The Routledge Handbook of Language Testing

Glenn Fulcher, Fred Davidson

Language testing and language management

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9780203181287.ch34
Bernard Spolsky
Published online on: 13 Mar 2012

How to cite: Bernard Spolsky. 13 Mar 2012, Language testing and language management from: The Routledge Handbook of Language Testing Routledge
Accessed on: 12 Dec 2020
https://www.routledgehandbooks.com/doi/10.4324/9780203181287.ch34
The chicken or the egg?

I have been puzzling about a title for this chapter, and it will become clear in the course of the chapter that I remain uncertain—confused, even—about the relationship between language testing and language policy, which is why I have chosen the neutral connector “and”. My first thought was to consider language testing as a method of language management (Spolsky, 2009), that is to say, a way in which some participants in a speech community attempt to modify the language beliefs and language practices of others. Let us take a simple example: a school class in language, in which the teachers set and mark tests that require pupils to use a version of the school language (or other language being taught) that he or she hopes that pupils will use. In essence, this assumes a belief, unfortunately widespread among those responsible for education, that testing is a good way to teach, rather than a way to gather data that needs further interpretation. Rather than take on the difficult problems of dealing with circumstances such as poverty which interfere with school success, a politician can simply propose that students who fail to pass a test (whatever that means) should be punished or even that their teachers who fail to bring them to some arbitrary standard should be fired. The general level of ignorance about testing displayed by politicians, the press, and the public is sometimes hard to believe (Taylor, 2009). How, for example, is a “pass” determined? I even recall a newspaper headline that complained half the students had failed to reach the average.

Krashan (n/d) regularly draws our attention to the huge gap in achievement between middle class and lower class schools in the United States: the former produce results as good as any other nation, while the latter pull down the national average, and international ranking results from the high percentage of child poverty in the United States. Under successive presidents of all political persuasions, US education policy has followed a brutally ineffective approach, testing and punishing rather than teaching (Menken, 2008). The same strange but convenient belief can be applied to all educational levels and topics, whether reading in the early years or mathematics in high school, so that its application to language specifically is not the point. Additionally, a decision to write a test in the standard language, whether pupils know it or not, is simply a consequence of an earlier decision to use it as the medium of instruction. A high-stakes test in the national language may modify the sociolinguistic ecology of a community, but language management may not have been the first direct goal of the testing (Shohamy, 2006).
Some historical examples

In this light, the significant effect that the Chinese Imperial Examination system had in establishing the status of written Chinese and the Beijing way of pronouncing it as the ideal version of language, paving the way for the Putonghua (common language) campaign (Kalmar et al., 1987) that is now seen as a method to unite a multilingual society, was not its original intended result, but rather an inevitable outcome of the choice of language in which this restricted higher-status elite examination was written (Franke, 1960). While the purpose of an examination may be neutral as to language, the choice of language will promote a particular variety of language and so advance its wider use. As long as the Cambridge tripos was an oral examination in Latin, it bolstered the status of those able to speak Latin: when it became a written examination, it recognized the significance of the English vernacular. We must be prepared therefore to distinguish between the language-related results of testing and the intention to use a test to manage language policy.

This may also be illustrated by the Indian Civil Service examination which helped establish the power and importance of examinations in nineteenth-century England. In arguing for the value of an examination as a method of replacing patronage in selecting candidates for government office, Macaulay (1853, 1891) made clear his neutrality on language issues. If the English public schools taught Cherokee rather than Latin and Greek, he said in his speech in Parliament, the candidate who could compose the best verse in Cherokee would be the ideal cadet. In practice, the examination when it was established did not include Cherokee, but besides Latin and Greek, it incorporated compulsory papers in German, French, Italian, Arabic and Sanskrit (Spolsky, 1995: 19). Similarly, the testing system which the Jesuits brought back from China for their schools in the seventeenth and eighteenth century had tested a syllabus that was determined centrally and did not focus in particular on attempting to change language choice, although of course it rewarded use of the appropriate version of the school languages and punished incorrect grammar (de La Salle, 1720).

But the Jesuit system was later adapted to language management during and after the French revolution. Under the Jacobins, the secularized school system was given the specific task of making sure that all pupils in schools under French rule inside France or its empire came out speaking and writing Parisian French, a task that took some seventy or eighty years to realize in schools of all but a few peripheral regions, and was never finally achieved in the colonies (Ager, 1999). This was a clear case of language management as I define it. Cardinal Richelieu’s goal of uniting the multidialectal provinces of France under a Parisian king, implemented by his establishment of the Académie Française (Cooper, 1989), was taken over by the Jacobins and carried on by Napoleon and successive French governments to establish monolingualism not just in continental France but ideally in the many colonies that France came to govern during the nineteenth century. The French example, with its many government committees and agencies responsible for promoting francophonie, and its timely constitutional amendment setting French as the sole language of the Republic passed just before Maastricht when the establishment of the European Community threatened to encourage other languages, showed the nationalist language focus of these activities quite clearly. Accepting the close tie between language and nationalism, a policy that insisted on the use of a single language and that used an elaborate high-stakes examination system to achieve it was clearly an example of language management.

Tests for language diffusion

My own first experience of language management testing was when I was at secondary school in New Zealand and took part in the oral examinations conducted by the Cercle Française, which, I
have since learned, was an agency of the French language diffusion enterprise (Kleineidam, 1992). The New Zealand school examinations, including the national examinations conducted towards the end of secondary school, did not include oral testing in foreign languages. The Cercle Française tests were given outside school to selected pupils at two or three levels; they consisted of a dictation and a short conversation, and were conducted not by schoolteachers but by lecturers from the local university. Participation as far as I recall was voluntary, but it was an honor to be selected. This international program was part of French government efforts to encourage the learning of French even in those countries that had not come under French colonial rule. It was distinct from the normal examination of language subjects associated with regular schooling.

Although school-based testing of national and foreign languages can be seen as an attempt to force or encourage the learning of the language, it is simply the application of testing to another school subject: it is the decision on which languages to include as school subjects (just like the decision on medium of instruction) that is language management rather than the test itself. Thus, the European Union’s policy calling for the teaching of two foreign languages is a management practice intended to ensure that a language other than English is included in the syllabus (Ammon, 2002). When testing is used to manage (or replace) teaching for all subjects, it is somewhat stretching the scope to include it in this chapter. In other words, it is a matter of the power of testing rather than of the power of language testing (Shohamy, 1993). The voluntary French oral tests were focused specifically on a language goal, and significantly, were under the direction of a foreign agency.

The gatekeeping function of high-stakes tests

When many years later I became directly involved in the testing of the English language proficiency of foreign students in or seeking admission to US universities, I became aware of the further management potential of language testing. I came to realize the non-educational effect of requiring English language competence in students applying to American universities, and wrote a paper drawing attention to the probable way in which it limited admission to those potential students whose parents’ financial or political standing had given them an opportunity to attend a school which taught English efficiently (Spolsky, 1967). Later, as I carried out research on the development of the Test of English as a Foreign Language (Spolsky, 1995), I learned more about this phenomenon, marked by the tendency of students from European countries to do better on the test (I discount, as some Canadian and other universities have learned to do, the distorted results of some Asian candidates who memorize possible examination questions).

The Test of English as a Foreign Language (TOEFL) is a clear case of a high-stakes test with social effects, and with a language management origin (Spolsky, 1995: 55–59). It was the third attempt to comply with a request from the US Commissioner of Immigration to correct a loophole in the 1923 US Immigration Act, the purpose of which had been to exclude immigrants from Asia, Africa, and eastern and southern Europe by setting quotas based on the years in which most immigrants had come from northern Europe. The loophole was a provision granting automatic student visas to candidates coming to study in the United States; the Commissioner feared (rightly) that many would use this as a way to enter the country and then remain illegally. The US government was not yet prepared (as the Australian government had been some years earlier) to establish its own testing system, but asked the College Entrance Examination Board, an association established by elite private schools to standardize admissions procedures, to develop a test which would guarantee that candidates planning to study in the United States had already achieved proficiency in English. In 1926, the Commissioner requested “that all schools indicate in the certificate of admission the exact knowledge of the English language a
student must have before he can be accepted”; this led to a request the following year to the College Board to develop a special examination. A committee set up by the Board suggested an outline for the examination, and a grant of US $5,000 from the Carnegie Endowment for International Peace helped to cover costs. With the hesitant assistance of some American embassies, in 1930 30 candidates took the examination in eight different countries; the following year, 139 candidates were examined in 17 countries including a group of 82 engineering students in Moscow. In 1932, the test was offered in 29 countries but in the increasingly difficult world economic depression, there were only 30 candidates; in 1933, the number dropped to 17 and in 1934, after only 20 students could afford the US$10 fee, the examination was “at least temporarily” discontinued, as no one could afford to prepare a new form. (We may also add that it was thus not available to certify the English proficiency of Jewish professionals seeking to escape from Nazi Germany.) There are a number of interesting features about this first attempt: the dependence on the support of foundations, the intelligent design of the test but practical limitations in its implementation, and the lack of willingness on the part of prospective users to pay for it.

The second attempt to plug the immigration loophole came in 1945, again with encouragement and limited support from the Department of State. Advised by leading language testers (including Charles Fries and his young student Robert Lado), one form of the English examination for Foreign Students was prepared in 1947 and administered overseas at some Department of State Cultural Centers. The test, developed by the College Board, was passed on to the newly independent Educational Testing Service at Princeton and available for a few years until financial problems led it too to be dropped (Spolsky 1995: 155–57).

In 1961, there was a new attempt. In the meantime, several American universities had developed English tests of their own. Lado’s Michigan tests were the best known, but the American University Language Center tests, originally written by A. L. Davis but later improved by David Harris, were also widely used by the US government. Many US consulates conducted their own primitive tests before granting student visas, but all these tests lacked the financial resources to guarantee security by offering new forms at each administration. To consider a possible solution, a meeting of interested parties was called by the newly established Center for Applied Linguistics in 1961 which developed a plan for what eventually became the TOEFL (Center for Applied Linguistics, 1961). After a few years of independence, and as a result of complicated maneuvering, the test was taken over completely by Educational Testing Service and grew into a major source of income (Nairn, 1980). Under user pressure, tests of spoken and written proficiency were later added, followed by a computerized version of the test. But the underlying problem continues: an Associated Press report in The New York Times on 8 March 2010 says that “A California man was charged Monday with operating a ring of illegal test-takers who helped dozens of Middle Eastern citizens obtain United States student visas by passing various proficiency and college-placement exams for them, federal authorities said.” This is a story repeated throughout the world, for all high-stakes tests.

**Exploiting the demands for English tests**

This corroborates how the value of the test finally became apparent to prospective users: the status of English as the major international language for education and commerce moved the demand for the test to the public, so that by now there are competing international English tests offered by public and private organizations. The first of these was the English Language Testing Service test (later renamed International English Language Testing Service test) developed in the late 1970s by the University of Cambridge Local Examinations Syndicate at the request of the British Council, with the participation later of the International Development Program of
Australian Universities (Criper and Davies, 1988; Alderson and Clapham, 1992). This test and its later forms have proved financially successful, and others have rushed into the business, including Pearson Education, a major international firm selling teaching and testing materials. There are now English testing businesses in many Asian countries.

At this stage, we have moved beyond the organized language management situation (Neustuptný and Nekvapil, 2003), with governments attempting to encourage people to learn their language, and to a stage where the demand for testing comes from the learners. That is to say, whereas the French oral examination that I took was supported by the French government to encourage pupils in other countries to learn French, these major industrial testing services intend to profit from demand from people in non-English speaking countries who see the value of learning the language and having a way of certifying that proficiency. These commercial tests then serve what Nekvapil and Nekula (2006) would call simple management, namely an individual seeking to improve his language and communication skills, rather than the complex management envisioned in the title of this chapter which assumes an authority aiming to manage the language choice of others: we are dealing then with top-down exploitation of bottom-up demand.

A similarly neutral attitude to language management lay behind the development of the Council of Europe Framework of Language Learning and Teaching (Council of Europe, 2001). While the Council and the European Union do have a language management policy, namely that all should learn two additional languages, they do not specify which languages they should be. True, calling (unsuccessfully as it turns out) for two additional languages is an effort to make sure that European foreign language learning should not be restricted to English teaching. But it is a weak management plan, as compared say to the Soviet insistence on satellite countries teaching Russian or the French willingness to accept bilingualism provided only that it includes French.

Phillipson (1992) has proposed that the spread of English is the result of imperialism of the core English-speaking countries rather than the result of the growing demand for knowledge of a useful language. Many argue against his position (Fishman et al., 1996). For me, the clearest evidence is the weakness of government diffusion agencies. The British Council for some years now has seen the teaching of English as a source of income rather than as a political and cultural investment, although admittedly in the 1930s there were a few who saw it as a way of spreading English influence; and the US Information Agency discouraged those centers which allocated too large a proportion of the budget to English teaching. This contrasts with other diffusion agencies like the network of activities encouraging Francophonie, or the language-teaching programs of the Goethe Institute, or the developing community of Lusophone countries with its post-imperial concerns, or the growing number of Confucius Institutes set up by the Chinese government to persuade Africans to replace English and French by Mandarin Chinese.

An instructive exception was the attempt of John Roach, assistant secretary of the University of Cambridge Local Examinations Syndicate who in the 1920s and 1930s concentrated on the development of the English language test, which, he argued before he left the Syndicate in 1945, would be an acceptable way of spreading English and its associated way of life (Roach, 1986). His ideas were ahead of his time, opposed by the establishment including the British Council: but he managed to keep the test alive into the 1940s so that it was ready for its later profitable growth as the demand for English spread.

School examinations for language management

I need to return to the general issue of school examinations as a potential language management device. My argument is that language tests and examinations can serve to focus attention in situations where society accepts the decision of government to require the learning of the
language and even more in those cases where education is conducted in a language other than the home language of the pupils. Members of minority groups and children of immigrants can be forced by these examinations to attempt to change their language practices; otherwise, their failures can be blamed on them rather than on a misguided educational policy.

Few educational systems allow for the fact demonstrated in many research studies that students need several years of instruction in a language before they can learn in it efficiently (Walter, 2003, 2008). In these circumstances, language tests become a powerful method of filtering out all but a select elite of immigrants and minority children; in Africa, in most countries where the medium of instruction is English or French, it reduces the possibility of providing a good education to the mass of speakers of indigenous languages (Heugh, 2005; Heugh et al., 2007). There are of course other ways of forcing the learning of the school language: American Indian children were punished for using their own home languages (McCarty, 1998, 2002), just as the enforcement of English in Welsh schools was accomplished by corporal punishment rather than testing (Lewis, 1975). In school then, language testing is one of several methods of implementing language management policies. Phillipson is right in his judgment of the effect of these policies, but perhaps overstates the intentionality.

Language tests to manage immigrants and asylum seekers

A similar analysis may be applied to the use of language tests in non-school situations. The infamous Australian immigrant language test (Davies, 1997) is a case in point (see Kunnan, this volume). As a number of studies have shown, the test was intended to exclude prospective immigrants who were judged unsuitable on other grounds by the immigration official. He was instructed to give a dictation test in a language which he assumed the immigrant did not know. The test failed if the immigrant passed. While the circumstances and results were similar, this was quite different from two other kinds of language tests administered to prospective immigrants, the test given to an asylum seeker to check his or her identity, and the test given to a prospective citizen to guarantee knowledge of the official language.

The asylum seeker’s test, now becoming common in Europe and providing employment for self-certified dialect experts (Eades et al., 2003; McNamara, 2005), is perhaps the closest thing to the shibboleth test in the Bible. I quote the passage from the book of Judges:

The men of Gilead defeated the Ephraimites for they had said, “you Gileadites are nothing but fugitives from Ephraim ...” The Gileadites held the fords of the Jordan against the Ephraimites. And when any fugitive from Ephraim said, “let me cross,” the men of Gilead would ask him, “are you an Ephraimite?”; if he said, “no,” they would say to him, “then say shibboleth”; that he would say “sibboleth” not being able to pronounce it correctly thereupon they would seize him and slay him by the fords of the Jordan. Forty-two thousand Ephraimites fell at that time.

(Judges, 12: 14)

In the current European asylum tests, the experts claim to be able to determine the origin of the asylum seeker from aspects of his dialect (Blommaert, 2008; Reath, 2004). As pointed out by Kunnan (this volume) the identifications are doubtful and clearly miss such facts as that the speaker may pick up a particular pronunciation from a short stay in another country or from a foreign teacher. Again, in the asylum seeker’s test, the test is a method of implementing an immigration decision to refuse admission to a person who cannot prove that he or she is from a specific country or group.
Similarly, language tests for citizenship depend on a government decision to require knowledge of the national language. The relationship between language and citizenship has been widely debated. Those who argue for multicultural citizenship favor linguistic diversity, but linguistic homogenization has been a common method of establishing civic identity (Patten and Kymlicka, 2003: 11–12). Our concern however is not specifically with multilingual societies, but with immigrants seeking citizenship in a new society. It is commonly agreed that in the long term they should integrate by learning the national language. But there are those who argue that excessive emphasis on language as a gateway to community membership is associated with “old-fashioned cultural assimilation.” To expect immigrants to develop fluency in the state language rapidly is seen by liberals as a challenge to their civic rights. The question then becomes, at what point in the process should the system require proficiency in the state language: in order to emigrate, in order to receive voting rights, or as in some Constitutions (this is notable in the linguistic clauses of the constitution of former British colonies in the West Indies) in order to be a candidate for legislative office. Whichever decision is made, one way of implementing it is likely to be a language test. The tests then will be seen by some as a barrier to immigrant rights (Milani, 2008; Nic Shuibhne, 2002).

Language tests to manage employment

Related to the case of language proficiency for citizenship, there are complex issues in requiring language proficiency for employment in various positions. There are occupations that seem to require specific language proficiency. I start with one that has been quite recently recognized: the ability of airplane pilots and ground personnel, especially traffic controllers, to communicate with each other (see Moder and Halleck, this volume). The International Civil Aviation Organization has established a policy requiring air traffic controllers and pilots to have certified proficiency in English or in a relevant national language. This step was taken only after a number of cases of accidents attributable to communication failure. Setting the requirement has been comparatively simple, but finding a way to implement it has led to a complex and largely unsatisfactory search for appropriate instruments. Alderson (2009, 2010) has been studying the process, and finds a multiplicity of tests, few with evidence of standards, reliability, or validity.

This is perhaps the appropriate point to comment on the quality of any of the tests used for language management. I have already raised questions about the validity of the test used for asylum seekers. Serious questions can in fact be raised about many tests used in education. Apart from the normal questions about test validity, special problems arise when the same test is expected to measure the language proficiency of native speakers and of language learners, as currently in the United States. The range is simply too large to develop an appropriate instrument: this is one of the most significant flaws in the current attempt to use test results to judge the efficiency of various schools and teachers. It is a fundamental principle that a poor test is a useless or dangerous way to manage language policy. This tendency will be exacerbated if countries, like the United States, adopt a policy that establishes a “common core” of educational standards (focused on English and Mathematics), and if as anticipated national examinations are developed to assess students, schools, and states in meeting those standards. If this happens, the result will be a major revolution in further centralization of US education and the kind of narrowness of focus that central standardized tests have produced, a fact bewailed 140 years ago by an English educator (Latham, 1877) who saw this as an emulation of the French policy.

Returning to occupational language qualification tests, one area where language proficiency is or should be a requirement is the health profession. Doctors and nurses need to be able to communicate with their patients both for satisfactory diagnosis and monitoring of treatment.
The question then arises, who should be tested? Many health agencies put the burden on the patient, who is expected to bring a child or other bilingual capable of answering questions and passing on instructions. Foreign and foreign-trained doctors and nurses are regularly expected to pass language tests, although the focus of these tests remains a serious problem. Should they be tested in medical language, or in communication in the various vernaculars with patients? Having listened to an Israeli doctor (fluent in Arabic and Hebrew) trying to take a medical history in an emergency room from an elderly patient limited to Russian and depending on a relative who could manage English, one realizes the attraction of requiring plurilingual proficiency in health personnel: the other choice is of course the expense of providing qualified interpreters in a large range of languages (Roberts et al., 2001).

Similar problems are faced by police departments in increasingly multilingual cities. In 2005, for instance, 470 New York police department employees were reported to have passed tests in more than 45 languages, including Arabic, Urdu, Hindi, Pashtu, Farsi, Dari, and Punjabi: another 4,000 were waiting to be tested. A number of US police departments encourage certification in sign language. Again, qualified personnel reduce the need for extensive and expensive interpretation services. In a related policy, the European Union required new candidates for membership to ensure that officials (especially border and customs officials) were proficient in Western European languages as well as in the Russian they needed during Soviet rule.

In the business world, international firms and companies with dealings with speakers of other languages commonly have a language policy usually implemented by hiring employees with relevant language proficiency. Occasionally, they may have specially designed language tests to measure the kinds of proficiency associated with the task, but they are just as likely to rely on casual interviews or the results of academic language tests. There have so far been only a few studies of language management in the workplace; they show the complexity of situations and include cases where testing is used (Neustupný and Nekvapil, 2003).

**Language tests—whether and whither?**

Putting all this together, one might ask about its relevance to this Handbook and to language testing practice. For prospective language testers, it might be read as a warning that you are heading into a dangerous field, one where there will be jobs, but where you will face, like workers in the armament industry, ethical problems. Tests, like guns, can be and regularly appear to be misused. For students of language policy, it shows the existence of a powerful instrument to manage other people’s language, to force them to learn and use new varieties, to trap them into authority-valued or -condemned identities. For all of us, it raises serious issues in educational, civil and political spheres, cautioning us that tests are as easily misused as employed for reasonable goals.

I would certainly not want to be interpreted as claiming that testing is necessarily bad: just as there are morally and ethically justifiable uses for guns, so there are morally and ethically justifiable uses for language tests. The problem seems to be rather that both tests and guns are potentially so powerful as to be commonly misused. If we analyze the problem of the US educational system in the twenty-first century, there is good reason to suspect that it can be cured only by serious social and economic engineering and the eradication of the poverty that produces the most serious failures. It is readily understandable but regrettable that educators, politicians, and the naive general public are attracted to a much simpler approach: raise the standards of school tests and punish any teacher or school that fails to achieve good results. This approach may seem more attractive, but in the long run it will not work. I am therefore depressed about the short-term future, and can only hope for greater wisdom in the future. Similarly, it is beginning to seem logical to deal with the demographic changes produced by immigration and asylum.
seeking by setting barriers controlled by tests rather than developing intelligent plans for teaching necessary language skills to immigrants and their families and providing appropriate support until they are gained. Again, the short term solution is unlikely to succeed, but it will take some time before the general public and politicians are prepared to undertake more basic solutions.

Thus, I deplore the misuse of language tests for management as a superficially attractive but fundamentally flawed policy. Language tests can play a major role when used diagnostically in improving language teaching or in planning for language support systems, making them an excellent instrument for intelligent and responsible language management, but the misuse remains a “clear and present danger”.

Further reading

McNamara, T. and Roever, C. (2006). Language testing: the social dimension. Malden, MA: Blackwell Publishing. While language tests are usually analyzed using statistical measures and citing psychometric principles, this important work adds social dimensions, thus giving sociolinguistics a role in testing alongside statistics.


Spolsky, B. (1995). Measured Words: the development of objective language testing. Oxford, UK: Oxford University Press. The first half of the book is a history of language testing up to 1961, the year that a meeting took place in Washington, DC that led to the creation of TOEFL. The rest of the book tracks the meeting and its results, and shows how what started as a test controlled by language testers became part of a major testing industry. It traces the early years of the industrialization of language tests.

References


