

---

## Assessing Multilingual Competencies: Adopting Construct Valid Assessment Policies

ELANA SHOHAMY

*Tel Aviv University*

*School of Education*

*Tel Aviv, Israel 69978*

*Email: elana@post.tau.ac.il*

---

All assessment policies and practices are based on monolingual constructs whereby test-takers are expected to demonstrate their language proficiency in one language at a time. Thus, the construct underlying these assessment approaches and/or scales (e.g., the CEFR) is of language as a closed and finite system that does not enable other languages to “smuggle in.” This view is in stark contrast to the current understanding of multilingual competencies for which various languages and aspects “bleed” into one another in creative ways as manifested by a growing number of users, especially immigrants, who are born into one language and acquire additional language(s), resulting in multilingual competencies. This is manifested in codeswitching and in the simultaneous use of different language functions (e.g., reading in one and speaking in another in the process of academic functioning). Yet, this multilingual functioning receives no attention in language testing practices. Further, multilingual users who rarely reach language proficiency in each of the languages that is identical to that of their monolingual counterparts are always being compared to them and thus receive lower scores. Consequently, they are penalized for their multilingual competencies, sending a message that multilingual knowledge is a liability. Given the current policies of cultivating multilingualism in schools and societies as expressed in the articles in this special issue, I critique the current monolingual assessment approaches within a political and social context. I argue that these approaches are rooted in nation-state ideologies that are still attempting to promote national collective agendas of “wishful thinking” and ignore the reality of how languages are being used. This is followed by empirical data pointing to the cost of the continued use of monolingual tests for individual students, especially those who are bilingual, as is the case with immigrants. All of these will lead to initial proposals and examples for the adoption of different types of multilingual testing and assessment policies and practices in various contexts. These approaches, I argue, are more construct valid, as they enable the manifestation of fuller knowledge in integrated ways, thus highlighting the advantages, rather than the problems, that multilingual users possess.

---

BETWEEN 1999 AND 2003, T. LEVIN, B. Spolsky, and I worked on a research project for the Israeli Ministry of Education, the aim of which was to compare the academic achievements of immigrant students from the former USSR and Ethiopia with Israeli Hebrew native speakers. The

main aim of the research was to answer the question of “how long,” that is, how many years it took immigrant students to close the gap between them and the Israeli-born students. The tests used for the comparison were all in Hebrew, the dominant official language of Israel and the language of instruction in all Israeli Jewish schools. As expected, the results of the study pointed to large and significant gaps between the immigrant students and the native Israelis. It takes about 9–11 years for students from the former USSR to close the gap, whereas the Ethiopian students never close it, not

even in the second generation. Eight years later, I rethink the ethical dimensions of the comparative study and my own compliance with research, the aim of which was to compare groups of incomparable conditions. Carrying out these tests in a language that the immigrants have not yet acquired is likely to yield lower scores, which will lead to wrong conclusions regarding their true levels of academic achievements. Although it is clear that the immigrant students did acquire vast amounts of academic knowledge prior to migration, it is not possible to access that knowledge via Hebrew tests. The lack of Hebrew proficiency masks that picture. For these students who acquired academic knowledge via Russian or Amharic, these types of comparisons are of low validity and fairness, as the students do not have the appropriate channel to demonstrate their academic knowledge. Such results are also likely to have effects on their identity, confidence, and self-concept. Using such deficient/subtractive research designs denies immigrants the fair opportunity to demonstrate their true academic knowledge. These types of studies are common worldwide and it is easy for researchers to fall into subtractive research designs, believing that pointing out that the obvious gaps would lead to improved policies toward the groups—they rarely do. The expectation to demonstrate academic knowledge via national languages is the dominant model everywhere. The use of a dominant national language as the means of demonstrating the academic achievements on tests is an example of buying into research designs that fit national language ideologies of one nation, one language, and hence mask the real trait that is the target of the measurement. Researchers and language testers face dilemmas as to whether they should refuse to participate in such studies or whether new methods of assessment should be invented to highlight the true academic knowledge in which multilingual students have more than one way to see beyond just temporary accommodations. This experience is the impetus for the current article. It is about ways to address multiple ways of seeing.

The articles included in this special issue of *The Modern Language Journal* all argue for the expansion of the language learning construct beyond monolingual views of language, toward different forms of multilingualism. There are two main approaches to multilingualism. One approach calls for the legitimacy and encouragement of teaching and learning of multiple languages within the same space (i.e., classrooms, schools, regions). This approach views each language as a closed and homogenous construct and encourages the

teaching of multiple independent languages in the same context. The data used to support this argument build mostly on the positive transfer among multiple languages. The second approach deconstructs the notion of a language as a finite construct, viewing it as an abstract notion that is used as a means for negotiating and creating meanings. Language, accordingly, is made up of hybrids and fusions and nonnative varieties that continuously cross over in creative ways with undefined boundaries and open forms of negotiations (Canagarajah, 2006). Translanguaging, as it is termed by García, Sylvan, and Witt (this issue), is one such example of moving freely within, between, and among languages. These views also follow notions advanced by Makoni and Pennycook (2006) of “disinventing languages” as well as what Li Wei and Martin (2009) refer to as “languages that bleed into one another.” Creese and Blackledge (2010), for example, examined such multilingual varieties of immigrant students in complementary schools in the United Kingdom (see also Li Wei, this issue). There are clearly multiple ways of moving within a multilingual continuum. The unique aspect of these articles is that, together, they recognize and legitimize the multilingual phenomena as part of language education.

Although dynamic, diverse, and constructive discussions of multilingual teaching and learning are currently taking place within the language education field, the phenomenon is completely overlooked in the assessment field that continues to view language as a monolingual, homogenous, and often still native-like construct. There seems to be a lack of coordination between the two disciplines of teaching and testing. This is in spite of the fact that language tests need to build on an updated language construct of what it means to know a language as the fundamental step in the creation of tests of high construct validity. Yet, although there is ample and convincing research about multilingual constructs, it is intriguing to examine the reasons that language tests are still not addressing these issues, especially given their powerful effects on learning and their ability to dictate and perpetuate language realities in educational institutions. Thus, overlooking the construct of multilingualism is likely to result in language tests of limited evidence of validity.

Traditionally, the field of language testing is understood to consist of two major components: one focusing on the “what,” which refers to the constructs that need to be assessed (also known as “the trait”) and the other pertaining to the “how” (also known as “the method”), which addresses the

specific procedures and strategies used for assessing the “what.” The trait is known to be defined by the language learning field, which provides definitions of the essential elements of language that language testers can use to design and develop valid language tests. The “how,” in contrast, is derived mostly from the field of testing and measurement that has, over the years, developed a broad body of theories, research, techniques, and practices about testing and assessment. An examination of the developments in the language testing and assessment discipline since the 1960s reveals, in fact, that its theories and practices have always been closely related to definitions of language proficiency. Thus, discrete-point testing viewed language as consisting of lexical and structural items so that the language test of that era presented isolated items in objective testing procedures. In the communicative era, tests aimed to replicate interactions among language users utilizing authentic oral and written texts, and in the performance testing era, language users were expected to perform tasks taken from “real life language activities” (Shohamy, 2009a).

Accordingly, the introduction and wide use of a multilingual construct in much of the literature of language learning in the past decade should have had a more direct carryover to the field of language testing. Multilingualism should have provided testers with new and broader definitions of the language that needs to be addressed by testers in the creation of language tests. Such tests are needed, given the multilingual practices, especially within the context of migration and globalization. It follows that there is a need to address the multiple versions and perspectives of the “what” of languages, along with the approaches developed for the “how.” Thus, the special focus on multilingualism today and the diverse contexts in which language testing and assessment are currently anchored require the incorporation of these theories into the field of language testing. Testers need to be socially responsible and accountable to ensure ethicality and professionalism. In other words, as language testers seek to develop and design methods and procedures for assessment (the “how”) and its multiple facets and dimensions, they also need to become mindful of the emerging insights regarding the trait (the “what”). This is related to the power of tests, the responsibility that language testers hold, and their central functions in education, politics, and society. It is being realized that language testing is not occurring in homogenous, uniform, and isolated contexts but rather in diverse, multilingual, and multicultural societies, a reality that poses new

challenges and questions to testers with regard to what it means to know language(s) in education and society. Thus, the claim made here is that for tests to be construct-valid, they need to be based on a construct that follows current understandings and theories of language. Given the multilingual approaches to teaching, learning, and understanding of language, language testing policies, procedures, and tasks need to reflect these approaches. I will now attempt to explain what I see as the roots of this continued monolingual view of language tests within a political and social context. This will be followed by some empirical data pointing to the cost of the continued use of monolingual tests for individual students, especially those who are bilingual, as is the case with immigrants. All of these will lead to an initial proposal and recommendation for the adoption of different types of testing policies in various contexts.

#### THE SOCIAL AND POLITICAL DIMENSIONS OF TESTS

In the past two decades, a sociopolitical perspective of tests has been introduced (McNamara & Roever, 2006; Messick, 1994; Shohamy, 2001, 2006, among others). Accordingly, tests are no longer viewed only in scientific and technical terms. Rather, there is strong attention to the uses of tests and their impact on people, education, and society. As a result, language testers are asked to deal with broader issues of the tests they develop and to examine their impact and consequences as well as to understand the motivations for introducing tests. Tests, then, are not only viewed anymore as naïve measurement tools, but also as powerful devices contextualized within broad social, political, and economic realities. Tests and language go along with the arguments about the political dimensions of language itself, used mostly by central authorities as an ideological tool for the creation of national and collective identities (Ricento, 2006; Shohamy, 2006). The use of monolingual tests needs to be contextualized within this political and social reality in which they operate.

Take for example the introduction of language tests for citizenship in a growing number of countries in the past decade. It has been argued that these tests are used as devices to gatekeep immigrants and asylum seekers and thus to exclude people whom the state is not interested (Extra, Spotti, & Van Avermaat, 2009; Hogan-Brun, Mar-Molinero, & Stevenson, 2009; Shohamy,

2009b). Other examples include the introduction of tests in educational systems often motivated by an agenda of perpetuating certain political ideologies. These tests are given in monolingual national languages. It is understood today that language tests are used as disciplinary tools to create and impose language ideologies and policies according to the agendas and authority of the nation-state. It is the power of tests that enables those in authority to use them as a major tool to perpetuate such agendas, given that those who are affected by these tests comply with the test demands, given the tests' consequences on their personal lives (Shohamy, 2001). Tests, then, serve as media through which messages regarding nationalism are being communicated to test-takers, teachers, and educational systems regarding language priorities. Thus, by conducting language tests in a given language, messages are being transmitted regarding the priority of dominant languages while marginalizing others, in line with national ideologies and agendas.

It is within this widespread practice of the use of tests as political tools by government and central agencies that the use of monolingual tests needs to be interpreted. The national ideologies of most nation-states worldwide are still driven by the promotion and perpetuation of "one nation, one language" for the sake of a national collective identity, a phenomenon that is even more dominant, given the large number of immigrants that move to many European countries and elsewhere. National languages still serve as symbols and devices to promote collective national identities. It is within this context that monolingual tests in national languages serve as institutional tools to perpetuate and impose such ideologies. Although there is some promotion of additional languages (i.e., foreign languages, mostly English), they are viewed as added languages and rarely have an equal status in relation to national languages. Although there are some content-based programs that will promote English, such as Content Language Integrated Learning (CLIL), given the role of English as an international language, this is rarely the case with other languages. In other words, the connection of national ideologies with national dominant languages and the continued role of language as a symbol of identity are manifested in and channeled to monolingual tests.

Thus, although multilingual teaching and learning are currently promoted, encouraged, and practiced, especially in the European context, there are no voices that argue in favor of multilingual tests. Yet, as argued earlier, tests need to reflect such language definitions so as to ensure

construct validity. Even in the cases of immigrant students in schools, who are in the midst of the long process of acquiring the dominant language of the new place they immigrated to, these languages are still being viewed as liabilities and are being overlooked for the sake of learning the preferred powerful and prestigious language of the nation. Rarely do these tests incorporate the languages of "the others" and/or receive any recognition. This is evident not only in the fast-growing policies, as mentioned earlier, where immigrants are forced to pass language tests in national languages in condition of residence and citizenship, but also in the international global tests such as the Programme for International Student Assessment (PISA). This test requires all students to perform exclusively via the national languages of the countries. Finally, it is important to examine the widespread instruments, such as the Common European Framework of Reference for Languages (CEFR) and the American Council on the Teaching of Foreign Languages (ACTFL) guidelines, as the main criteria and yardsticks for judging the quality of the languages produced on tests. These are based exclusively on monolingual homogeneous constructs that forbid any other languages to "smuggle in" or "bloom."

The conclusion of the aforementioned discussion is that given their power, monolingual tests in national languages serve as tools that perpetuate the monolingual *de facto* policies of the nation. These policies, in return, have detrimental effects on learning. Thus, not only are the existing tests not congruent with the various multilingual approaches, they in fact work against these approaches (Cheng, 2009; Davies, 2009; McNamara & Roever, 2006; Menken, 2009; Shohamy, 2001; Spolsky, 2009).

#### SOME DATA REGARDING THE COST OF MONOLINGUAL TESTS

As is the practice worldwide, immigrant students are being compared with native speakers on their academic achievement via tests in the monolingual dominant language. Such is the case with the No Child Left Behind Act (2001) in the United States, in which all students are expected to be tested on their academic proficiency. This includes immigrant students who are in the process of acquiring the dominant languages of their new American residence and at the same time are tested in English in the short time after their arrival in the United States. These second-language learners, it is argued, are bound to be marginalized in comparison to native speakers, given the

length of time it takes to acquire a new language. The main point made here is that as long as such national tests will require nonnatives (i.e., immigrant students) to be like natives and to be measured on monolingual tests in the dominant language, these students will be not be able to achieve valid scores. As a result, hierarchies of immigrants and natives are being created and perpetuated, and messages regarding marginality are being delivered and reinforced. In other words, these monolingual tests bear costs for multilingual test-takers, as will be demonstrated in the empirical evidence presented next.

First, monolingual tests overlook various research findings that demonstrate that immigrants continue to employ their first languages (L1s) in various academic literacy situations, a long time after immigration, often even for a lifetime (Thomas & Collier, 2002). The languages that immigrants possess upon arrival in the countries to which they immigrate do play important roles in providing cognitive and educational advantages. Immigrants bring with them knowledge of the world, varied content, experiences, and interpretations. For example, in the domain of speaking, immigrants use mixed codes of L1 accents, grammar, lexicon, syntax, discourse, and pragmatics. Similarly, in writing, language users continue to use L1 syntax and vocabulary so much so that they almost always need to employ editors to improve and standardize their writing so it will be on par or close to that of native speakers. Clearly, the use of a mixed code in writing and in other areas is still considered a liability. In the domains of listening, contextualized information and schema provided in the L1 are often needed in order to process oral information. Finally, in reading, information about history, culture, and politics is needed for comprehension of most texts. Additional sources and resources that rely on multimodal sources are constantly in use in the form of images, sounds, and other clues. Even when decoding is practiced and words are read well, there is a need for cultural and contextual familiarity. A study by Haim (2010) demonstrated how significant transfers occur within three languages, from the L1 to the second language (L2) to the third language (L3) within Russian, Hebrew, and English (see also the article by Cenoz & Gorter, "Focus on Multilingualism," this issue). Strong predictions were found within the three languages in reading and writing according to years of residence. Clearly, the L1 plays a role, as well as the L2 and the L3. All these language performances need to be incorporated into multilingual tests.

In our own research (Levin, Shohamy, & Spolsky, 2003; Levin & Shohamy, 2008), as mentioned in the preface to this article, immigrant students, irrespective of age, time of migration, and length of residence, are being compared with students born in Israel and are measured on the same yardstick of monolingual Hebrew test. Yet, multilingual users rarely reach language proficiency identical to those born into one language, as they are not on par with the native speakers. The very question as to whether immigrants can ever be compared to native speakers, when the test is conducted in a language that will take them a long time to acquire, requires special attention. There seems to be a need to address the channel through which these students can best demonstrate the knowledge of content areas they possess. By ignoring it, the very academic competence gained by these students in their previous years of study in their countries of origin, where they obtained most of their academic learning, is overlooked. Yet, this knowledge is certainly part of their academic knowledge, regardless of the channel through which they can demonstrate it. Figures 1 and 2 display results that point to the gaps between immigrants and native speakers. It is shown that it takes immigrants from the former USSR 9–11 years to achieve the same academic level of a native speaker in mathematics and in Hebrew (Levin & Shohamy, 2008). These findings were identical for all the three grades tested (5th, 9th, and 11th), whereas immigrants from Ethiopia could never achieve such high levels as the native speakers, not even in the second generation.

Yet, the main point raised here is whether this is even a legitimate question to ask, as it is clear that immigrants cannot perform as native speakers in a language they do not know. Moreover, immigrant students continue to interpret the knowledge via two languages for a long period of time after immigration. Figure 3 thus portrays the results of a study in which students from the former USSR were tested in a bilingual Hebrew–Russian test. In this test, questions and distracters were presented in two languages versus a control group, which was tested in Hebrew only. The results showed that the students who received the tests in the two languages significantly outperformed those students who were subject to the monolingual "Hebrew-only" test, an advantage that lasted 12 years after migration.

Immigrant students therefore continue to rely on the language they acquired before immigration as a valuable resource for processing the new academic knowledge. Yet, on the tests that are administered in most educational systems, this

FIGURE 1  
Grade 9 Hebrew Standard Scores, According to Years of Residence

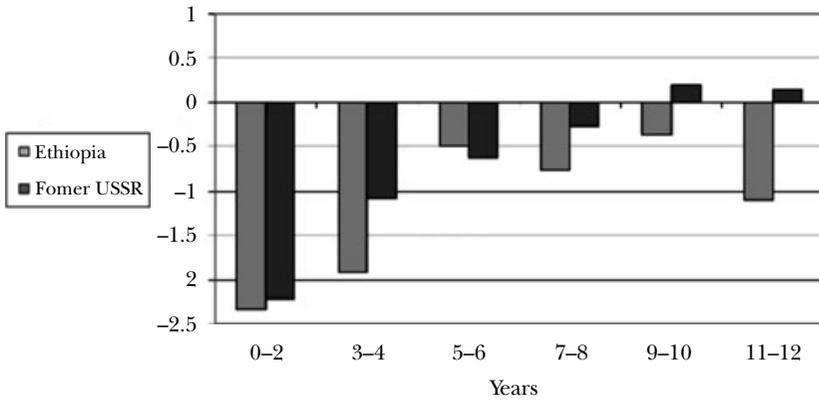
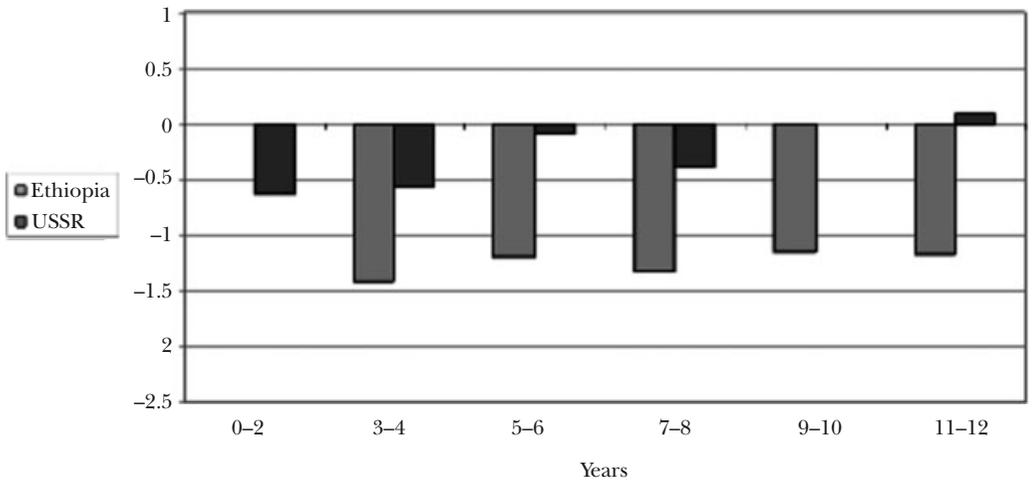


FIGURE 2  
Grade 11 Standard Scores, According to Years of Residence



knowledge is denied them. Wright and Xiaoshi (2008) demonstrated, for example, how students who performed well in math in their home countries performed poorly in math in the new country due to the new language. The knowledge was there, but the channel of expressing it was missing. Figure 4 demonstrates the notion that immigrants view academic knowledge via two perspectives—their own home language and the new language(s). However, upon immigration, they are expected to see with one eye only. Moreover, that eye happens to be the weak one.

In Figures 5–7, we can see further evidence as to the degree of the obstacle that the new language poses to academic achievement. In the following

experiments, performance of students on math tests was compared in three conditions. One included math tests that included symbols versus words. Specifically, it was shown that the use of symbols such as graphs and images on a test, instead of words, increased the performance levels of students in math. The results are presented in Figure 5. It can be seen that when students were presented with visual and symbolic signs in relation to just verbal information, they performed significantly better.

Lack of familiarity with the topics of the texts used on academic achievement tests provides yet another source that penalizes immigrant students on academic tests. In Figure 6, we can see that

FIGURE 3  
Math Scores, According to Years of Residence: In the Two Conditions

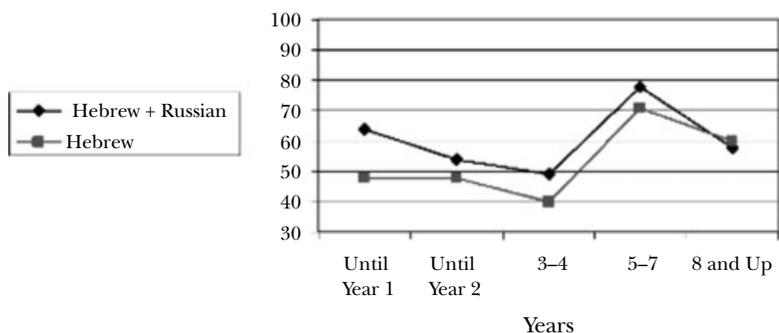


FIGURE 4  
A Portrayal of Pushing Immigrants to Use One Language Instead of Two



FIGURE 5  
Visuals and Symbolic Versus Verbal

Modes	Grade 5			Grade 9			Grade 11		
	IS	FSU	ETH	IS	FSU	ETH	IS	FSU	ETH
Verbal	53	47	30	43	42	15	57	44	22
Visual	62	56	39	51	51	21	48	40	25
Symbolic	71	65	45	61	63	34	67	50	32

Note. IS = Israel; FSU = Former Soviet Union; ETH = Ethiopia.

FIGURE 6  
Familiar Content

Familiar Topics	Israel (IS)	Former USSR (FSU)	Ethiopia (ETH)
*1st Language Maintenance	62	55	49
*Difficulties in Immigration	69	56	52
Less Familiar Topics	58	37	38

when students were tested on familiar topics, such as immigration and language learning, they performed significantly better, in comparison to the identical tests which contained questions on unfamiliar topics.

Another source of difficulty for immigrant students is related to cognitive processing. In Figure 7, we can see that immigrant students, in all grades, performed significantly better when they were provided with cognitive guidance for processing math texts when it was presented in the dominant language being acquired.

Through statistical procedures such as Differential Item Functioning (DIF), it is possible to identify the specific items that discriminate (i.e., differentiate) against students of different backgrounds, such as different language backgrounds. For example, in a study by Levi-Keren (2008), it was shown that by applying the DIF statistical methods and analyses, the very test items that differentiated negatively (and positively) between students from the USSR and Ethiopia and native speakers were tracked and identified. This was

FIGURE 7  
Cognitive Guidance

	Grade 5			Grade 9			Grade 11		
	IS	FSU	ETH	IS	FSU	ETH	IS	FSU	ETH
Process + Product	65	58	40	36	41	18	44	33	20
Product	57		30	22	23	04	32	28	11

Note. IS = Israel; FSU = Former Soviet Union; ETH = Ethiopia.

accompanied by a follow-up study that employed “think-aloud” protocols to identify the specific cultural and cognitive elements that caused these students to perform better or worse. The findings identified factors such as lack of background knowledge, unfamiliar topic, lexicon, and cognitive processing that are responsible for low performance. It is expected that such results will feed into models that will be instrumental in developing tests that are less biased against these students. Further work currently being conducted attempts to follow the process by which immigrants process monolingual texts, often showing that they do that using both the L1 and the L2 (Logan-Terry & Wright, 2010). Such studies are relevant to various types of multilingual students, immigrants, as well as various types of minority students and those who are born into one language, function in a number of other languages, and possess multilingual competencies throughout their lives.

The aforementioned data and arguments demonstrate that the widely used tests that are based on a monolingual construct result in biased and often discriminatory scores, as many factors are not incorporated. Often, these monolingual tests are also based on the native variety of the language, whereas it is clear here that very few L2 learners can ever reach such a standard. This phenomenon is even more relevant with regard to English, with its many native and nonnative varieties, as it is used as a lingua franca worldwide. Even in these cases, the national L1s of immigrants are generally being viewed as intrusions to the acquisition of Standard English, as is the requirement in most English language tests.

One context in which multilingual tests are recognized is in test accommodations. This context allows for immigrant students to obtain assistance in performing on academic tests in their L1, such as with the aid of a dictionary, translation of vocabulary items, and translations in the body of the text. The rationale behind the accommodation

policy is that they are needed on a temporary basis until the ultimate goal of perfect monolingualism is reached (Abedi, 2004, 2009). Yet, this very policy implies that immigrant students can perform bilingually until they start ignoring their L1 or ignore bilingualism and adopt monolingualism again in a different language. However, if, as the data suggest here, students in fact continue to use their L1 for a long time, what is the real role of accommodation? One also wonders if the use of accommodations does, in fact, contribute to monolingualism. Current research on school learning examines various types of accommodation; there is a trend to view accommodations as reflecting the complex construct of processing test items for L2 learners. Thus, when a student processes test items using the L1 and L2, it is viewed as a more valid construct and not as the “route to monolingualism.” Therefore, accommodations need to be viewed as integrated components of academic performance of bilingual students as L2 students continue to process language in two languages, as discussed earlier. This indicates that the view that reliance on L1 is only a temporary phase, used for a limited time only, should be rethought. There is an urgent need to think of a new term that will address the reality of how L2 learners process academic language on a permanent basis rather than as a temporary one. Accommodations need to reflect the wealth of factors that play a role in academic processing, not as temporary help but as an integral component of the newly defined language construct.

Given the aforementioned findings and arguments, a number of ideas for multilingual tests are proposed. These address multilingual competencies that many language speakers possess. Whether in the context of migration, minority groups using multiple languages, or any other multilingual situations of students acquiring additional languages in school learning, these competencies need to be incorporated in new ways of using tests. These tests will more accurately reflect the reality of the multilingual construct. Although the principle of the multilingual construct is important, it clearly is different in diverse language learning contexts. Yet, in all cases, these types of tests are likely to enable manifestation of true knowledge and will enhance bilingualism and multilingualism as advantages rather than as liabilities.

As was noted earlier, a number of reasons can account for this dominant phenomenon. One is that large-scale testing is ideological and political, so that tests are used as tools that define and perpetuate the varieties that the nation and its educational system echo and voice and are

FIGURE 8  
Bilingual Construction of a Recipe

**Class Recipe Book**
**Task**

**My Recipe for:**  
chocolat cookies

**Bushuk**  
**Ingredients:**

1. shogor -
2. sult -
3. chocolat -
4. Water -

**Kitchen Tools:**

1. tavhit -
2. churot plastik -
3. hot fahur -
- Kebara -

**What to do:**

1. Take the water and shfahit in the kebara.
2. Take the chocolat and shfahit in the kebara.
3. Take the shogor and take sult and shpochit in the kebara too.
4. Take the Meharbel and tearbol.
5. Take the isa and shpoch it on th Bushuk.
6. Take the churot.

**Picture:**



in the tabor

believed to serve the system. Then the testers themselves, mostly contracted and employed by national institutions, provide the devices to pursue and impose these agendas. Thus, testers are viewed mostly as technicians who follow the agendas of the nations and provide the technical dimensions of tests (i.e., items, analysis, reliability) while not arguing with the construct. In most cases, test institutions are branches of central governments or are contracted by them and thus help maintain these ideologies. In other words, testers do not design tests that are based on the reality of how languages are really used and learned by L2 learners, minorities, and even foreign lan-

guage learners, but rather take the side of the ideological institutions and therefore serve as servants of the ideology that is in charge of one of the most powerful devices in modern societies: tests. Testers choose to overlook the diversity and complexity of the language construct, what languages are, and continue to produce monolingual tests that reflect ideologies of those who decide how languages should be used.

Yet, language testers cannot just stand aside and ignore the role that tests play in creating monolingual policies. They should assume responsibility for their tests, especially because tests determine the prestige and status of

languages and are capable of suppressing language diversity. This finding was shown in a number of studies such as that of Menken (2008a, 2008b) and others who pointed to the phenomenon whereby the introduction of monolingual tests, such as through NCLB, brought about a growing number of monolingual programs giving increased attention to the dominant languages and suppressing other languages that had been studied and used until the tests were introduced. Thus, the use of native-like tests tends to standardize languages and perpetuate correctness and purity. The main argument here is that as long as language tests continue to be based on a limited and narrow view of language, these powerful tests will continue to contribute to a limited and narrow view of language.

### SOME PROPOSALS FOR MULTILINGUAL ASSESSMENT

How can language tests best reflect the wealth of factors that play a role in academic processing and thus be more congruent with the various types of the multilingual constructs?

In principle, when tests adopt a broad and expanded view of assessment, such as when assessment is used for diagnostic, feedback, and learning purposes beyond the narrow view of testing as it is used in large-scale tests, then any multilingual pedagogical strategy can be used as an assessment procedure. Thus, teaching and assessment are totally integrated and inseparable. One is embedded in the other and there are no clear and distinct divides. This means that testing leads to teaching that brings about further assessment based on diagnosis, as is argued by the theories of dynamic assessment (Poehner, 2008). In these cases, the multilingual approaches are integrated in multilingual assessment tasks in natural and organic ways.

Yet, as to large-scale assessment when testing is totally separable from teaching, which is the main critical focus of this article, one needs to follow the multilingual approaches as outlined in this issue and as outlined earlier here. It was claimed that there are two types of multilingual approaches: one where a number of homogenous languages are taught in one space, whereas the other focuses on the integration and mixing of a number of languages. These two approaches will have direct manifestations as to the strategies of testing and can actually be put on a continuum.

On one side of the continuum, language *x* is used for certain purposes, such as reading, whereas language *y* is used for writing, and lan-

guage *z* is used for discussions. In many classes in Arab schools in Israel, a text in history is read in English, students are asked to summarize the article in Hebrew, and the oral discussion takes place in Arabic, either Modern Standard or a spoken dialect. The extent to which these three (or four, including the Arabic varieties) languages, used in the same space, are actually kept homogenous and separate is probably too idealistic and unreal. On the other extreme side of the continuum, tests are based on the approach in which a mixture of languages and open borders among them is a recognized, accepted, and encouraged variety. This approach can be translated into multilingual tasks for which it is understood by the test-takers that mixing languages is a legitimate act that does not result in penalties but rather is an effective means of expressing and communicating ideas that cannot be transmitted in one language (see also Canagarajah, this issue). Figure 8 displays such a writing task—in this case, it is writing a recipe. The test-taker who authored the text wrote a very accurate recipe utilizing Hebrew and English. In cases like this, a monolingual rating scale would have penalized the student, as it violates the descriptions of “interference from first language.” What is viewed in monolingual tests and criteria as interference is viewed here as a more effective way to transmit language.

The other example, which falls somewhere between the two approaches, is the one mentioned previously about the math tests for which the test questions are presented in Hebrew and Russian (a bilingual test). Both the math text and the test questions were presented in two versions: Hebrew and Russian. Although on the surface, this version of a test seems to follow the approach of multilingual tests with two homogenous languages, when students were asked about the process they followed in responding to the text, they admitted to using a bilingual pattern of thinking while responding. Specifically, they claimed that they took some words from the Russian version, understood the syntax from the Hebrew version, and combined both in the process of meaning-making. This approach also shows that there are various strategies. They may still have used bilingual mixed processing strategies as well as monolingual ones while reading academic texts. This, as we mentioned previously, also happens in monolingual texts (Logan-Terry & Wright, 2010). There are those (Canagarajah, 2006) who claim that a focus on one language only is unrealistic, as the main goal should be negotiations and communication. To reach that goal, many routes can be followed: monolingual,

bilingual, and multilingual, as well as other communicative and negotiation strategies, including images and symbols, and other semiotic devices, as was demonstrated in the examples described in this article (see also Canagarajah, this issue).

It is clear that the processing and production of academic texts provide a useful tool for high-level academic performance. It is a challenge to the language testing profession to develop and invent tests and rubrics that will be based on a broader multilingual construct of language.

## CONCLUSION

One central question in the adoption of multilingual assessment procedures is whether it is even legitimate to compare groups with different language backgrounds or perhaps regardless of language backgrounds, we should shift to content-based tests in which students may react in any means they find appropriate, but language should not be the main focus (i.e., challenging the notion of language tests in general, which are detached from content). Can there be any test that is “language or languages”? Should we adopt separate tests for bilingual and monolingual students to avoid unjust comparisons, as all monolinguals cannot perform on bilingual tests?

Rather than ask “how long?” there is a need to carry out research that looks deeper into meaning construction of bilingual/multilingual test-takers in different settings. Academic knowledge is a complex construct and additional factors and dimensions provide resources for meaning construction: L1, cognitive skills, familiarity, previous experiences, and so forth. The data shown here point to the exclusion of the complex dimensions of academic language and thus provide invalid test scores. Language testing as a discipline has reached a point where it must examine its underlying constructs in line with developments in the field of language learning, teaching, and use. It is time for the language profession to think of a construct that will better match current research and knowledge and not serve ideologies of nation-states that are interested mostly in homogenous and monolingual citizens. These constructs need to incorporate the specific and unique competencies that demonstrate the advantages that bilingual students have in a number of areas, in spite of not being proficient in the *dominant* language. Language testers should take the role of resistance ‘from below’ to imposed ideologies and invent creative ways of testing, which reflect the diverse populations in this day and

age and provide equal participation not limited to dominant languages and dominant people.

---

## REFERENCES

---

- 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. (2009). Utilizing accommodations in assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 331–347). Berlin: Springer.
- Canagarajah, S. (2006). Changing communicative needs: Revised assessment objectives, testing English as an international language. *Language Assessment Quarterly*, 3, 229–242.
- Cheng, L. (2009). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 349–364). Berlin: Springer.
- Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *Modern Language Journal*, 94, 103–115.
- Davies, A. (2009). Ethics, professionalism, rights and codes. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 429–444). Berlin: Springer.
- Extra, G., Spotti, M., & Van Avermaat, P. (2009). *Language testing, migration and citizenship: Cross-national perspectives*. London: Continuum.
- Haim, O. (2010). *The relationship between Academic Proficiency (AP) in first language and AP in second and third languages*. Unpublished doctoral dissertation, Tel Aviv University.
- Hogan-Brun, G., Mar-Molinero, C., & Stevenson, P. (Eds.). (2009). *Discourses on language and integration*. Amsterdam: Benjamins.
- Levi-Keren, M. (2008). *Factors explaining biases in Mathematic tests among immigrant students in Israel*. Unpublished doctoral dissertation, Tel Aviv University. [in Hebrew]
- 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, 1–14.
- Levin, T., Shohamy, E., & Spolsky, B. (2003). *Academic achievements of immigrants in schools*. Report submitted to the Ministry of Education. Tel Aviv University. [in Hebrew]
- Li, Wei, & Martin, P. (2009). Conflicts and tensions in classroom codeswitching [Special issue]. *Bilingual Education and Bilingualism*, 13.
- Logan-Terry, A., & Wright, L. (2010). Making thinking visible: An analysis of English language learners’

- interactions with access-based science assessment items. *AccELLerate!*, 2(4), 11–14.
- Makoni, S., & Pennycook, A. (2006). *Disinventing and reconstituting languages*. Clevedon, England: Multilingual Matters.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.
- Menken, K. (2008a). Editorial introduction 7.3. Thematic issue entitled No Child Left Behind and U.S. Language Education Policy. *Language Policy*, 7(3), 191–199.
- Menken, K. (2008b). *English learners left behind: Standardized testing as language policy*. Clevedon, England: Multilingual Matters.
- Menken, K. (2009). High-stakes tests as de facto language education policies. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 401–414). Berlin: Springer.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher*, 23, 13–23.
- No Child Left Behind Act (2001). Public Law 107–110 (January 8, 2002). United States Congress. <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Poehner, M. E. (2008). *Dynamic assessment. A Vygotskian approach to understanding and promoting L2 development*. Berlin: Springer.
- Ricento, T. (2006). *An introduction to language policy: Theory and methods*. Oxford: Blackwell.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow, England: Pearson Educational.
- Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches*. London: Routledge.
- Shohamy, E. (2009a). Introduction to Volume 7: Language testing and assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. xiii–xxii). Berlin: Springer.
- Shohamy, E. (2009b). 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: Benjamins.
- Spolsky, B. (2009). Language assessment in historical and future perspective. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 445–454). Berlin: Springer.
- Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement: Final report. Project 1.1*. Santa Cruz, CA: Center for Research on Education, Diversity and Excellence (CREDE).
- Wright, W., & Xiaoshi, L. (2008). High-stakes math tests: How No Child Left Behind leaves newcomer English language learners behind. *Language Policy*, 7, 237–266.