

---

# Critical Language Testing

Elana Shohamy

---

## Abstract

Critical language testing (CLT) refers to the examination of the uses and consequences of tests in education and society (Shohamy 2001a, b; Spolsky 1995). The topic gained attention by various scholars and particularly Messick (1981, 1989), who argued for expanding the definition of construct validity as a criterion for evaluating the quality of tests, to include components related to tests use, such as values, impact, and consequences. CLT emerged from the realization that tests are powerful tools in education and society, which may lead to unintended consequences that need to be examined and evaluated. It is the power of tests, especially those of high stakes, that causes test takers and educational systems to change their educational behaviors and strategies as they strive to succeed in tests given their detrimental impact.

Ample research on CLT exists which focuses mainly on the uses of tests with regard to high-stakes tests such as the TOEFL, school leaving exams, entrance and placement tests, as well as international/comparative tests such as PISA and TIMSS. These studies pointed to the misuses of tests and their impact that goes far beyond learning and teaching into issues of identity, educational policies, as well as marginalization and discrimination against immigrants and minority groups. The chapter ends with a discussion of alternative testing strategies, developed over the past decade, which aim at minimizing the power and negative consequences of tests mostly by including democratic approaches of formative and dynamic assessment, multilingual testing, inclusive assessment, and bottom-up testing policies and tasks, all aiming to use tests in constructive and positive ways, diminishing their excessive power.

---

E. Shohamy (✉)  
School of Education, Tel Aviv University, Tel Aviv, Israel  
e-mail: [elena@post.tau.ac.il](mailto:elena@post.tau.ac.il)

---

**Keywords**

Consequential validity • Washback • Multilingual assessment • Ethicality • Democratic assessment • Values • Assessment literacy

**Contents**

Introduction and Early Developments .....	442
Major Contributions .....	444
Work in Progress .....	447
Responses: Minimizing and Resisting Power .....	449
Future Directions .....	450
Cross-References .....	451
Related Articles in the Encyclopedia of Language and Education .....	451
References .....	452

---

**Introduction and Early Developments**

In most countries worldwide, individuals are subject to tests, whether to enter educational programs, to pass from one level to the next, or to be granted certificates to practice professions. Tests determine whether students will be allowed to enter high schools and higher education and in many cases even kindergartens and elementary schools. In schools, classroom tests are used in all subjects and grades and have an effect on students' status in their classrooms as well as on their identities and self-concepts. Tests are used by teachers as disciplinary tools to control students' behaviors and the curricula and to upgrade the status and prestige of specific topics and subjects. High-stakes tests lead to rejections and acceptances, to winners and losers, and to successes and failures and hence have an impact on people's lives. For adult immigrants, tests determine whether they will be granted permission to immigrate and to obtain citizenship in countries they moved to or seek asylum.

Critical language testing (CLT) originated from a focus on the *uses* of language tests and the realization of their enormous power to influence education, societies, and even the status of nations as a result of performances on international tests. It is the power of tests and the detrimental decisions they bring about that grants them such status in society so that people change their behavior in order to succeed on tests (Shohamy 2001a). It is this very power that brings about decision makers and those in authority to introduce tests since they know that once a high-stakes test is introduced, it is most likely that principals, even if the curriculum has not changed, will start imposing the teaching of these topics, and students will be forced to learn them. Hence, there is a change in stakeholders' behaviors in an intensive effort to achieve high scores. In fact, in many schools the content that is included in these tests becomes the *de facto* curriculum and often overlooks the written curriculum that already exists as those who introduce tests often have different educational agendas (Cheng 2004; Cheng and Curtis 2004 and others).

Two examples that demonstrate the phenomenon are the following: The first in the context of migration (Extra et al. 2009), where adult immigrants, moving

to a new country are required to take tests of their proficiency in the language used in the new location as a condition for citizenship and residence. At times these tests are being administered still in their home countries and thus restrict the number of immigrants. Governments implement language testing regimes as a way to control the number of immigrants they allow to enter the country and/or of those who can stay there. In most nations nowadays immigrants are required to pass a test in the main official language of the country. This policy does not originate from research findings that demonstrates that proficiency in national languages is relevant for functionality; still, language tests become the tool for screening, leading to decisions as to whether immigrants are allowed to stay in the country or would be forced to leave. It also ignores situations when immigrants are at an age that they are incapable of learning the new language and/or cannot read or write in their own language or when there are no learning opportunities such as language courses where they can learn the new language (McNamara and Shohamy 2008; Shohamy and Kanza 2009). It is also known that many immigrants tend to be employed in their own communities and are very comfortable using their home languages which are functional for them in most domains of everyday lives. The test then is used primarily as a tool to screen immigrants, which brings about enormous criticism about the ethicality of these types of tests as they are used for purposes they were not intended to. The children of immigrants usually acquire the new language relatively fast in comparison to their parents because they are schooled in that new language as a medium of instruction, albeit, this too takes a long time (Levin and Shohamy 2008) as will be reported below.

The second case is the testing of immigrant and minority school students who lack high proficiency in the power language which is the medium of instruction in schools. In this case students are required to take standardized tests as mandated by national policies after a short time of being in the country. While research shows that it takes immigrants about 10 years to acquire a new language (Collier and Thomas 2002; Valdés et al. 2015; Levin and Shohamy 2008) and yet while they are still in the process of learning the new language, they are being tested in school content areas via the new language. Given that the students are not proficient in the language yet, they often fail these tests in the different academic subjects and become marginalized and discriminated against by their teachers and peers (Levin and Shohamy 2008; Levin et al. 2003).

In both of the above-described cases, language testing policies are used as disciplinary tools given that test takers have no choice but to comply with the policy demands. While test takers and regional educational systems comply with such disciplinary demands, they also resent them as they feel they were imposed on them without their voice being heard. It is the powerful uses of tests – their detrimental effects and their uses as disciplinary tools that are responsible for the strong feelings that tests evoke in test takers. It is the raising of critical questions about the testing policy and their impact and consequences as well as the intentions behind the introduction of these tests which is the essence of CLT.

## Major Contributions

**A social perspective.** The use of tests for power and control was argued convincingly by Foucault. In *Discipline and Punish: The Birth of the Prison* (1979) Foucault stated that examinations possess built-in features that enable them to be used for exercising power and control. Specifically he mentions that tests serve as means for maintaining hierarchies and normalizing judgment. They can be used for surveillance, to quantify, classify, and punish. Their power lies in that they can lead to differentiation among people and for judging them. Tests consist of rituals and ceremonies along with the establishment of truth and all in the name of objectivity, as Foucault puts it:

The examination combines the techniques of an **observing hierarchy** and those of a **normalizing judgement**. It is a **normalizing gaze**, a **surveillance** that makes it possible to **qualify**, to **classify** and to **punish**. It establishes over individuals a **visibility** through which one **differentiates** them and **judges** them. That is why, in all the mechanisms of discipline, the examination is highly **ritualized**. In it are combined the **ceremony of power** and the **form of the experiment**, the **deployment of force** and the **establishment of truth**. At the heart of the procedures of discipline, it **manifests the subjection** of those who are perceived as objects and the **objectification** of those who are subjected. (p. 184) (my emphasis)

In Foucault's biography, written by Eribon (1992), he provides evidence of Foucault's personal experiences and sufferings from tests, making him a "test victim." He shows that Foucault himself was a victim of tests, who failed on high-stakes tests. References are made to situations when tests played detrimental roles in his own life, possibly causing him to gain the special insight into the uses of tests as disciplinary tools. Foucault (1979) also noted that it is only in the twentieth century that testers made tests "objective unobtrusive" messengers, while in the past testers had to face test takers directly and to share the responsibility for the testing verdict.

The notion that tests represent a social technology is introduced by Madaus (1990) as an extension of the uses of tests as disciplinary tools. He claimed that tests are scientifically created tools that have been historically used as mechanisms for control and their power is deeply embedded in education, government, and business. The test is a means for social technology as it not only imposes behaviors on individuals and groups but also defines what students are expected to learn and know and can therefore be referred to as "de facto curriculum." It therefore guaranteed the movement of knowledge from the teacher to the pupil, but it extracted from the pupil a knowledge destined and reserved for the teacher.

Bourdieu (1991) claimed that tests serve the needs of certain groups in society to perpetuate their power and dominance; thus, tests were rarely challenged. Tests have wide support of parents, as they lead to the imposition of social order. For parents who often do not trust schools and teachers, tests provide indication of control and order, especially given their familiarity with tests in their own years of schooling. For many parents tests symbolize control and discipline and are perceived as indications of effective learning. It is often observed that raising the educational standards

through testing appeals to the middle classes, partly as it means gaining access to better jobs for their children, and for some it is also a code word for restricting minority access. The paradox is that low-status parents, minorities, and immigrants, who are constantly excluded by tests, have an overwhelming respect for them and often fight against their abandonment.

Hanson (1993) as well discusses the power of tests to affect and define people and notes that tests have become social institutions on their own, taken for granted with no challenging questions. Specifically, while a testing event is only a minute representation of the whole person, tests are used both to define and predict a person's ability as well as to keep them powerless and often under surveillance. He adds the following:

In nearly all cases test givers are (or represent) organizations, while test takers are individuals. Moreover, test-giving agencies use tests for the purpose of making decisions or taking actions with reference to test takers – if they are to pass a course, receive a driver's license, be admitted to college, receive a fellowship, get a job or promotion. . . . That, together with the fact that organizations are more powerful than individuals, means that the testing situation nearly always places test givers in a position of power over test takers. (Hanson 1993, p. 19)

The use of language tests as disciplinary tools by powerful political institutions is discussed by McNamara (1998) who notes that tests have become an arm of policy reform in education and vocational training as well as in immigration policies. Such policy initiatives are seen within the educational systems as well as in the workforce. A concern for national standards of educational achievement in a competitive global economy, together with a heightened demand for accountability of government expenditures, has propelled a number of initiatives involving assessment as an arm of government educational policy in the national, state, and district levels.

**A psychometric perspective.** Some psychometricians who themselves develop tests have been critical about them. Most notable is Messick, who was employed at the Educational Testing Service in the USA, a center that develops and researches tests. Messick (1981, 1996) was among those who drew attention to the topic of impact, claiming that tests' consequences should be incorporated into a broader perspective of a unified concept of validity. He argued that given that social values were associated with intended and unintended outcomes, the interpretations and uses which derive from test scores, the appraisal of the social consequences of tests should be subsumed as aspects of construct validity (1996, p. 13). Messick (1996) claimed that "[i]n the context of unified validity, evidence of washback is an instance of the consequential aspect of construct validity." Thus, Messick's concept of unified validity seems to be the bridge between the narrow range of effects included in washback and the broader one encompassed by "impact" which includes "...evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use. . . especially those associated with bias. . . unfairness in test use, and with positive or negative washback effects on teaching and learning" (p. 12). The term "consequences" is used mostly by Messick to encompass washback and construct validity but with a stronger focus on ideological values. This

is also how the term is used here to discuss the societal influences of tests in a larger scope. Messick notes that washback is only one form of testing consequence that needs to be weighed when evaluating validity, and testing consequences are only one aspect of construct validity, leading to the term *consequential validity*. An additional term often used to refer to the connection between testing and instruction is *systemic validity* (Frederiksen and Collins 1989) relating to the introduction of tests into the educational system, along with additional variables which are part of the learning and instructional system. In such situations tests become part of a dynamic process in which changes in the educational system take place according to feedback obtained from the test. Similar terms associated with the impact of tests on learning are *measurement-driven instruction*, referring to the notion that tests drive learning, and *curriculum alignment*, implying that the curriculum is modified according to test results.

Thus, although psychometricians developed sophisticated methods for test development and design, in terms of reliability and validity and quality of items and tasks, they tend to overlook the important dimension of consequences of tests. This leads to the need to pose questions that will incorporate the consequences such as: What are the tests being used for? What purposes are they intended for? Do they lead to decisions which are beneficial or harmful for people? Are they meant to evaluate the level of language proficiency or as sanctions for discipline and control? In other words, is a test really a pure measurement of language proficiency or is it used as a disciplinary tool for other agendas such as selection, expulsion, and differentiation leading to stigmas about different populations and their rejection from bastions of society?

**Language testing perspective.** In the book *Measured Words*, Spolsky (1995) surveyed the different language tests from a historical context. He brings up the cases of the different agendas that were associated with the TOEFL tests to prevent people from certain areas in the world to study in US educational institutions.

Shohamy (2001a) introduced the notion of CLT and focused on three studies which demonstrated how the introduction of high-stakes language tests brought about major changes in the behavior of the school: a test of oral proficiency in English as a second language which led to teaching to the test, a test in Arabic which turned the classes to prepare students for the test which meant studying the exact content of the test, and a national test for testing reading comprehension which in a year's time changed drastically the reading curriculum to texts with multiple questions; in all cases, there was narrowing of the curriculum and the test dominated most activities. In other words, once the test was administered, the teaching returned to non-testing activities. A study (Shohamy et al. 1996) examined the effect of these tests several years later and showed that the only meaningful change took place in the high-stakes tests while in the case of low-stakes tests these changes were totally overlooked. Shohamy and McNamara (2009) critiqued the tests for citizenship. In Shohamy (2009) there is a strong argument and a list of reasons against the use of tests for enforcing such policies.

Studies by Alderson and Wall 1993 examined a number of hypotheses whereby one could expect a change in the school learning policy of languages due to tests but found very little effect due to the tests.

Cheng (2004; Cheng et al. 2011; Cheng and Curtis 2004; Cheng and DeLuca 2011) conducted studies focusing on the washback of high-stakes tests, especially in China but also in Canada and elsewhere. They found major impact of tests on teaching. More information about these studies can be found in the chapter “► Washback, Impact, and Consequences Revisited” by Tsagari and Cheng in this volume, examining at least two different types of washback studies, one related to traditional standardized tests and the other in which modified versions of tests are examined as means for achieving more positive influence on teaching and learning.

Fulcher (2004) critiqued the growing number of rating scales which are expected to provide more accurate scores. Two of these well-known scales are the ACTFL scale used in the USA and the CEFR used mostly in Europe and elsewhere. Yet, major critiques have emerged from these scales, as to their linearity, and the scales not being appropriate for all learning settings. (see a chapter “► The Common European Framework of Reference (CEFR)” by Barni and Salvati, in this volume).

Shohamy's (2001a) brought up pleas for developing more democratic views of tests, increasing the responsibility of testers, minimizing the power of tests, protecting test takers, and posing a questions about the ethical roles of language testers.

One immediate outcome of the CLT was the development of a *Code of Practice* to protect test takers. Davies (this volume), who was very attentive to the notion of CLT, examined the professionalism of language testers who design tests and overlook their impacts. He posed ample questions about what it means to be an ethical and professional tester and their responsibilities. Davies served as the chair of the ILTA (International Language Testing Association) committee that developed the *Code of Ethics* and a *Code of Practice* to be used by language testers in the development and uses of tests, so testers become aware of their professional and ethical responsibilities. The real aim accordingly was to create tests which are more fair, considerate, constructive, and ethical in terms of their power.

---

## Work in Progress

Over the years, a large number of questions emerged that have fallen under the paradigm of CLT. With the introduction of language citizenship tests for immigrants in an expanding number of countries in Europe, Asia, and elsewhere, ample studies pointed to the harmful effects of those tests. Milani (2008), based on protocols about integration in the Swedish Parliament, pointed to the debates about the tests revealing a taste of discrimination, given the goal of integrating immigrants into the main society. A special issue of the journal *Language Assessment Quarterly* (Shohamy and McNamara 2009) focused on these tests in a number of countries such as Estonia, Latvia, the UK, the USA, and Israel. A number of comprehensive edited books (Stevenson 2009; Extra et al. 2009) were published as well. Unfortunately, this research did not yield major changes in terms of government policies, and the

problem continues as more countries adopt these tests. Recently, Norway is joining these countries with the introduction of new citizenship tests as of January 2017. These policies get stricter as the wave of immigration expands in Europe and elsewhere. Likewise in schools, while tests such as the ones mandated by the NCLB act ceased to exist in the USA, the new policy of the Common Core represents a new testing policy with higher cognitive demands introduced in schools, thus creating injustices for immigrants (Abedi 2001, 2004; Abedi and Dietal 2004; Shohamy and Menken 2015). Thus, these tests, as Valdés and Poza point out, discriminate against newcomers and minority groups (see their chapter “► [Assessing English Language Proficiency in the United States](#)” in the present volume).

At the same time, the work on CLT continues (see the chapter “► [Washback, Impact, and Consequences Revisited](#)” by Tsagari and Cheng in this volume). Indeed, the notion of the “power of tests” puts enormous responsibility on the shoulders of those who wield the tests. Yet, at the same time there are also new approaches that attempt to respond to the power of tests, to minimize and challenge it by focusing on tests geared for more effective learning rather than tools for punishment. Further, with the changes toward multilingualism, there is more of an emphasis on the meaning and essence of language in this day and age with regard to globalization, multilingualism, language varieties, and mixture of languages to include immigrants and minority groups in different types of multilingual tests.

These directions responded to questions such as the following:

- Do the tests reflect the bi-/multilingual uses of language in this day and age in the context of plurilingual societies?
- Do tests have realistic goals in terms of their levels of proficiency, considering the dynamic and fluid nature of language?
- Are the validation procedures based on realistic norms and not on the native speaker?
- Do they consider all components that contribute to performance, beyond language per se?
- Are we ethical when we design tests based on definitions and goals provided by central agencies?
- Are language tests open to monitoring by society, critiqued, and sanctioned?
- How can immigrants and minority groups be included in spite of their language proficiency, given that they are educated, talented, good people, but have difficulties with language, or it takes them long time to acquire it?
- How can immigrants who have to pay big amounts of money for language courses get resources that will help them learn the languages? And is the “almost” native speaker realistic for all people, regardless of age, background, etc. (see the chapter “► [Assessing English as a Lingua Franca](#)” by Jenkins and Leung, which discusses the ELF variety that most English nonnative speakers use, as well as the chapter “► [High-Stakes Tests as De Facto Language Education Policies](#)” by Menken, both in this volume)?
- Do all immigrants need to pass a language test where there is no evidence that knowledge of “the” language necessarily contributes to good citizenship? That is,



how can we minimize the use of language tests, preventing them from being a major tool in creating immigration policy?

In the section below, a list of additional new initiatives which can defy or minimize the power of tests will be briefly described.

---

## Responses: Minimizing and Resisting Power

The topics below include strategies of assessment initiatives and practices which can lead to more positive outcomes of tests which are more fair, just, and mostly educational. These go beyond standardized tests into language testing that proposes means for diverting tests to learning and less for judgment.

**Dynamic assessment:** An approach whereby testing and teaching are connected and hence minimize the power of tests, based on Vygotsky's sociocultural theory whereby the emphasis is the use of tests for learning (see the chapter "► [Dynamic Assessment](#)" by Poehner, Davin, and Lantolf in this volume; and see also Levi 2015; and Levi 2016).

**Assessment literacy** increases the basic knowledge about assessment including CLT and focuses on their consequences and impact, which needs to be addressed and is part of language testing along with other factors (see chapters "► [Language Assessment Literacy](#)" and "► [Training in Language Assessment](#)" by Inbar-Lourie and Malone in this volume).

**Test accommodation and differential item functioning (DIF)** provide a tool for assisting immigrants and minority groups who are not familiar with the new language to obtain assistance and thus enhance the achievement in academic and content subjects, especially for the early years of migration while learning the new language (see chapter "► [Utilizing Accommodations in Assessment](#)" by Abedi in this volume). Further, the focus is on the technique of DIF as a strategy to identify the test items and tasks which discriminate against students of different backgrounds. Removal of such items and tasks results in tests which are more fair to larger pool of test takers.

**Formative/alternative assessment** attempts to develop assessment strategies which are more constructive than standard external items, often developed by local agents at the schools and not by central agencies (see the chapter "► [Task and Performance-Based Assessment](#)" by Wigglesworth and Frost and the chapter "► [Using Portfolios for Assessment/Alternative Assessment](#)" by Fox, in this volume).

**Multilingual/translanguaging and ELF tests.** An approach built on the nature of the language construct as it is being viewed today, where languages are mixed and people use them in very creative ways. Shohamy (2011) demonstrated how the use of multilingual tests in testing mathematics of immigrant students (in Hebrew and Russian on the same test) result in higher mathematics scores than of those students who were tested in monolingual (Hebrew) tests in Israel. A case in point can best be demonstrated with English away from the concept of the native speaker (see the chapter "► [Assessing Multilingual Competence](#)" by Lopez, Turkan, and Guzman-Orth and also the chapter "► [Assessing English as a Lingua Franca](#)" by Jenkins and Leung in this volume).

**Tests for indigenous contexts.** Given that the essence of testing is that it grants importance to languages and provides a message that they should be empowered, there is a call to include indigenous language within the repertoire of testing (see the chapter “► [Language Assessment in Indigenous Contexts in Australia and Canada](#)” by Wigglesworth and Baker in this volume).

**Full language repertoire (FLR).** This refers to expansion of assessment to include all the languages a person knows, regardless of each language level of proficiency. This is especially relevant with immigrant students who arrive in a new location, so the languages they know from the past will not be overlooked and ignored, but rather they should be incorporated into the whole language repertoire, viewing these languages as significant resources.

Other themes included in this volume that have the potential to reduce the power of tests and focus more on learning include the following: the chapter “► [Assessing Students’ Content Knowledge and Language Proficiency](#)” by Llosa recommending a focus on content and less on language proficiency, the chapter “► [Culture and Language Assessment](#)” by Scarino with emphasis on culture within assessment, and qualitative methods of validation by Lazaraton. Chapter “► [Assessing the Language of Young Learners](#)” by Bailey especially warns about the overuse of tests with regard to young learners. Other studies demonstrate the extent to which language tests are instrumental for control. Tsagari and Cheng show how significant it is to examine the consequences of tests so to limit their powerful status, which is related to the chapter “► [High-Stakes Tests as De Facto Language Education Policies](#)” by Menken demonstrating that tests should avoid dictating the curriculum but rather reflect it. These are manifested in the use of tests as “de facto” curriculum, approaches which should be minimized. All these are warning signs regarding the ethics, professionalism, rights, and codes as described in the chapter “► [Ethics, Professionalism, Rights, and Codes](#)” by the late Alan Davies. The existence of the Common European Framework of Reference (CEFR) requires testers to be even more cautious about the power of tests as these scales provide an extra tool to bring about homogeneity, as can be seen in the chapter “► [The Common European Framework of Reference \(CEFR\)](#)” by Barni and Salvati. The dangers of using tests which are not appropriate to specific students are being critiqued by Poza and Valdés (chapter “► [Assessing English Language Proficiency in the United States](#)”).

All in all, many of the chapters in this volume discuss and propose a number of ways to focus on learning and hence to minimize the power of tests by using the strategies described above and in many of the chapters.

---

## Future Directions

CLT led to ample questions about the quality of tests, their consequences, and the difficulties they impose on test takers and systems. Tests often offer simplistic solutions for complex issues. The research in this field attempted to explore areas where tests are misused by examining their consequences and the intentions of those who introduced them. The responses are varied so that what is considered negative

or positive is constantly debated given Messick's views that these are related to the values of test takers, educational systems of nations, and ideologies of governments and regimes to use tests for power and control. Yet, the obligation of those engaged in language testing is to adopt CLT approaches to try to look beyond the tests themselves and toward their uses; in other words, a good test may be necessary but not sufficient. It is the obligation of all those working in test development and use to constantly ask questions as to intentions and uses of tests in education and society with regard to the multiple groups for whom national languages are second languages.

It is encouraging to see that in the past decade a number of different types of assessment strategies and procedures have been developed and implemented. These strategies are currently being used to broaden the construct of testing tests and provide successful ways of "talking back" to the power of tests that can minimize their power and protect test takers and parents, teachers, and principals, enhancing the uses of assessment procedures to minimize discrimination and marginalization and maximize learning, fairness, ethicality, equality, and justice. The purpose is not to eliminate tests but rather to see the values behind them as well as their hidden agendas in the area of accountability and the learning of languages and to reflect perspectives of languages in this day and age.

---

## Cross-References

- ▶ [Assessing English as a Lingua Franca](#)
- ▶ [Assessing Multilingual Competence](#)
- ▶ [Ethics, Professionalism, Rights, and Codes](#)
- ▶ [History of Language Testing](#)
- ▶ [High-Stakes Tests as de facto Language Education Policies](#)

---

## Related Articles in the Encyclopedia of Language and Education

- Linda von Hoene: [The Professional Development of Foreign Language Instructors in Postsecondary Education](#). In Volume: Second and Foreign Language Education
- Bonny Norton, Ron David: [Identity, Language Learning and Critical Pedagogies in Digital Times](#). In Volume: Language Awareness and Multilingualism
- Alastair Pennycook: [Critical Applied Linguistics and Education](#). In Volume: Language Policy and Political Issues in Education
- Hilary Janks, Rebecca Rogers, Katherine O'Daniels: [Language and Power in the Classroom](#). In Volume: Language Policy and Political Issues in Education
- Stephen May: [Language Education, Pluralism, and Citizenship](#). In Volume: Language Policy and Political Issues in Education

## References

- Abedi, J. (2001). *Assessment and accommodations for English language learners: Issues and recommendations* (CRESST Policy Brief 4). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4–14.
- Abedi, J., & Dietal, R. (2004). *Challenges in the no child left behind act for English language learners* (CRESST Policy Brief 7). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Allderson, J. C., & Wall, D. (1993). Does washback exist? *Applied linguistics*, 14(2), 115–129.
- Blackledge, A. (2009). “As a country we do expect”: The further extension of language testing regimes in the United Kingdom. *Language Assessment Quarterly*, 6(1), 6–16.
- Bourdieu, P. (1991). *Language and symbolic power* (trans: Gino Raymond and Matthew Adamson). Cambridge, MA: Harvard University Press.
- Cheng, L. (2004). The washback effect of a public examination change on teachers’ perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 147–170). Mahwah: Lawrence Erlbaum Associates.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3–18). Mahwah: Lawrence Earlbaum Associates.
- Cheng, L., Watanabe, Y. & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah: Lawrence Erlbaum Associates. Code of Practice.
- Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment*, 16(2), 104–122.
- Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students’ and parents’ views of SBA in Hong Kong. *Language Testing*, 28(2), 221–249.
- Collier, V., & Thomas, W. (2002). Reforming education policies for English learners means better schools for all. *The State Education Standard*, 3(1), 30–36.
- Davies, A. (1997). Demands of being professional in language testing. *Language Testing*, 14, 328–339.
- De Jong, J., Lennig, M., Kerkhoff, A., & Poelmans, P. (2009). Development of a test of spoken Dutch for prospective immigrants. *Language Assessment Quarterly*, 6(1), 41–60.
- Eades, D. (2009). Testing the claims of asylum seekers: The role of language analysis. *Language Assessment Quarterly*, 6(1), 30–40.
- Eribon, D. (1992). *Michel Foucault* (trans: Betsy Wing). Cambridge, MA: Harvard University Press.
- Evans, B., & Homberger, N. (2005). No child left behind: Repealing and unpeeling federal language education policy in the United States. *Language Policy*, 4(1), 87–106.
- Extra, G., Spotti, M., & van Avermaet, P. (Eds.). (2009). *Language testing, migration and citizenship: Cross-national perspectives*. London: Continuum.
- Foucault, M. (1979). *Discipline and punish: The birth of the prison* (trans: from the French by Alan Sheridan). New York: Vintage Books.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational researcher*, 18(9), 27–32.
- Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1(4), 253–266.
- Government Accountability Office. (2006). *No child left behind act: Assistance from education could help states better measure progress of students with limited English proficiency*. Washington, DC: Author.

- Gysen, S., Kuijper, H., & van Avermaet, P. (2009). Language testing in the context of immigration and citizenship: The case of the Netherlands and Flanders. *Language Assessment Quarterly*, 6(1), 98–105.
- Hanson, F. A. (1993). *Testing testing: Social consequences of the examined life*. Berkeley: University of California Press.
- Inbar-Lourie, O., & Shohamy, E. (2009). Assessing young language learners: What is the construct? In M. Nikolov (Ed.), *Contextualizing the age factor: Issues in early foreign language learning*. Berlin/New York: Mouton de Gruyter.
- Kunnan, A. J. (2009). Testing for citizenship: The U.S. naturalization test. *Language Assessment Quarterly*, 6(1), 89–97.
- Levi, T. (2015). Towards a framework for assessing foreign language oral proficiency in a large-scale test setting: Learning from DA mediation examinee verbalizations. *Language and Sociocultural Theory*, 2(1), 1–24.
- Levi, T. (2016). Developing L2 oral language proficiency using concept-based Dynamic Assessment within a large-scale testing context. *Language and Sociocultural Theory*, 3(2), 197–220.
- Levin, T., & Shohamy, E. (2008). Achievement of immigrant students in mathematics and academic Hebrew in Israeli school: A large-scale evaluation study. *Studies in Educational Evaluation*, 34(1), 1–14.
- Levin, T., & Shohamy, E. (2012). Understanding language achievement of immigrants in schools: The role of multiple academic languages. In M. Leikin, M. Schwartz, & Y. Tobin (Eds.), *Current issues in bilingualism: Cognitive and socio-linguistic perspectives add page numbers* (pp. 137–155). Springer: Literacy Studies.
- Levin, T., Shohamy, E., & Spolsky, B. (2003). *Academic achievements of immigrants in schools, Report submitted to the Ministry of Education (in Hebrew)*. Tel Aviv: Tel Aviv University.
- Levin, T., Shohamy, E., & Inbar, O. (2007). Achievements in academic Hebrew among immigrant students in Israel. In N. Nevo & E. Olshtain (Eds.), *The Hebrew language in the era of globalization* (pp. 37–66). Jerusalem: Magnes Press, the Hebrew University.
- Madaus, G. (1990, December 6). *Testing as a social technology*. Paper presented at the Inaugural Annual Boisi Lecture in Education and Public Policy, Boston College.
- McNamara, T. (1998). Policy and social considerations in language assessment. *Annual Review of Applied Linguistics*, 18, 304–319.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell.
- McNamara, T., & Shohamy, E. (2008). Language tests and human rights. *International Journal of Applied Linguistics*, 18(1), 89–95.
- Menken, K. (2006). Teaching to the test: How standardized testing promoted by *No Child Left Behind* impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521–546.
- Menken, K. (2007). High-stakes tests as de facto language policies in education. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education, Language testing and assessment* (Vol. 7, pp. 401–414). Netherlands: Kluwer.
- Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Clevedon, Avon: Multilingual Matters.
- Menken, K. (2010). *No Child Left Behind* and English language learners: The challenges and consequences of high-stakes testing. *Theory into Practice*, 49(2), 121–128.
- Menken, K. (2013). Restrictive language education policies and emergent bilingual youth: A perfect storm with imperfect outcomes. *Theory into Practice*, 52(3), 160–168.
- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher*, 10, 9–20.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 447–474). New York: ACE/Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *ETS Research Report Series, 1996* (1), i–18.

- Milani, T. M. (2008). Language testing and citizenship: A language ideological debate in Sweden. *Language in Society*, 37(01), 27–59.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142–173.
- Schissel, J. (2012). *The pedagogical practice of test accommodations with emergent bilinguals: Policy-enforced washback in two urban schools* (Unpublished doctoral dissertation). Philadelphia: University of Pennsylvania.
- Schupbach, D. (2009). Testing language, testing ethnicity? Policies and practices surrounding the ethnic German *Aussiedler*. *Language Assessment Quarterly*, 6(1), 78–82.
- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing*, 14, 340–349.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24, 331–345.
- Shohamy, E. (2001a). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Pearson Education.
- Shohamy, E. (2001b). Democratic assessment as an alternative. *Language Testing*, 18(4), 373–391.
- Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches*. Abington: Oxon Abingdon.
- Shohamy, E. (2009). Language tests for immigrants: Why language? Why tests? Why citizenship? In G. Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Eds.), *Discourses on language and integration* (pp. 45–59). Amsterdam: John Benjamins.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *Modern Language Journal*, 95(3), 418–429.
- Shohamy, E. (2015). *Critical language testing and English Lingua Franca: How can one help the other? Waseda Working Papers in ELF (English as a Lingua Franca)* (Vol. 4, pp. 37–51). Waseda ELF Research Group Waseda University: Tokyo.
- Shohamy, E., & Kanza, T. (2009). Citizenship, language, and nationality in Israel. In G. Extra, M. Spotti, & P. van Avermaet (Eds.), *Language testing, migration and citizenship: Cross-national perspectives*. London/New York: Continuum.
- Shohamy, E., & McNamara, T. (2009). Language tests for citizenship, immigration and asylum. *Language Assessment Quarterly*, 6(1), 1–5.
- Shohamy, E., & Menken, K. (2015). Language assessment: Past and present misuses and future possibilities. Bi-multi-lingual assessment. The Routledge handbook on bilingual education. In W. E. Wright, S. Boun, & O. Garcia (Eds.), *The handbook of bilingual and multilingual education* (1st ed.). London/New York: Routledge.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298–317.
- Solano-Flores, G., & Trumball, E. (2003). Examining language in context: The need for new research paradigms in the testing of English-language learners. *Educational Researcher*, 32(2), 3–13.
- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.
- Stevenson, P. (Ed.) (2009). 'National' languages in transnational contexts: Language, migration and citizenship in Europe. In: *Language ideologies, policies and practices: Language and the future of Europe* (pp. 147–161). London: Palgrave Macmillan.
- Valdes, G., & Figueroa, R. (1996). *Bilingualism and testing: A special case of bias*. Norwood: Ablex.
- Valdés, G., Menken, K., & Castro, M. (2015). *Common Core, bilingual and English language learners: A resource for educators*. Philadelphia: Caslon Publishing.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41–69.