

Choosing Language Competence Descriptors for Language Assessment: Validity and Fairness Issues

Samira Elatia

University of Alberta - Canada

samira.elatia@ualberta.ca



Synergies Europe n° 6 - 2011 pp. 165-175

Dates de soumission/acceptation : 3 avril - 13 septembre 2011

Résumé : Dans l'évaluation des langues, est-ce qu'un seul ensemble de descripteurs de compétences langagières peut être utilisé quel que soit le mandat et le contexte de l'examen qui est en cours d'élaboration et d'usage? Cet article aborde la question complexe de la validité et de l'équité en faveur d'un ensemble de descripteurs de compétences, à savoir le Cadre européen commun de référence pour les langues (CECR), dans le développement de tests de langue. Dans plusieurs contextes d'évaluation pédagogique et langagière en dehors de l'Union Européenne, le cadre est utilisé sans une compréhension approfondie du mandat et du contexte dans lequel il a été développé et utilisé. En prenant comme référence le cadre européen commun de référence, les Niveaux de compétence linguistique canadiens (NCLC), et les Lignes directrices ACTFL (American Council on the Teaching of Foreign Language), cet article soutient que même si la compétence de base est presque identique entre ces descripteurs, les implications politiques, sociales et culturelles de l'objectif et les organismes responsables de leur développement sont différents. La notion de multiculturalisme qui est la base de ces descripteurs diffère grandement. Les descripteurs sont enracinés dans la realpolitik du contexte dans lequel ils sont utilisés comme ils sont élaborés et parrainés par les agences gouvernementales pour exercer un mandat social, culturel, politique et linguistique dans une région géopolitique bien précise. En outre, l'économie et la philosophie de l'éducation au sein de chaque contexte jouent un rôle majeur dans le développement et les objectifs de ces descripteurs.

Mots-clés : Descripteurs de compétence, évaluation des langues, validité, équité, construit, CECR, NCLC, ACTFL

Abstract : In language assessment, can one set of language competence descriptors be used regardless of the mandate and the context of the language test being developed and used? This article addresses the complex issue of fairness in favoring a set of guidelines, namely the Common European Framework of Reference for Languages (CEFR), in the development of language tests. In many educational and language assessment contexts outside of the EU, the Framework is being used without a thorough understanding of the mandate and the context in which they have been developed and used. Taking as references the Framework, the "Canadian Language Benchmarks" (CLB) and the "ACTFL guidelines" ('American Council on the Teaching of Foreign Language' guidelines), this paper argues that while the core competence in language skills is almost identical among these descriptors, the political, social, and cultural implications of the purpose and the agencies behind their development are quite different. The very notion of multiculturalism on which each of these guidelines is built differs greatly. Guidelines descriptors are deeply rooted

in the realpolitik of the context in which they are used. They are developed and sponsored by governments to serve a social, cultural, political and linguistic mandate within a very precise geopolitical region. Furthermore, economics and the philosophy of education within each context play a major role in the making and the objectives of these guidelines.

Keywords : Competence descriptors, language assessment, validity, fairness, construct, CEFR, CLB, ACTFL

Language Competence Descriptors in Language Testing

In language assessment, the construct intended to be measured is linked closely to the level of the examinee's competence and knowledge at a given skill in whether it relates to production, interaction, or comprehension of elements in and of a language. In almost all cases, the construct is defined by a set of language competence descriptors chosen by the test developers. These descriptors or guidelines, by attempting to describe what a person is able to do with the language at different levels of competence, help different stakeholders lay the foundation for what a person can do with the language that s/he is either in the process of learning or has already learned. The language competence descriptors provide a point of reference to different stakeholders and decision makers in language assessment and as such they play a major role in language test development. Competence descriptors are neither benchmarks nor standards; they are agreed upon levels that indicate what a language speaker is able to do with the language s/he is learning or has learned. Using competence descriptors help establish standards and criteria for assessment.

When the first Task Force on Testing Standards (TFTS) of the International Language Testing Association submitted its report (1995), it included language competence descriptors as standards. The fact that language competence descriptors made up more than half of the standards (table 1 below) collected in this initial report is a testament to their importance in the test development process: from crafting specification and writing items, deciding on outcomes and scores, inferences to making decisions on test results, and deciding on an examinee's level of competence in a language.

Standard definition/type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Guideline	58	52.7	58	52.7
Other	10	9.1	68	61.8
Performance	15	13.6	83	75.5
Test	27	24.5	110	100.0

Table 1: Breakdown of Type of Standards (TFTS Report 1995, p.18)

Within the language assessment field, language competence descriptors are important for the validity of the assessment instrument. In the process of test construction, defining the construct is primordial and these constructs "are selected from models, embodied in

frameworks that relate constructs to contexts, and operationalized in test specifications that articulate purpose in practice (Davidson and Fulcher 2007 : 232).” Establishing language competence descriptors that are targeted becomes essential in narrowing down the intended construct and subsequent test specifications¹. Davis (2004) stressed the ethical importance of defining what the test is targeting as a basis for establishing validity. In ILTA’s Code of Practice (2007), the relationship between construct and the competence to be assessed is addressed in Articles 1 and 2 of Part A:

Art.1: The test developer’s understanding of just what the test, and each sub-part of it, is supposed to measure (its construct) must be clearly stated.

Art.2: All tests, regardless of their purpose or use, must provide information which allows valid inferences to be made. Validity refers to the accuracy of the inferences and uses that are made on the basis of the test’s scores. [...] The test score inference or interpretation can be valid only if the test construct offers as accurate as possible a picture of the skill or ability it is supposed to measure. (ILTA Code of Practice, 2007 : 2)

Consequently, the choice of the appropriate language competence descriptors becomes an important component in the development of the test. It is an essential part of the mandate of the assessment tool. Davidson and Lynch (2001) define the mandate “as that combination of forces which help decide what will be tested and to shape the actual content of the test (p.77).” The mandate is motivated by either “pedagogical and research forces such as changes in the theoretical base for teaching and learning languages or common practices at a particular institution” and “by forces outside the test (ibid).” As such, it would be unprofessional to choose a set of language competence descriptors without providing strong validation arguments to support the choice. Unfortunately, in many language assessment situations, choosing a language competence descriptor rests on what is widely used at the moment, that is, what is fashionable. Such a practice is a threat to the validity of the construct and of the whole testing enterprise. When deciding on which language competence descriptors to use, it is important to use them neither outside of the context and mandate in which the language competence descriptors are developed nor without understanding what they stand for and what are the social, political and historical backgrounds that led to their development as well as the agencies that stand behind their development.

Defining Performance and Competence Language Descriptors

Performance descriptors or guidelines are statements of expected ability and/or competence in a language. They serve as guidelines for curriculum design in language teaching, as scales for establishing item difficulty and establishing language constructs, and criteria to be achieved at different levels of both teaching and assessment, and they serve as the genesis of language test tasks and items. They also serve in determining a person’s proficiency in a language independently from any language learning program. These include a person who is not learning the language at the moment, or has not learned the language in a conventional academic setting, but would like to know his or her proficiency level. It is not exclusively for educational needs as it serves many other purposes such as work, immigration, promotion, and/or just a personal interest in knowing one’s competence level in a language.

These guidelines are used by test developers as language performance descriptors and they refer to them primarily for the purpose of test construction and data analyzes. For both purposes, there is an intricate relationship. Around the world, many sets of competences descriptors are being developed, but for almost a decade now, the “Common European Framework of Reference for Languages” dominates the field and is widely used. Other examples of competence descriptors include the “Canadian Language Benchmarks” and the “American Council on the Teaching of Foreign Languages” guidelines,” which were widely used in the 1980s and 1990. In this article, the *Benchmarks* and the *Guidelines* are used as a point of comparison to the *Framework*.²

Mandate and Context of *Benchmarks*, *Guidelines* and *Framework*

It would be erroneous and simplistic to assume that the *Benchmarks*, *Guidelines*, and *Framework* are equivalent, and that they can be used interchangeably. The core which describes the competence level may be similar, but the overall presentation, mandate, and context are different from one set to another. Most importantly, the purpose for which each has been developed differs. It is these differences that stakeholders and decision makers need to pay close attention to when choosing which one to use. Each of these language competence descriptors is a product of a unique context. Understanding these contexts and what gave birth to these descriptors sheds light on why and how they should and should not be used outside of their context.

Canadian Language Benchmarks

The Canadian Language Benchmarks provide “a national framework of reference for the development of language learning programs, curricula and materials relevant to the needs of adult newcomers to Canada during the process of settlement and integration (Language, 2011 para.1)” for both official languages: French and English. They are used for “describing, measuring and recognizing the second language proficiency of adult immigrants and prospective immigrants for living and working in Canada (ibid).”

Stage I	Stage II	Stage III
<i>Basic proficiency</i> Benchmarks 1 to 4	<i>Intermediate proficiency</i> Benchmarks 5 to 8	<i>Advanced proficiency</i> Benchmarks 9 to 12

Table 2: Breakdown of Levels and Sublevels in CLB

The *Benchmarks* were published in 2000 and are currently undergoing revisions. They are composed of 12 benchmarks, divided into three categories - basic, intermediate, and advanced proficiency - covering the four classic breakdowns of language skills: listening, speaking, reading, and writing (Table 2 above).

The *Benchmarks* serve a distinct purpose for the Canadian society in which French and English are the only two official languages within a multicultural setting that allows all Canadians to maintain their cultures and their own native languages be they Urdu, German, Mandarin, Cree, or any another language.

The Centre for Canadian Language Benchmarks is a national not-for-profit organization established in 1998, funded by the Canadian federal government, and “governed by a

nationally representative, multi-stakeholder board of directors including representation from government, ESL and FSL experts and language assessors. (Language 2011, para. 2).” The Center strongly warns about the use of the *Benchmarks* outside of their context without valid research to back up such use. The Center offers guidelines and directives to other individuals and institutions in and outside of Canada that would like to use the *Benchmarks* as a reference.

In a situation in which a language test would be chosen by Canadian immigration officials to assess the language competence of potential immigrants, it would be obvious that the *Benchmarks* would be selected as a point of reference in the test since it is in Canada. If decision makers favor another set of competence descriptors such as the *Framework* and use a test that is calibrated using the *Framework* levels, unfairness and validity threat hang over the decision process. Strong arguments need to be made to justify such a choice since there are already the *Benchmarks*, and they are developed in Canada, by a Canadian agency, to serve this very purpose. It would be questionable to use anything else that does not address the mandate and context of this testing situation. In fact, the decision makers would in reality violate the mandate of the test and of the immigration regulations.³ In England and the Netherlands, when officials were deciding on language assessment tools of immigration and citizenship, the *Framework* was selected for establishing the appropriate level of competence (De Jong et al 2009, and Blackledge 2009) without hesitation since both countries adhere to the Convention of Europe that mandates the use of the *Framework* as a way to standardize language competence descriptors across Europe. Furthermore, in the immigration situation, Saville (2009) stresses the need for “Understanding test *purposes* and related *contexts* (p.26),” otherwise several issues would arise from clashes between policy and decision makers on the one hand and testing practitioners and examinees on the other hand.

American Council on the Teaching of Foreign Languages

There are four main levels of the ACTFL *Guidelines*. These are superior, advanced, intermediate, and novice. However, one more level - “distinguished” - is added to reading and listening competence. The *Guidelines* are further broken down into subcategories: low, mid, and high. However, depending on which skill is being assessed, some of the fundamental levels are not broken down into sublevels, or there are only two sublevels within a level.

Unlike the *Benchmarks* or the *Framework* that are linked directly to governmental agency, *Guidelines* are the product of a language teaching association, and stem from a different perspective than the Canadian and European geopolitical situations. The *Guidelines* are closely linked to Interagency Language Roundtable (ILR) scales. In the 1950s, the US government realized the need for a standardized national system to evaluate the foreign language competence of civil servants, and launched the Foreign Service Institute (FSI) to tackle this issue. The US government discovered that there was no agreed upon method or scale to assess the competence of its workers, and by having the FSI handle the language question, it provided a national body that oversees the language competence descriptors regardless of the academic institution or environment in which language learning occurred. The first scales of the FSI ranged from ‘no functional ability’ at 0 to ‘equivalent to an educated native speaker’ at 5. Later the ILR developed descriptors of language skills level for government use. ACTFL developed the *Guidelines* based on the ILR but for academic purposes and from an academic perspective.

In a monolingual yet multicultural society, the purpose of ACTFL is to have a “national organization dedicated to the improvement and expansion of the teaching and learning of all languages at all levels of instruction (ACTFL 1986, p.3).” The ACTFL focuses on teaching foreign languages (neither first, second, minority, nor regional), in a manner that shares the same philosophy as the ILR’s or the FSI’s. It is constituted of 9,000 members who are mostly foreign language educators and administrators from elementary through graduate education, with some representation from industry and governments.

ACTFL’s main purpose is to provide leadership for the teaching and learning of foreign languages in the US. It sets Proficiency Guidelines, and holds a leadership role in the creation of national standards for language competence. According to its vision statement of 2005:

Believing that language and communication are at the heart of the human experience, that the US must nurture and develop indigenous, immigrant, and world language resources, and that the US must educate students to be linguistically and culturally prepared to function as world citizens, ACTFL is uniquely positioned to lead this endeavor (ACTFL, 2001).

Thus, ACTFL provides a “citizen of the world” perspective to promoting language learning that is far from an integration model for immigrants such as that of the Canadian model. The humanitarian and the cosmopolitan focus that is targeted by this notion of “citizen of the world” is in itself a political project or agenda in which the American citizen stands clearly identified vis-à-vis the world. This notion is further solidified by the word “foreign” in ACTFL: s/he, American, monolingual, appreciates others that s/he consider ‘foreign’ to his or her linguistic and cultural identity.

Not just anyone can use the *Guidelines*. In order to be able to use the *Guidelines*, a person needs to be certified as an ACTFL rater. The process is a lengthy and costly one. First, the individual needs to attend a workshop organized by ACTFL in which initial training and materials are offered. Second, s/he proceeds to the certification process that could take up to a year from the date of the workshop since it requires rating a large number of samples at all competence levels and sublevels of the *Guidelines*. Consequently, using the *Guidelines* is very restricting and unless one is well informed and trained, s/he cannot use the *Guidelines*. Unfortunately, that is exactly what happened for many years during which many decision-makers would choose the *Guidelines* as their point of reference in language test development, yet they did not have the appropriate training and expertise to use them. It should be mandated that all raters, regardless of the scale being used, must undergo extensive training.

Common European Framework of Reference for Languages

The Language Policy Division within the Council of Europe oversees the developments of the *Framework* as part of its activities “to promote linguistic diversity and language learning in the field of education that is carried out within the framework of the European Cultural Convention. (Council of Europe, para 3)” This convention, which was ratified by 49 states in 1954, was established “to foster among the nationals of all members, and of such other European States as may accede thereto, the study of the languages, history and civilization of the others and of the civilization which is common to them all.” The role of each state is “to take appropriate measure to safeguard and to encourage the development of its national contribution to the common cultural heritage

of Europe,” and to “encourage the study by its own nationals of the languages, history and civilization of the other Contracting Parties [...] and to promote the study of its language or languages, history and civilization in the territory of the other contracting parties.” (Articles 1 and 2 of the Convention). Furthermore, the 1998 European Charter for Regional or Minority Languages whose aim is

to protect and promote the historical regional or minority languages of Europe” was adopted “on the one hand to maintain and to develop Europe’s cultural traditions and heritage and on the other, to respect an inalienable and commonly recognized right to use a regional or minority language in private and public life (Council of Europe, 1992).

There are three main levels of the *Framework* (Table 3 below). Each of these levels is then divided into two secondary levels. This is the fundamental breakdown of the CEFR. However, the secondary levels can be broken down into more sublevels depending on specific contexts and needs: for instance A1.1, B1.2.

Level “A”		Level “B”		Level “C”	
<i>Basic User</i>		<i>Independent User</i>		<i>Proficient User</i>	
A 1	A2	B1	B2	C1	C2

Table 3 : Fundamental Structure of The Framework Competency Levels

The *Framework* is developed and used within a European perspective of multiculturalism whose main purpose is the establishment of a common multicultural European identity in that the languages and cultural heritage of non-European immigrants are not considered part of this identity and as such do not benefit from any direct official support (ElAtia and Kibbee, 2001). The use of one common *Framework* allows for an easier comparison of language competencies and proficiencies across Europe, mainly for labour and work but also for educational purposes. It ensures, as it is stated in the official publication of the Council of Europe, a reliable comparison between different academic programs in Europe. Language learning within Europe becomes a socialization process in the way that Bernstein (1973) viewed learning a language as a socialization act. It targets mainly people who have learned languages in a school environment, and language learning becomes a social capital, following Bourdieu (1977), allowing a European ‘national’ to invest in this capital with the possibility of economic, geographic, and labor mobility within Europe.

The European notion of multiculturalism differs from that of Canada, in that under the Canadian Multiculturalism Act, languages of the immigrants are promoted and may receive official support, and that of the United States that views multiculturalism as a system in which the individual identifies self with an appreciation of others. In Europe, multiculturalism is restricted to Europe and the native European languages. Given this difference in the fundamental perception of multiculturalism, it is biased and discriminative to say the least, that one set of language competence descriptors would be used in other contexts in which the social foundation, the definition of cultural identity, and the use of languages differ immensely.

Fairness and Validity Issues in Selecting a Set of Guidelines

At the core, all three language competence descriptors provide descriptions for different levels of language competence for specific skills on a specific scale. In comparing the

basic level of the writing skill across these guidelines, even though the wording, the scales and the breakdown of levels change from one set to another, there are elements common to all three sets. These descriptors of a basic level for writing skill are, thus, the fundamental elements of the writing skills at this level of competency in a language. According to the three guidelines, a person at the elementary or beginner level of language competence in writing should be able to:

- make lists
- write short notes
- write very short texts (a paragraph)
- use simple sentence structure
- use basic vocabulary of everyday usage
- use simple tenses (past, present and future)

While the core competence in the writing skill is almost described in a similar way, the appeal of the *Framework* stems from its wording. The *Framework* uses a simplified straightforward language that would first, facilitate translations into other European languages and second, ensure that different stakeholders within the European context are able to relate to it. Several countries within the European Union are using the *Framework* and the wording needs to remain simple and general that would facilitate mobility within Europe, mostly for labor resources. Yet, for each level, this general description gets minutely detailed in the sub-levels: it becomes more complex and would require a language specialist's understanding.

The officials of the *framework* identified as many uses of/for it as possible. Each of these possibilities is specified with further subdivisions in each level of competence. Many users of the *Framework* outside of Europe adopt it without adapting it to the new context and mandate and without paying much attention to these minute differences and hence violate an important element that the *Framework* warns against; in the same way that the CLB center warns against using the *Benchmarks* outside of their context and mandate without further validity studies

[...] alternative uses may be appropriate and encouraged as long as the planner understands that the Benchmarks can serve only as a guideline and that due consideration must be given to the developmental and linguistic needs of the learners. However, use of the assessment tools is not appropriate unless validated for use with that population (Language 2001, para. 6).

Language competence descriptors are written in a generalized language that does not prescribe details of what needs to be assessed, and it is because of this that they serve to narrow down the construct to be assessed, but they are not themselves the construct. The purpose, the audience, the context, and the mandate of any assessment tool differ to a great extent and decision makers in the test development process need to be aware of this issue. The social context and the pragmatics of the assessment tool make it unique to the examinees and examiners (McNamara and Roever, 2006). Regrettably, many decision makers opt to use the *Framework* arguing its validity, yet in reality they are seeking "recognition and equivalence" as Fulcher (2011) put it. They are indeed violating the mandate of the test, and by using the *Framework*, and supporting their claims of validity by quoting studies that have been done in other contexts, they try to justify the link to the purpose and mandate of their testing situation (Kaftandjieva, 2007).

In the case of Alberta, a western Canadian province, the *Framework* has been chosen by some school districts to measure teachers' competence in French. The assessment tools used to evaluate this competence are the DELF and DALF tests (*Diplôme d'études en langue française* and *Diplôme approfondi de langue française*). Both DELF and DALF were developed by the *Centre international d'études pédagogiques* of the French Ministry of Education, and are based and aligned with the *Framework* levels. Even though these two tests are administered in Alberta by the local "Alliance Française," the whole development process takes place in France: test specifications, items drafting, scoring rubrics, and calibration. The danger of such a decision is that these tests fail to address a unique situation proper to Alberta that would cause the test items to be biased against examinees and hence it is discriminative against many competent teachers. It would be the unfair due to language variation and socio-linguistic biases (ElAtia, 2010). The DALF and DELF tests fail to recognize the situation of French in Alberta and that of its minority context, in which pidginization has developed (ElAtia, 2011). These teachers themselves are either francophone from Africa, from Europe, from Eastern Canada, or Franco-Albertans. Among these four groups of native speakers of French, there are linguistic varieties and differences that a test developed in Paris, using the Metropolitan variety of French, does not address nor is aware of. Moreover, most of these teachers were students at immersion schools or learned French as a second language, in an English dominant environment. This group of teachers speaks a different variety of French than the others. The threat of bias and unfairness is looming and many proficient teachers will be discriminated against because of the use of a test that is based on the *Framework* without any consideration of the Canadian context and mandate. This practice violates Standard 1.1 of the Standards for educational and psychological testing (AERA, APA, NCME 1999) that strongly emphasizes clear definition of the population, purpose, and the construct of a test.

Conclusion

Among these three language competence descriptors, the political, social, and cultural implications of the purpose and the agencies behind their development are quite different. The *Framework descriptors* are deeply rooted in the *realpolitik* of the context in which they are used. They are developed and sponsored by European governments to serve a political, social, cultural, and linguistic purpose. Furthermore, labour mobility and the alignment of educational philosophies and legislation within the European context play a major part in the making and the objectives of these competence descriptors.

Given these three contexts, if performance descriptors are used as standards in language testing, they may serve exactly the same political control function as a descriptor for language competence. As such, it would be unethical for stakeholders who are unaware of these mandates to use them outside of their context without adaptation and extensive validity and fairness studies and analyses. Within minority contexts, the position of a language is complex, and without proper adaptation of the competence descriptors, adopting the *Framework* as is for assessment purposes would be unfair. In the situation of western Canada, where French is spoken in a minority context within a diverse socio-linguistic environment that includes many varieties of French and differing French competencies, the use of any language competence descriptor in constructing language tests must be strongly validated.

There are many performance descriptors around the world, and they are all linked to a political vision, nested within a cultural and social framework that necessitates further studies when applied to other settings. They are also used for specific context and follow a certain mandate. History, culture, and perception about language learning are at the core of any language performance descriptor, and as a result they cannot be disassociated from their context and used elsewhere without an acute awareness of these differences. Although at the core, the performance descriptors describe more or less the same skills, it would be erroneous to consider them similar and to use them interchangeably without being aware of the mandate and context behind their development.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. 1999. *Standards for educational and psychological testing*. Washington, DC : American Psychological Association.
- American Council on the Teaching of Foreign Languages. 1986. *ACTFL Proficiency guidelines*. Washington, DC : ACTFL.
- American Council on the Teaching of Foreign Languages. *Mission Statement*. <http://www.actfl.org> [Retrieved March 10, 2011].
- Alderson, C., Clapham, C., Wall, D. 1995. *Language Test Construction and Evaluation*. Cambridge : Cambridge University Press
- Bachman, L. 1990. *Fundamental Considerations in Language Testing*. Oxford : Oxford University Press.
- Blackledge, A. 2009. « “As a Country We Do Expect”: The Further Extension of Language Testing Regimes in the United Kingdom ». *Language Assessment Quarterly*, vol. 6, pp. 6-16.
- Bourdieu, P. 1977. *Outline of a Theory of Practice*. Cambridge and New York: Cambridge University Press.
- Bernstein, Basil. 1973. *Class, Codes and Control: Applied studies towards a sociology of language*. London : Routledge.
- Centre for Canadian Language Benchmarks/Centre des niveaux de compétence linguistique canadiens. 2000. *Canadian Language Benchmarks*. Ottawa: CCLB. www.language.ca [Retrieved August 30, 2011].
- Centre for Canadian Language Benchmarks. 2010. *Appropriate Uses of the Canadian Language Benchmarks Guidelines and Assessment Tools*. http://www.language.ca/display_page.asp?page_id=784 [Retrieved August 30, 2011].
- Council of Europe. 2001. *Common European framework of references for languages: learning, teaching and assessment*. <http://www.coe.int/> [Retrieved March 26, 2011].
- Council of Europe. 1992. *European charter for regional or minority languages*.
http://www.coe.int/t/dg4/education/minlang/Default_en.asp [Retrieved March 26, 2011].
- Davis, A (Ed.). 2004. The Ethics of Language Assessment. *Language Assessment Quarterly*, vol. 1, issues 2 & 3.

Davidson, F., Lynch B. 2000. *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications*. New Haven, CT : Yale University Press.

Davidson, F., Fulcher, G. 2007. « The Common European Framework of Reference (CEFR) and the design of language tests: A Matter of Effect ». *Language Teaching*, vol. 40, issue 3, pp. 231-241.

De Jong, J., Lennig, M., Kerkhoff, A., Poelmans, P. 2009. « Development of a Test of Spoken Dutch for Prospective Immigrants ». *Language Assessment Quarterly*, vol. 6, pp. 41-60.

ElAtia, S. 2010. A study of the effect of Sociolinguistics Bias in the Adaptation and Translation of Bilingual Educational Tests across Canada. Social Sciences and humanities Research Council of Canada, Standard Grant Proposal.

ElAtia, S., in press. Enhancing Foreign Language Writing in Content-Based Courses. *Journal of Theory and Practice in Education*.

ElAtia, S., Kibbee, D. 2001. « Language protection and cultural policy in France ». *The Journal of the Foundation of the Endangered Languages*, vol. 3, issue 1, pp. 28-35.

Fulcher, G. 2011, March 15. OET & CEFR. Message posted to LTEST-L, the Language Testing List-serv, archived at <http://lists.psu.edu/archives/ltest-l.html> [Retrieved March 26, 2011].

Fulcher, G. & F. Davidson. 2007. *Language Testing and Assessment*. London & New York : Routledge.

International Language Testing Association. 1995. *Report of the Task Force on Testing Standards*. www.iltaonline.com [Retrieved March 15, 2007].

International Language Testing Association. 2007. *ILTA: Guidelines for Practice*. www.iltaonline.com [Retrieved March 15, 2011].

Kaftandieva, F. 2007. Quantifying the quality of linkage between language examinations and the CEFR. In : Carlsen, C. and Moe, E. (Eds.) *A Human Touch to Language Testing*. Oslo : Novus Press

McNamara, T. & Roever, C. 2006. *Language Testing: The Social Dimension*. Malden, MA : Blackwell Publishing.

Saville, N. 2009. Language Assessment in the Management of International Migration: A Framework for Considering the Issues. *Language Assessment Quarterly*, vol. 6, pp. 17-29.

Notes

¹ For more information on test construction, and development, the following references provide ample information: Alderson and al, 1995; Bachman, 2001; Davidson and Lynch, 2001, and Fulcher and Davidson, 2007.

² In this article, it would be referred to the "Common European Framework of Reference for Languages" as *framework*, the "Canadian Language Benchmarks" as *benchmarks*, and the "American Council on the Teaching of Foreign Languages guidelines," as *guidelines*.

³ Few months ago, a heated debate among language testing community on the L-Test listserv addressed the issue of Immigration Canada choosing the IELTS (International English Language Testing System) for immigration purposes. The IELTS is developed in The UK by the University of Cambridge ESOL Examinations. Fairness and validity issues were raised.