

# **Language Tests for Social Cohesion and Citizenship**

**Issues of quality assurance**

***An Outline for Policymakers***

**Henk Kuijper**

**ALTE**

Presentation held at the Intergovernmental Seminar 'The Linguistic Integration of Adult Migrants'

Council of Europe, Strasbourg, June 26th-27th 2008

## **Introduction**

Many European countries are introducing or formalising linguistic requirements for the purposes of migration, residency and citizenship, and national governments often require language tests or other formal assessment procedures to be used.

It is not the purpose of my presentation to promote or reject the use of such language tests but, where mandatory testing is either in place or under consideration, it is necessary to offer professional guidance based on good testing practice to ensure that the needs of stakeholders are met and that tests are fair for test takers. Test fairness is a particularly important quality when tests are related to migration, residency or citizenship.

At first glance this issue of test quality is not different from any other consideration language testers are confronted with when delivering high stake tests. But at the same time tests for immigration, integration and citizenship can be used – even more than other high stake tests – as political instruments to influence migration streams and for immigrants already living in a country to influence their legal status: because for a growing number of countries civil rights like entrance into a country - including the right to marry or to family reunion - residence permits, citizenship, social security payments and even – as I know - the right to be allotted a house are related to language achievements or attending language courses. These language tests are profoundly influencing the lives of the test takers and thus very high stake. Tests therefore must be of excellent quality. How can this be achieved?

This seminar shows that there are numerous issues related to the use of language tests. I will focus on the guidelines for language tests, following the ALTE contribution titled: Language tests for social cohesion and citizenship.

## **Use and impact**

The important starting questions for politicians and test developers before designing a test or assessment procedures are:

- what use will be made of the test results?
- what will be the impact for the migrant?
- what will be the consequences of a test on society?
- is it more appropriate to use another form of assessment, rather than test?
- could it be appropriate to use several methods of assessment at the same time?

The use of the results of language tests for immigration and integration is an extremely important feature for the context and construction of language tests. Considering the development of using language tests and their consequences over the last ten years we can observe a stricter policy on language testing for immigrants and a shift from using language test to facilitate language learning to connect sanctions to these test results. This means that the impact for immigrants grows more and more negative as civil rights and financial consequences come joining the game. But also consequences for society must be taken into account. The facilitating approach will enable us to adapt courses and tests tailored to the needs of different groups of immigrants, based on work and educational backgrounds and aspirations. The goals set can vary and be set rather high in special cases. At the contrary the sanction based approach carries the danger that, because of the strong consequences and because of reasons of equal rights, the language levels set are rather low to not exclude immigrants with lower socio-economic and educational backgrounds from civil rights.

For instance in the Netherlands during the facilitating approach learning and test facilities were existing where immigrants could reach levels up to B1 in immigration courses. Since the new legislation, with its financial and civil rights consequences the level to acquire for everyone was lowered to A2. Fine from the perspective of civil rights, not that fine because A2 is too low a level to really function in work and society. For the same reason flexibility of test methods across candidates is more difficult to achieve when sanctions are related to test results, as differentiation could lead to arguments of incomparability.

At the other hand setting too high levels, like B1 or even B2 for tests with legal consequences would exclude many immigrants from for instance citizenship of countries they are living in for already a longer period, resulting in the long run in societies with two different kinds of inhabitants.

It may be clear that these prior issues will have to be discussed extensively, as there is a great danger that tests would have to serve conflicting functions.

### **Validity and reliability issues**

Once policymakers have decided language tests for immigration, integration and citizenship have to be developed and language testers have accepted the task of developing such tests a number of further questions have to be answered. These questions can be classified into two main categories:

- The validity issue: what is the assessment aimed at and does the test reflect this aim
- The reliability issue: Are the tests set up in a way that the test specifications are measured in a reliable way to ensure fairness for candidates

The validity of the test involves the following activities:

- Determining test purpose and real world demands on test takers
- Determining the linguistic demands
- Determining the appropriate level of difficulty
- Producing test specifications

### **Test purposes**

In determining test purpose and real world demands on test takers it is not sufficient to state that a test is for the purposes of migration and citizenship, because even within this area, there is a wide range of reasons for testing migrants, ranging from motivating learners to ascertain whether their competence is sufficient for participation in well-defined social situations as study or work, to taking decisions which affect their legal rights, such as their right to remain in a country or acquire citizenship of it.

Extensive needs analysis is necessary for reaching a sound overview of possible test purposes. It is also necessary to take into account the fact that there are various subgroups of migrants with their own specific needs. Those, for instance, who want to join the job market as soon as possible often have different needs from those who are planning to raise young children at home. In a needs analysis, it is good practice for language test developers to define the relevant contexts and situations of the target groups. In planning such needs analyses, policy makers should be sure to set aside sufficient resources. And delegates from different sections of society should be involved in the definition of the needs.

### **Linguistic demands and level of difficulty**

For determining the linguistic demands and the appropriate level of difficulty we face similar problems. If, for instance, a language test were designed to measure whether the test-taker had

the language proficiency necessary to follow a vocational training course, we might expect it to test the ability to follow lessons and workshops, to communicate with teachers and fellow students, to read relevant literature, to write assignments, etc. This analysis could then help to determine the appropriate level of language proficiency. If, on the other hand, a language test were designed for the certification of translators or interpreters, a comparable needs analysis of the profession would show the required level of proficiency to be much higher and the linguistic demands would be very different in content.

In contrast, deriving linguistic requirements is far less straightforward in the case of migrants and candidates for citizenship. The relation between language proficiency and the ability to integrate into society or exercise the rights and responsibilities of citizenship is looser and far more difficult to pin down. After all, if language proficiency were the only factor in play, all native inhabitants of a country would be fully integrated citizens. As this is not the case, it can be deduced that other factors are also important. Another difference is that language tests for study and work are, in most cases, taken by groups of candidates which are homogeneous with respect to educational background and cognitive skills, while tests for integration and citizenship must cater for a full range of possible candidates. They must therefore be accessible to people with a high level of academic education but also to those with little or no literacy skills.

### **Reliability issues**

But let us assume we were able to produce proper test specifications that can cope with the great variety of test takers and the conflicting functions and purposes of tests.

Then we have to assure that these high stakes tests are meeting highest quality standards, especially because consequences are severe. To achieve high reliability is necessary for reasons of fairness for the candidates.

The test format must be appropriate for the target group. One can imagine that with the great variety of test takers this will be difficult to achieve.

When language tests are leading to very important decisions, extended pretesting is necessary to confirm that the items actually work as intended. We have to be sure whether

- the test measures the target language,
- the test differentiates effectively between stronger and weaker candidates,
- the test does not result in bias towards a particular candidate profile, etc).

Therefore it is necessary to pretest the materials under test conditions on candidates with a similar demographic profile to that of the live test candidates. Furthermore we have to assure that test administration and marking and grading is executed in the best way possible. Test providers should ensure that the test is taken under conditions which are equally fair for all candidates. To this end, procedures must be developed to minimise differences in administration.

High quality data analysis has to be incorporated to monitor the test qualities and the candidate's responses, to detect bias for groups of candidates and to investigate the impact on immigrants, language education and society. I have the impression that politician might underestimate the importance of and the resources necessary for data analysis and research.

Procedures for candidates with special requirements have to be developed either in adapting test administration or in rules for exemption.

## **Main issues**

Using tests for migration and citizenship purposes is considerably more complex than it may seem at first. I have attempted to outline the issues that need to be considered and, by implication, the issues for which policy makers should take responsibility. The questions of what type of assessment is necessary for the intended purpose, and what it can be expected to measure should be considered first. Where it has been decided to use a test, it is vital that the test meets the highest requirements I outlined before. It must not be forgotten that the outcomes of a test can have important consequences for the candidates, for larger groups of people, and the society as a whole. Most important are the consequences related to the civil and human rights of the test taker. When using a language test for migration and citizenship, policymakers must work with the test providers on several aspects after the decision to use a test. These aspects include the definition of the precise purpose of the test and the allocation of resources for successful completion of all stages of test development and test use.

At all times test fairness for immigrants should be considered of prime importance.