert aspects of language policies. Shohamy (2006a) writes about explicit and implicit policies and introduced the term mechanisms as the mediators that create *de facto* language policies. The notion of mechanisms is discussed within both covert

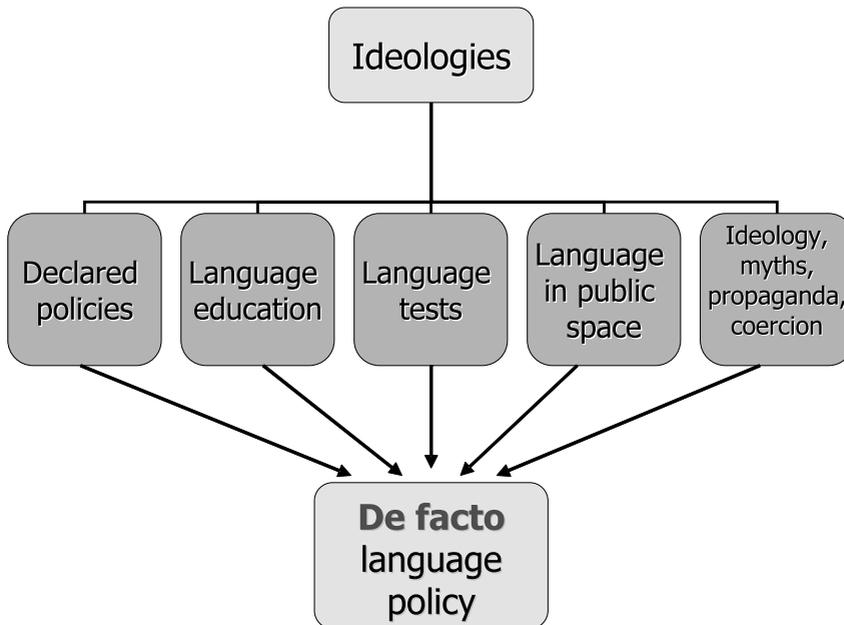


Figure 2. Language tests as a mechanism affecting language policy

and overt policies as capable of turning language ideologies into practice (see Figure 1). Figure 2 points to the very mechanisms that are widely used in most societies for the purposes of turning ideologies into practice; yet it is often the case that practices, via the mechanisms, turn into ideologies. Tests represent such mechanisms through which *de facto* language policies are created and perpetuated.

Earlier in this paper it was shown how language tests were used as mechanisms through which language education policies were manipulated and implemented. In these cases the decision-makers who introduced the tests admitted that these tests were introduced with the clear intent of affecting and imposing new language policies regarding reading comprehension, English speaking and Arabic respectively (Shohamy, 2001). Tests were then used as mediators, as mechanisms for creating language education policies to control and manipulate *de facto* language policy. It was clear that the language supervisors who introduced the tests realized the power of the tests and their potential to affect the behaviour of those who are affected by the tests. The tests were introduced to *cause* such a change in their behaviour in accordance with their own language ideologies. In the case of the reading comprehension tests, these were used to bring about special attention to the teaching of reading comprehension in schools; in the case of the introduction of the English oral tests the national supervisor claimed that the only way teachers would teach oral language in their classes would be by imposing oral language tests at the end of secondary school. The tests then served as mediators to influence classroom behaviour and specific pedagogy. In the case of the Arabic test, the national supervisor felt that the only way Arabic letters would be taught efficiently was if there were to be a test that all students had to pass at a certain date, so that teachers would have no choice but to prepare their students for this test.

In the examples that follow it will be shown how the introduction of language tests can affect various dimensions of language policy in a number of domains: determining prestige, status and hierarchy of language, suppressing diversity and standardizing and perpetuating language correctness and homogeneity.

Determining prestige, hierarchy of languages and suppressing diversity

Given the power of tests, they serve as important tools for determining the status of languages. Thus, it is the act of testing that grants status to the objects that get tested. This means that when those in authority decide to include a certain language as part of school or university exit examinations it automatically grants it high status as it provides an indication that it is being valued highly by 'the authorities'. This is especially relevant given the high-stakes decisions that derive from the results of tests such as graduation or eligibility for entering higher education. Conversely, when those in authority make decisions *not* to test certain languages the message that is being delivered is that the untested languages are marginal and irrelevant. In most nation states, therefore, tests of national and official languages are given priority in national tests so as to perpetuate the languages' high status within the hierarchy of languages. In Israel, entrance tests to higher education are conducted in Hebrew even though all

Arabic students conduct their schooling through Arabic; the act of testing further perpetuates the low status of the language within Israeli society. Similarly, in the US the status of English versus all other languages used is reinforced with the No Child Left Behind (NCLB) tests. Evans and Hornberger (2005) and Menken (2005, 2006) demonstrate how the administration of the NCLB tests via English only contributes to the perception of the irrelevance of immigrant languages and devalues and even abolishes other languages. In the case of immersion programmes whereby content is taught via other languages (e.g. Navajo in some schools in Arizona), the NCLB provides a direct message about that language, namely that Navajo has no meaningful worth, and prestige. It is often the case that through the administration of tests in official languages students are not rewarded for mastering other languages; in fact they are being penalized. Similarly, when immigrant students are required to take tests in English soon after their immigration, not only does it send a message about the marginality of these languages but also that any knowledge that is manifested through another language has no value so that the knowledge is devalued as well (Menken, 2005, 2006). Furthermore it fails to acknowledge that learning a new language is a long process. Such testing policies are implemented in spite of a large body of research pointing to the length of time it takes immigrant students to perform well academically in the new language (Thomas & Collier, 1997). In a large study conducted in Israel on a national sample (Levin *et al.*, 2003, 2007 in press), it was shown that the length of time needed for immigrants to reach academic levels equal to those attained by native speakers ranged from 9 to 11 years in both Mathematics and in Hebrew (the new language acquired by these immigrants). Yet most national tests administered to immigrants overlook such findings and create unrealistic policies with regard to school achievement that have major negative consequences for these students and their lives.

The effect of language testing policy in suppressing the learning of foreign languages is demonstrated in Byrnes (2005) where teachers and scholars report on the changes that have been taking place at schools in the US as a result of the No Child Left Behind (NCLB) programme. They show how these tests affected the teaching and learning of foreign languages as these are not included in the NCLB. They claim that this policy is having a major impact on the number of students enrolled in foreign language classes and even foreign language teachers are now required to transfer their expertise to the teaching of English as a second language in order to accommodate the demands of these tests. This brings about a reduction in the number of foreign language classes offered, a phenomenon that they claim will be difficult to reverse.

Language tests thus contribute to and perpetuate assimilative agendas rather than interactive ones, as they perpetuate the status of dominant languages and the suppression and elimination of the unique knowledge of minority groups. As noted earlier, this occurs even when official curricula may contain statements recognizing diverse knowledge; yet the tests over-ride these statements and explicit official policies. A number of scenarios are presented in Shohamy (2004) showing how language tests impose monolingual language policies and thus suppress multilingual diversity because the languages of minority groups are not valued and appreciated—the tests serve as tools to affirm and reinforce these hierarchies.

Standardizing, homogenizing languages and perpetuating criteria of correctness

One of the most salient uses of tests affecting language policies is in perpetuating language homogeneity, a construct which is detached from the reality of what languages are and how they are being used, especially in multilingual societies. Most tests impose homogenous criteria of correctness and thus deliver a message that languages are uniform, standard and follow the same written norms. Thus, Milroy and Milroy (1999) note that tests tend ‘to characterize speech in terms of writing’ (pp. 139–140) while spoken norms are ignored on most tests, which leads to the issue that ‘normal spoken features tend, as a result, to be stigmatized’ (p. 140). Thus, tests are capable of influencing the definition of languages and of creating new genres. They also note that language tests allow only for standardized answers in terms of grammar and lexicon. ‘They do not usually attempt to measure these more subtle but important aspects of language ability’ (p. 142). As such, tests over-simplify language which, they claim, is characteristic of most doctrines of correctness, ‘and it is clearly these popular correctness doctrines, rather than careful linguistic analysis, which underlie the design of many language tests...’ (p. 142). In fact, they argue, results of tests are often used to support arguments that non-standard English-speaking candidates lack language ability in some unspecified way. Scores are thus used to lend apparent objectivity to judgements based on prescriptive ideologies... and ‘...unjustifiable notions of correctness are inherent in the design of tests...’ (p. 145). Moreover,

This widespread use of standardized testing as a mode of assessment is predicated on the assumption that standardized tests are objective; this assumption in turn does not appear to take into account differences in communicative competence between different groups of speakers. (p. 145)

Another context in which uniformity is established is through rating scales. Anchored historically in definitions provided by US Government agencies, such as the FSI (Foreign Service Institute), the DLI (Defence Language Institute), and the Peace Corps, different proficiency scales have been designed as criteria for determining language quality. These are aimed at criteria that would accord with viewing language development as a progression and a hierarchy of development as learning progresses along the second/foreign language continuum from novice (minimal amount of language), through ‘some language’ to ‘some more,’ and via ‘advanced’ to a ‘professional’ level. Different terms describe this progress. For example, the Council of Europe (2001) proficiency scales differentiate between *Breakthrough* level (A1), *Waystage* (A2), and a mastery level (C2) with specific criteria describing the skills associated with each level; the Association of Language Testers of Europe (ALTE) has translated these specifications into similar descriptions that are part of scales (North & Schneider, 1998; North, 2000, 2004).

These scales affect *de facto* language policy. First, these scales define a presumed hierarchical nature of second language learning, as though it followed a prescribed and controlled linear order without an empirical basis for this assumption. For example, it is not clear whether these hierarchies represent the reality of the process of second language (L2) learning, whether all learners proceed along the same route,

whether a particular level is in fact higher than the previous one as stipulated by the guidelines, and whether these levels potentially represent deep-seated ideologies rather than the reality of language development. It is especially troubling since there is little convincing evidence for the claim that L2 learning actually works in a clean, linear, and homogeneous order of progression that is similar for all learners. And yet the proficiency descriptions have deeply influenced the views of language learning that schools and universities have adopted in terms of language policies.

The *Common European Framework* (CEF) seems to take on a similarly powerful position in educational decision-making in most nations in Europe (Morrow, 2004). In terms of policy, Fulcher (2004) writes that, over time, the guidelines have created a 'false' truth for teachers and bureaucrats, with no evidence of their validity and they serve as 'prescriptions' that dictate proficiency levels in a way that is detached from reality. Instead of *defining* levels of language proficiency they have become institutionalized and reified as 'language': the main danger, Fulcher (2004) claims, is that teachers are beginning to believe that the scales represent an acquisitional hierarchy, rather than a common perception. These scales serve as testing tools which prescribe proficiency. In Shohamy (2006b) it is claimed that such descriptors within these scales are especially problematic for *advanced* language proficiency use, which includes cognitive abilities, content knowledge, context awareness, input processing capacities, interactive abilities, and multilingual performance options—components that are not being addressed in these scales. Further, they are detached from a variety of contextual variables such as the purpose of the assessment, the specific uses of the language, the context in which the language has been learned, the age of the learners, the learning conditions, the specific languages learned and assessed, and especially the multiple functions of different languages in different contexts, and tend to view language learning in homogenous terms that can be generalizable from one domain to another. There are therefore doubts as to whether such broad and generic testing descriptions are relevant and valid for different language learning contexts and uses, such as foreign language learning, second language learning, immersion programmes, bilingual programmes, immigration contexts, indigenous languages, specific grade levels, instructed learning, content based instruction, tertiary education, elementary and high schools, and for capturing the variety of language needs that characterize different workplaces. This shows the problems that arise when test criteria such as rating scales affect language policy, and definitions of 'what it means to know a language' when such rating scales presuppose a hierarchy of both development and performance, adhere to generic descriptions and claim to be universally applicable, detached from the contextualized nature of language and language performance in multilingual environments.

Discussion

As has been shown in this paper, the fact that tests are such powerful tools in society means that they are capable of affecting a number of important societal domains. They are therefore used extensively by central authorities to create washback and impact and *de facto* language policies and language education policies that override

and contradict existing declared policies and regulations. They often do this in covert ways such that the public at large is not aware of this powerful role that tests may play in these domains and often focuses on tests *per se*. Yet, with the recent turn towards examining tests in broader ways—especially by focusing on their uses—these issues emerge and call for broadening the scope of the quality of tests. It was shown how in the changing nature of nation states, language tests become central tools for manipulating the linguistic repertoire, teaching material, language standards and language diversity. The results are that language policy *documents* often become no more than declarations of intent that can easily be manipulated and stand in stark contradiction as the ‘tested language’ obtains prestige and recognition. Such use of tests often leads to violations of language and personal rights, as in the case of language tests for citizenship, a phenomenon that is now widespread in many countries. The introduction of these language tests serves as the main gate-keeping tool in immigration processes, as well as in the repatriation of those who have been living in a country for some time. Thus, the use of language tests for policy-making serves governments as a means of carrying out and implementing a variety of public policies resulting in unethical immigration policies: the language tests serve as tools for denying residence, entrance to educational institutions and the workplace. Their use is especially problematic for those who are of an age and in a context where they are unable to acquire, or master, new languages at the high level of language proficiency that most of these tests require. The ramifications of expulsion from countries, as well as the denial of social benefits, are far-reaching and represent a serious violation of human rights. It is through language tests that various language policies are being implemented in schools, universities and societies (see Garcia Skutnabb-Kangas & Torres-Guzman, 2006, for this manifestation in schools). The washback and impact of these tests is that they influence and redefine knowledge, language priorities, hierarchies, standards, choice of languages, methods and content studied, as well as the criteria for language standards and correctness. In most political systems in the world today, tests drive the system and serve as mechanisms that affect *de facto* language policies.

Tests as mechanisms are considered covert policies and it is rare that these issues are raised and openly discussed. Yet those tested have no choice but to comply with every testing policy even though the consequences may be detrimental. As noted above, there is ample research pointing to the length of time needed to acquire a new language, yet all immigrant students are required to be tested in the new languages, with no option of refusal, while the scores can lead to ramifications such as preventing graduation, attending higher education institutions or obtaining jobs.

As has been shown tests play a major role in the competing language policy battles of central agencies seeking implementation and control of homogenous language policies and the desire for individuals and groups for language freedom; between the drive of nations for a common unifying language and multiple use of languages; between a monolingual ‘one language for all’ policy and multilingual tolerance; between resentment of control by centralized agencies and the need for control to maintain status and social order; between homogenous language knowledge and fluid and more open language boundaries of hybrids and fusions resulting from intensive

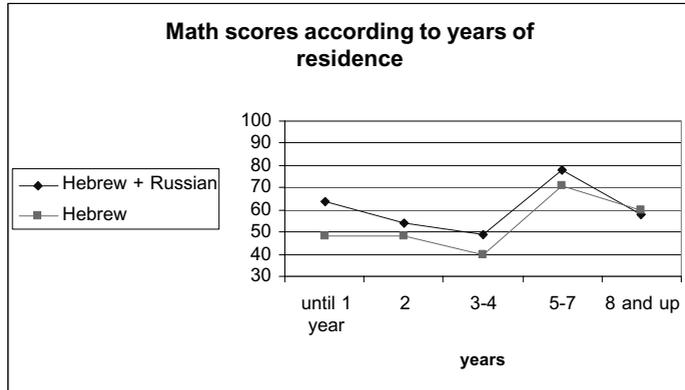


Figure 3. Bilingual test accommodations as enhancing achievement in mathematics

contacts in pluralistic societies. Given the power of tests and their roles in manipulating language policies, there is a need to apply democratic principles and practices in the use of language tests in the domain of language policy where practice can affect ideologies. Tests, then, can also serve as tools for such *negotiations* over language policies. They can include different voices in the assessment, examine the uses of tests from a critical language testing perspective; consider the development of shared and collaborative assessment models that will incorporate such diversity; allow the resistance to tests which is ideologically based and for which test-takers are required to pay a high price. One such example is the use of test accommodations (see Fox & Cheng, this issue) that can be used to reflect more accurately the knowledge of test-takers.

For example, in Figure 3 it is shown how immigrant students obtained higher scores in mathematics when the test questions were presented in a bilingual form (Hebrew and Russian) compared with a control group that received the identical test in a monolingual version, Hebrew only version. This advantage was sustained for more than 10 years of residence in the new country.

Empirical data of this sort can be helpful in creating different testing policy, based as they are on the real language processing of immigrant students who continue to make meaning in two languages long after the point of immigration. Denying immigrants accommodations during the test-taking process may also be viewed as a violation of language rights, by denying students opportunities to demonstrate their true academic potential.

Similarly, the imposition of language citizenship tests as a condition for residence needs to be challenged and negotiated as these tests represent an imposition, with no questions asked. Tests, in this context, also represent a strong form of violation as it is difficult to prove the connection between language proficiency and good citizenship. In terms of language correctness and the homogeneity that language tests are prescribing, this needs to be examined empirically especially in the multilingual and global world where language hybrids and fusions are used extensively (Canagarajah, 2006). Tests that are based on homogenous criteria of correction and that do not

recognize diversity, let alone the use of multilingual and multi literacy codes are no longer realistic. It is, perhaps, the use of multilingual tests that can offer a more valid type of assessment for immigrant populations as well as for students in so many countries, now English plays a major role in the language repertoire.

Tests clearly have an impact and strong washback on many areas beyond classroom teaching: they can affect actual language policy, at educational or societal levels. There is, thus, a need to examine further these washback and impacts in several directions. It is clear that it is often the case, as with the teaching of immigrants in the US, that the requirements produced by the washback of the NCLB are in violation of students' personal rights, as they do not reflect the actual processing of language learning. Such impact may create considerable harm, based as they mostly are on 'wishful thinking' and the misuse of tests. Thus washback, which is a built-in phenomenon, is based on a false assumption. It is important to examine the ways in which tests may be creating false expectations with regard to criteria of language correctness and use of language that they impose. For example, if immigrants continue to use linguistic hybrids and fusions as a result of interacting in bilingual societies, there is a need to create tests that reflect this diversity, as opposed to the other way round where tests are given a free reign to impose unrealistic expectations.

The message of this paper is not a call for the abolition of tests, nor a denial of their influence on language policy but, rather, a call for a view of tests as tools that serves as a basis for negotiating more democratic testing and a closer connection to real policies. How can washback be rejected?. This paper further claims that language tests should mediate ideologies and practices in more open, democratic and negotiable ways, and prevent the use of tests as powerful mechanisms capable of imposing draconic policies that have no empirical base. This happens especially when language tests violate diversity, when a false view of language development is being dictated through tests, when language is viewed in isolated ways detached from actual use of multilingual codes in communities, or when there are empirical data about the advantage of different accommodations that is being denied. Building on the power of tests and on the phenomenon of washback *per se*, without viewing the whole picture, is very problematic and dangerous. Yet it is possible to build on the washback phenomenon in more constructive and negotiable ways. Thus, language tests can become a useful tool for negotiating between language ideologies and language practices as, according to Davies (1997), this is an indication of professional and ethical behaviour.

Notes on contributor

Elana Shohamy is professor and chair of the Language Education Programme, School of Education, Tel-Aviv University, Israel. Her recent research focuses on political, social and educational dimensions of language tests and a variety of topics related to multilingualism within a framework of critical pedagogy.

References

Alderson, J. C. & Wall, D. (1993) Does washback exist? *Applied Linguistics*, 14, 115–129.

- Andrews, S., Fullilove, J. & Wong, Y. (2002) Targeting washback—a case study, *System*, 30, 207–223.
- Bachman, L. (2005) Building and supporting a case for test use, *Language Assessment Quarterly*, 2(1), 1–34.
- Broadfoot, P. (1996) *Education, assessment and society: a sociological analysis* (Buckingham, Open University Press).
- Byrnes, H. (2005) Perspectives, No Child Left Behind, *Modern Language Journal*, 89(2), 248–282.
- Canagarajah, S. (2006) Changing communicative needs, revised assessment objectives: testing English as an international language, *Language Assessment Quarterly*, 3(3), 229–242.
- Cheng L. & Y. Watanabe with A. Curtis (Eds) (2004) *Washback in language testing: research contexts and methods* (Mahwah, New Jersey, Lawrence Erlbaum Associates).
- Cheng, L. (1998) Impact of a public English examination change on students' perceptions and attitudes toward their English learning, *Studies in Educational Evaluation*, 24(3), 279–301.
- Cheng, L. (2005) *Changing language teaching through language testing: a washback study. Studies in language testing: vol. 21* (Cambridge, Cambridge University Press).
- Council of Europe (2001) *Common European Framework of reference for language learning and teaching* (Cambridge University Press, Cambridge).
- Davies, A. (1997) Demands of being professional in language testing, *Language Testing* 14(3), 328–339.
- Eades, D., Fraser, H., Siegel, J., McNamara, T. & Baker, B. (2003) Linguistic identification in the determination of nationality: a preliminary report, *Language Policy*, 2, 179–199.
- Evans, B. & Hornberger, N. (2005) No child left behind: repealing and unpeeling federal language education policy in the United States, *Language Policy*, 4(1), 87–106.
- Fox, J. & Cheng, L. (2007) Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers, *Assessment in Education*, this issue, 9–26.
- Fulcher, G. (2004) Deluded by artifices? The Common European Framework and harmonization, *Language Assessment Quarterly*, 1, 253–66.
- Garcia, O., Skutnabb-Kangas, T. & Torres-Guzman, M. E. (Eds) (2006) *Imagining multilingual schools* (Clevedon, Avon: Multilingual Matters Ltd.).
- Hamp-Lyons, L. (1997) Washback, impact and validity: ethical concerns, *Language Testing*, 14, 295–303.
- Hawkey, R. (2006) *Impact theory and practice: studies of the IELTS Test and Progetto Lingue 2000* (Cambridge University Press, UCLES, Cambridge).
- Kunnan, A. (Ed) (2000) Fairness and validation in language assessment: selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida. *Studies in Language Testing: Volume 9*, (Cambridge, Cambridge University Press).
- Levin, T., Shohamy, E. & Spolsky, B. (2003) *Educational achievement of immigrant students. Research report submitted to the Ministry of Education* (Tel Aviv University, School of Education) [in Hebrew].
- Levin, T. & Shohamy, E. (2007, in press) The role of academic language in understanding the mathematics achievements of immigrant students, in: C. S. Sunal & K. Mutua (Eds) *The enterprise of education: research on education in Africa, the Caribbean, and the Middle East* (Tuscaloosa, Alabama, The University of Alabama, Info Age Publishing).
- Little, D. (2005) The Common European Framework and the European Language Portfolio: involving learners and their judgements in the assessment process, *Language Testing*, 22(3), 321–336.
- McNamara, T. & Roever, C. (2006) *Language testing: the social dimension* (Malden, MA and Oxford, Blackwell).
- Menken, K. (2005) *When the test is what counts: how high-stakes testing affects language policy and the education of English language learners in high school*. Doctoral Dissertation. Teachers College, Columbia University.
- Menken, K. (2006) Teaching to the test: how standardized testing promoted by the No Child Left Behind Act impacts language policy, curriculum, and instruction for English language learners, *Bilingual Research Journal*, 30(2), 521–546.

- Messick, S. (1981) Evidence and ethics in the evaluation tests, *Educational Researcher* 10, 9–20.
- Messick, S. (1994) The interplay of evidence and consequences in the validation of performance assessment, *Educational Researcher*, 23, 13–23.
- Messick, S. (1996) Validity and washback in language testing, *Language Testing*, 13(4), 241–57.
- Messick, S. (1998) Validity, in: R. L. Linn (Ed.) *Educational measurement* (3rd edn.) (Washington, DC, The American Council on Education and the National Council on Measurement in Education), 13–103.
- Milroy, J. & Milroy, L. (1999) *Authority in language. Investigating standard English* (3rd edn.) (London, Routledge)
- Morrow, K. (Ed.) (2004) *Insights from the Common European Framework* (Oxford, Oxford University Press).
- North, B. (2000) *The development of a Common Framework Scale of language proficiency* (New York, Peter Lang).
- North, B. (2004) Relating assessments, examinations, and courses to the CEF, in: K. Morrow (Ed.) *Insights from the Common European Framework* (Oxford, Oxford University Press), 77–90.
- Qi, L. (2005) Stakeholders' conflicting aims undermine the washback function of a high-stakes test, *Language Testing*, 22, 142–173.
- Rea-Dickins, P. & Gardner, S. F. (2000) Snares and silver bullets: disentangling the construct of formative assessment, *Language Testing*, 17(2), 215–243.
- Saif, S. (2006) Aiming for positive washback: a case study of international teaching assistants, *Language Testing* 23, 1–34.
- Schiffman, H. (1996) *Linguistic culture and language policy* (London, Routledge).
- Shohamy, E. (2001) *The power of tests, a critical view of the uses of language tests* (Essex, England, Longman Pearson).
- Shohamy, E. (2003) Implications of language education policies for language study in schools and universities, *The Modern Language Journal*, 87, 278–286.
- Shohamy, E. (2004) Assessment in multicultural societies: applying democratic principles and practices to language testing, in: B. Norton & K. Toohey (Eds) *Critical pedagogies and language learning* (New York/London, Cambridge University Press), 72–93.
- Shohamy, E. (2006a) *Language policy: hidden agendas and new approaches* (London, Routledge).
- Shohamy, E. (2006b) Re-thinking assessment for advanced language proficiency, in: H. Byrnes, H. Weger-Guntharp & K. A. Sprang (Eds) *Educating for advanced foreign language capacities: constructs, curriculum, instruction, assessment* (Washington, DC, Georgetown University Press), 188–208.
- Stobart, G. (2003) The impact of assessment: intended and unintended consequences, *Assessment in Education*, 10(2), 139–140.
- Thomas, W. & Collier, V. (1997) *School effectiveness for language minority education students* (Washington, DC, National Clearinghouse for Bilingual Education).
- Watanabe, Y. (1996) Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research, *Language Testing*, 13, 318–333.