LANGUAGE ASSESSMENT QUARTERLY, *1*(4), 253–266
Copyright © 2004, Lawrence Erlbaum Associates, Inc.

COMMENTARY

Deluded by Artifices? The Common European Framework and Harmonization

Glenn Fulcher
University of Dundee, Scotland

This commentary provides a critical and historical review of the *Common European Framework of Reference: Learning, Teaching, Assessment* (CEF). It is presented within the context of political and policy issues in Europe, and considers the role that the CEF is likely to play in that context, which is beyond the control of language testers. The dangers of institutionalization through political mandate are explored for test providers, test takers, and score users. It is argued that the CEF should be treated as just one of a range of tools for reporting test scores.

The Council of Europe (COE), not to be confused with the European Union (EU), was founded in 1949 and currently has 45 member states. In Article 1 of the founding Statute (Council of Europe, 1949), its purpose is to

> achieve a greater unity between its members for the purpose of safeguarding and realizing the ideals and principles which are their common heritage and facilitating their economic progress ... by discussion of questions of common concern and by agreements and common action in economic, social, cultural, scientific, legal and administrative matters and in the maintenance and further realization of human rights and fundamental freedoms.

Based in Strasbourg, it is made up of the Committee of Ministers (the foreign ministers of the member states) and the Parliamentary Assembly (not to be confused with the elected European Parliament), which consists of representatives appointed by national parliaments. It is supported by a Secretariat of over 1200 offi-

Request for reprints should be sent to Glenn Fulcher, Center for Applied Language Studies, University of Dundee, Dundee, DD1 4HN, Scotland, UK. E-mail: g.fulcher@dundee.ac.uk

cials. It has oversight of the European Court of Human Rights (not to be confused with the European Court of Justice, which operates under the aegis of the European Union) and the Commission on European Rights. What is now the EU began as the European Coal and Steel Community, which was founded in 1951 with six members, becoming the European Economic Community in 1957. It underwent expansion in 1973, 1981, 1986, and 1995. In 2004, 10 new countries joined, bringing the total number of members to 25. What began as a trading block that would also bring Germany and France into a new relationship after the War, has now developed a much wider political agenda. The COE and the EU are separate institutions that increasingly cooperate on matters of common interest, which are identified through meetings between executives of the (unelected) EU commission and the COE twice yearly. Since 1987, the COE and the EU have jointly sponsored projects that reflect the values of both organizations; some of these are in the sphere of education, language learning, and the recognition of qualifications.

The *Common European Framework of Reference: Learning, Teaching, Assessment* (CEF; COE, 2001) is a COE project that brings together cross-national institutions that are actively encouraged and funded to achieve harmonization, with the overtly political aim of encouraging a common view of European Citizenship. As Milanovic (2002, p. 3) says, the CEF "offers language test designers and those involved in producing examinations the possibility of moving collectively towards a shared language testing system that is motivated by the core values of the Council's own notion of European citizenship."

However, current disagreement over the need for, and content of, a European Constitution shows how sensitive the populations of member states are to these harmonizing tendencies, which are frequently interpreted as further steps toward political union with common policies on all issues from defense to taxation. In some more Euroskeptic states like Britain, this has also led pro-European politicians to offer referendums rather than continue implementing harmonization that cedes sovereignty to European institutions without recourse to the people, and in the EU parliamentary elections on June 10, 2004 no less than 12 British candidates were elected from the UK Independence Party. In education there has been less dissent, and harmonization is increasingly seen as a necessity. Kelly (2004) writes,

> Higher education provides a microcosm of the broader picture as Europe attempts to respond to global changes by creating a European Higher Education Area, capable of matching the US and Japan by 2010. That is a tall order and, as the EU Commission has warned, unlikely to be met without urgent reforms. In response, the package of changes known as the Bologna Process[1] is likely to be accelerated, with resulting opportunities and threats. (p. 1)

[1]The Bologna Process is a joint COE/EU project to harmonize higher education throughout Europe. The European Space for Higher Education is a commitment by 29 countries to reform their higher education systems to create convergence at the European level. Related documentation can be found online at www.enic-naric.net

In language testing and assessment, there is also a desire for harmonization using the CEF as a tool. In some quarters this has led to an over-simplified approach to validity issues. As the language learning press in Europe increasingly propound,

> The idea is relatively simple. Using a notional and functional taxonomy drawn from the Common European Framework, tests in languages are created so as to give unified, reliable and valid test results no matter what the language. (EL Gazette Examinations Supplement, 2004)

The question remains why it is necessary to harmonize—which in the case of higher education means introducing a common structure, credit rating, and content comparability for degree programs across Europe—to have a system of qualification recognition. It could equally be argued that harmonization means less diversity, and less choice, with one degree program looking very much like another. Rather than tackling the difficult questions, higher education appears to be increasingly driven by the political mandate set out in the Bologna declaration.

THE COMMON EUROPEAN FRAMEWORK

The CEF was published in 2001 and is rapidly becoming the standard reference document for teaching and testing languages in Europe. The framework consists of six levels, each associated with a set of descriptors, three of which are also defined by fixed point descriptive documents that pre-date the CEF.

Despite its relatively recent impact upon Europe, it has a history going back 30 years. In the 1970s, the COE initiated a number of symposia to evaluate the possibility of a European credit scheme for language learning, related to fixed points on a framework (van Ek, 1975). The first of the so-called fixed points was the *Threshold Level* (describing an independent language user), which was published in 1985 (van Ek, 1975) and republished in 1990 (van Ek & Trim, 1990a). Threshold was intended to be a statement of what a language learner needed to be able to do to communicate independently in the target language, with the emphasis firmly on social situations and interaction. This descriptive document has now been attached to CEF level B1 (see Figure 1). Also in 1990, Waystage (van Ek & Trim, 1990b) was published. This is described as the halfway point to *Threshold*, and has been attached to CEF level A2. The publication of these documents came just as a new COE project, "Language Learning for European Citizenship," was getting underway to develop a "comprehensive, transparent, and coherent common European framework" of language learning and assessment. In 2001 the last fixed point, the Vantage Level (upper intermediate learners) was published, and has been attached to CEF level B2 (van Ek & Trim, 2001).

It should be remembered that these documents were written and published before the CEF, and it was decided that they should help define three of the levels on

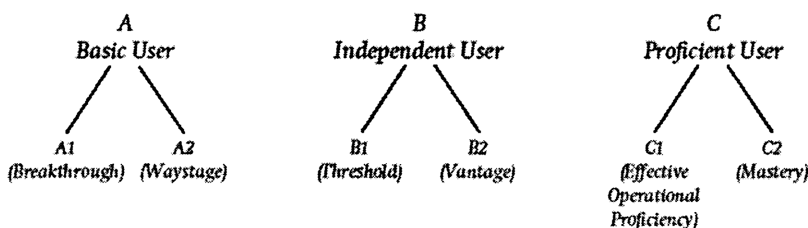

FIGURE 1 Common European Framework of Reference levels.

the new scale to preserve continuity in the work of the COE (John de Jong, e-mail to LTEST-L, March 17, 2004; North, 2000). The documents themselves are purely descriptive, and the distance between Waystage and Threshold is not based upon any empirical evidence, but the intuition of the authors. The content consists of lists of situations, functions, linguistic elements, competencies, and (since 1990) discourse and strategic abilities. As there is no attempt to state how well learners at a level will be able to perform in any of the elements within a list, or whether the system would be compensatory or all or nothing, it is difficult to know what might constitute being a "threshold person." Alderson (2004) has also undertaken a detailed content comparison of Threshold and Waystage, and shows that, apart from different structural organization, the content is reproduced in both levels, sometimes (as with learning to learn and compensation strategies) with identical wording. Where the two levels differ it is often in level of detail or claims that would be difficult to justify, such as the ability of Threshold level learners to use high-fall intonation that is not present in Waystage. It is, therefore, arguably the case that it would be difficult to detect a meaningful difference in the move from A2 to B1. The rationale for attaching these documents to CEF levels is found in the term "natural levels," which are "the conventional, recognized, convenient levels found in books and exams" (North, 1992, p. 12). Despite the problem of circularity, and the rider that "natural" is not associated with "nature here," the use of the term does nothing to undermine the popular view among teachers in Europe that these levels describe linguistic and communicative reality.

The published rationale for the development of the CEF, incorporating the earlier fixed-point descriptions, was twofold: first, to support the introduction of a European Language Portfolio upon which learners progress in acquiring European languages could be recorded (Schärer, 1992), and, second, to provide a means of comparing existing language tests for certification of learning that would have Europe-wide recognition (North, 1992). The official political impetus (Recommendation R(82)18 and r(98)6 of the Committee of Ministers of the Council of Europe) was to encourage mobility, to promote understanding and cooperation, and to overcome prejudice and discrimination (COE, 1996, 2001). The first draft of the

CEF was subtitled "Language learning for European citizenship." The final CEF (2001) contains sets of scales, and a number of descriptive chapters that provide an encyclopedic list similar to those in the fixed-point documents. These lists are unrelated to the scales.

The Development of the CEF Scales

It is important to understand how the scales in the CEF were developed to evaluate their limitations and potential use. This is described in North (2000), and summarized in Fulcher (2003a, pp. 107–113). Essentially, there were four phases in the developmental work: (a) intuitive analysis, (b) qualitative analysis with informants, (c) quantitative analysis using questionnaires, and (d) replication. There were a number of discrete steps within each phrase of the process.

In Phase A, the designers collected 30 existing rating scales. In total, these contained 2000 proficiency level descriptors. These were put into sentence length descriptors and placed intuitively into the six a-priori levels. Each descriptor was coded by the communicative activity to which it referred, and where gaps in the system were perceived, new descriptors were written to fill the gaps. In Phase B, teachers were asked to evaluate the descriptors for relevance to their learners, and then told to put them in piles according to whether they represented low, middle or high proficiency levels. The descriptors that were classified most consistently by the teachers were then compiled into questionnaires with approximately 50 items to cover the 6 levels of the intended scale. Some of the items overlapped with the level above and below. These questionnaires were presented to teachers for data collection in Phase C. A Likert-type item was attached to each descriptor on each questionnaire, asking teachers to decide which descriptors defined a level that was below, at, or above the level of their students. Multifaceted Rasch analysis was then used to provide difficulty estimates for each of the descriptors and then misfitting descriptors removed from the pool. From the surviving descriptors, two-thirds of those on spoken interaction, production, and competence are taken from Eurocentres rating scales (North, 2000, p. 337).

The descriptors were then scaled, and cut points set to place them into the six apriori levels, with two levels each within elementary, intermediate and advanced. In the CEF, these are now known as levels A: Basic User, B: Independent User, and C: Proficient User.

Finally, in Phase D, the process was repeated with a new group of teachers, to discover if the same sequences could be replicated. North (2000, p. 339) reports a correlation of .99 between the sequence produced in the development and replication study, and further studies are summarized in North (2000, pp. 573–574), primarily using self-assessment.

The approach to scaling was principled, and appears from current evidence to be stable. However, as North and Schneider (1998, pp. 242–243) admit, the result-

ing CEF scale is essentially atheoretical. Specifically, "what is being scaled is not necessarily learner proficiency, but teacher/raters' perception of that proficiency—their common framework" (North, 2000, p. 573). The word "common" in the title, therefore, refers to agreement between teachers in the study samples when sequencing descriptors. The levels into which the descriptors are placed are "natural," in that they reflect the way European teachers, publishers and testers think of language levels in terms of elementary, intermediate and advanced. However, it was always intended that "common" should have a wider implication. Namely, that the CEF "seeks to make it easier for teachers, learners, publishers and testers to communicate across languages, educational sectors and national boundaries" (North, 2004). Hence the requirement for creating benchmark work samples at each of the levels of the CEF, and training teachers to think of their students in terms of these levels.

The requirement to think in terms of the framework through training is part of the process of institutionalization that is familiar from the spread of the American Council on the Teaching of Foreign Languages (ACTFL) Guidelines in the US during the 1980s (Lowe, 1986), so that eventually there is little dissent from the common interpretation.

The Problem With Framework

The notion of a scale that describes levels of proficiency across languages for use in teaching and testing is not new. (This is the other use of common in the language of frameworks.) The first such scale was developed during the 1950s by the Foreign Service Institute for use by the American military. By the 1960s, this had evolved into the *Interagency Language Roundtable* scale, which is still in use today. Its Civilian counterpart, the ACTFL Guidelines, has become the standard framework for all modern language teaching and testing in the US (Liskin-Gasparro, 2003). Within the UK, the English Speaking Union published a framework for the same purpose in 1989 (Carroll & West, 1989), although it is no longer in use because of institutional objections to where tests were placed against scales and the resulting comparisons that were made between tests.

The real problem is that the meta-objective of providing proficiency descriptors that are applicable across languages requires a framework so abstract that it is not a framework, but a model. Milanovic (2002) is explicit in referring to the CEF as reflecting communicative theories of language testing and the Bachman (1990) model. North (2004) also states that

> the CEF draws on theories of communicative competence and language use in order to describe what a language user has to know and do in order to communicate effectively and what learners can typically be expected to do at different levels of profi-

> ciency. It doesn't try to define what should be taught (content specifications), let alone state how it should be taught (methodology).

Similarly, North (2000, p. 571) writes, "The CEF consists of a taxonomic descriptive scheme, the core of which is not dissimilar to Bachmans (1990) model of communicative language ability ... together with a set of common reference levels."

However, the taxonomy in the CEF document is related to the fixed-point documents. These draw on other models and theory intuitively. The CEF level descriptors are atheoretical. This is why there is no link between the taxonomy and the levels, so that

> the main chapters illustrating the CEF, Chapters 4, 5 and 7, which give details of themes, purposes, activities, strategies, texts, processes, competences and tasks, do not specify which of these might be expected to apply to which level, rather than another. (Alderson, 2004, p. 1)

A similar problem was faced by users of the ACTFL Guidelines (see Chalhoub-Deville & Fulcher, 2003), where the intention was to develop tasks that could be done at each level of the scale, latterly analyzed in terms of the functional trisection (Liskin-Gasparro, 1984, p. 35). This is not possible when an intuitive taxonomic model is placed alongside, but unrelated to, a set of scaled descriptors.

Chalhoub-Deville (1997) provides the conceptual tools for deconstructing the term "framework" in the same way that we have attempted to deconstruct the term "common" in the preceding paragraphs. Models (or any attempt to produce encyclopedic descriptions of proficiency) are by their very nature abstract and far removed from specific testing purposes and contexts. She considers the Unitary Competence Hypotheses, the cognitive/academic language proficiency and basic interpersonal and communicative skills (CALP/BICS) model, communicative competence, and communicative language ability. But the CEF might equally have been cited as an example of a description operating at this level of abstraction. True frameworks, on the other hand, need to mediate between the abstract and the context of a particular test; a framework operationalizes those parts of a model that are relevant to a specific test purpose, and makes possible the production of test specifications.

Indeed, this is the problem that is now being recognized in practice (Weir, 2004). Alderson et al. (2004) argue,

> The CEF, being such a comprehensive description of language use, can also be considered, implicitly at least, as a theory of language development, but the CAN-DO scales for reading and listening present a taxonomy of behaviors rather than a theory of development in listening and reading abilities. Moreover, it is far from clear that the abstract statements in the CEF can be turned into items that illustrate or exemplify the different CEF levels. (p. 2)

The aim of the Dutch CEF project is to "develop a frame for analysis of tests and specifications" (Alderson et al., 2004, p. 3). The use of the word "frame" shows that this is different from the framework in that it is more detailed, and can relate directly to test specifications. The real problem lies in the use of the word "framework" in the CEF, for it is not a framework in the sense of a document that allows the generation of test specifications, but much more like a model without actually being a model. Perhaps Spolsky (E-mail to LTEST-L, March 16, 2004) has characterized the CEF more accurately as a useful heuristic that may help curriculum developers or test developers think about what they may eventually include in a framework.

THE DANGERS OF INSTITUTIONALIZATION

The key problem is that once a framework is institutionalized, the process of reification is quick and not easily reversed. Although the CEF documentation itself (even in the first consultation draft of 1996) states that "The construction of a comprehensive, transparent and coherent Framework ... does not imply the imposition of one single system" (Council of Europe, 1996, p. 3), it is rapidly becoming "the" system. For teachers, the main danger is that they are beginning to believe that the scales in the CEF represent an acquisitional hierarchy, rather than a common perception. They begin to believe the language of the descriptors actually relates to the sequence of how and what learners learn.

For many producers of tests, one of the dangers lies in the desire to claim a link between scores on their tests and what those scores mean in terms of CEF levels, simply to get recognition within Europe. They do not have any choice in this, for if institutions begin to believe that the CEF is the truth against which all else must be measured, failure to claim a link to the CEF would equate to a commercial withdrawal from continental Europe. Some test providers claim this link through standard setting (Tannenbaum & Wyile, in press); most are prepared to make an intuitive guess and print this in guides and on web sites. Apart from the political forces at work, the CEF will be more successful than the ESU Framework precisely because test producers can fit their tests to that part of the scale that matches their target markets and ambitions to be compared with prestigious large-scale tests.

Even more critical are dangers for score users. The most important relates to the stated aim of making it possible to compare tests through the mediation of the link to the framework. By producing maps, the users are encouraged to think that a score of "X" on a UK test is equivalent in meaning to a score of "Y" on a US test, and "Z" on an EU test. This has long been a European goal. As Carroll and West (1989) said of the ESU project:

> How can students decide which is the most appropriate examination for them to take? And how does a future employer find out from all this confusing information just what standard of English has been acquired by an applicant presenting a Certificate? (p. 1)

The problem with this use of the CEF is that it is not possible to use a description at the model level to meaningfully link tests that have been designed for different purposes, and hence with a variety of different construct definitions. Specific tests may relate to the model level only in that they represent possible selections of constructs from the currently defined universe of constructs. Linking to the CEF cannot, therefore, provide equivalence of meaning across tests. The provision of widely available maps that do not carry health warnings (e.g., Morrissey, 2003) encourage score and test misuse. A much more considered position is that of Taylor (2004), in which she observes that the ESU framework "implied greater degrees of comparability than was actually justified, it risked oversimplification or misinterpretation and this limited its usefulness to the users for whom it was intended" (p. 4). Taylor realizes that test users, like other consumers, will compare products across a range of criteria, but that does not mean that tests can be considered equivalent in score meaning through linkage with the CEF.

Linking Tests to the CEF

Linking tests to the CEF is not a simple matter. The CEF scale has no underlying theory, and there are no content specifications attached to the levels. Further, many tests that now claim to be linked to the CEF do not themselves have a theoretical basis, or known reliability. The linking is mostly intuitive. To educate test producers, the Council of Europe has commissioned Guidelines, *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching and Assessment* (COE, 2003), the preliminary pilot version of which is now available on the COE web site. This document focuses on linking assessments to the CEF through what Mislevy (1992) called social moderation, and which others agree is the weakest form of linking assessments (Linn, 1993). However, this is in keeping with the meaning of "common" previously outlined. As North (2000, p. 570) states, "Social moderation is, however, potentially powerful because it implies the development and validation of a *common* view of a construct" (p. 570, emphasis added).

The Manual aspires to encourage transparency on the part of examination providers, as well as creating networks of institutions and professionals who will be involved in linking tests to the CEF. It rightly stresses the need for theoretical and empirical evidence to support the validation of the claim, and usefully states that it is the responsibility of test providers to investigate the reliability and validity of

their own tests before attempting to establish a link. It recommends that the claim for a link should be based on specification of test content and mapping this to the content of the CEF (scales and fixed point documents), the standardization of judgments pertaining to judging test item difficulty in relation to CEF levels in relation to benchmark samples, and empirical validation to the CEF through an independent measure. The assumption underlying the document is that linkage to the CEF will result in comparability of qualifications across Europe, and that, through the standardization process, language professionals will increasingly see the world through the common interpretation of the CEF levels.

Power and Control

The institutionalization of the CEF throughout Europe has been rapid. This is primarily because, unlike the ACTFL Guidelines in the United States, there is a strong political agenda at work. Some test producers explicitly acknowledge this in their claims. For example, the European Consortium for the Certificate of Attainment in Modern Languages (ECL; 2004) Web site states,

> In accordance with the EU unification policy the Member States that formed the consortium wanted to ensure equivalency and recognition of the certificates in each language without having to be validated nationally (nostrification).

Although this represents the strongest end of the cline, it is not unusual to find such sentiments expressed. This is what Davidson and Lynch (2002) would call the test mandate, which is an external political force for harmonization of systems and thought across Europe that would ultimately make political union more palatable to the citizens of member states.

Such mandates are also difficult to control, because their nature changes over time, and language testing may be forced into new service. For example, citizenship appears to have been originally conceived as a vague notion relating to communicating with fellow-Europeans in their own language to reduce prejudice and injustice and encourage mobility. Although this was certainly a noble aim, it is now being conceived of literally in some parts of Europe. Language testing for asylum seekers and immigration is on the increase (Fulcher, 2003a), and governments are increasingly using language requirements as a method of control. In the U.K., for example, the Immigration and Asylum Act of 2002 requires potential immigrants to demonstrate sufficient knowledge of English, Welsh or Gaelic. In the development of a German test for immigration, Ruebeling (2003) notes that "quite often during the meetings of the team of test developers concerns were voiced whether the experts should carry out this development at all" (p. 2). But the tests are nevertheless being produced to be pitched at what is considered the appropriate CEF level for immigration—even though there does not appear to be any research in this

area. Not surprisingly, we are also beginning to see unsubstantiated claims about the number of hours of study required to reach a B1, in the same way that such claims emerged in the early years of the ACTFL Guidelines.

The CEF has become a potent symbol of the power of language testing within Europe. As a centralized transnational framework, it is virtually impossible for institutions to resist seeing the world through the same spectacles as the COE (Shohamy, 2001). As its power and influence grow, voices of dissent or warning are increasingly unwelcome. It is, therefore, even more important that issues are raised in journals and the press (see Fulcher, 2004).

CONCLUSION

The CEF is a fascinating case study in the way political and social agendas can impact on language testing, and how language testing can be made to serve those agendas. It is sometimes difficult to tease these agendas out from the laudable desire to increase communication across the continent and reduce prejudice. And because the use of the CEF is tied to the political notion of being European[2] it is inevitable that there will be many who will disagree with this analysis, which is not dissimilar in nature to that of Mill (1859, pp. 80–81):

> Europe is, in my judgement, wholly indebted to this plurality of paths for its progressive and many-sided development. But it already begins to possess this benefit in a considerably less degree. It is decidedly advancing towards the Chinese ideal of making all people alike.

Technical language testing issues relating to models and frameworks aside, we should therefore recognize that language testers will react differently to the institutionalization of the CEF on the basis of their political beliefs. The COE has done much to establish individual freedoms and rights; among those are the right to dissent and to oppose moves to political unification that puts the individual voter at two or more removes from the institutions that make decisions affecting their lives. This becomes even more pressing when there is the possibility of ceding authority to non-elected bodies. These tendencies take democracy too far away from the individual, and create a growing sense of disenfranchisement. Citizens of European countries have a right to be concerned about national sovereignty if giving it away reduces the value and impact of their vote on the political system, and hence upon the legal, social and financial frameworks

---

[2]For example, there was considerable disquiet on the discussion list of the European Association for Language Testing and Assessment (EALTA) over the ETS Standard Setting Study to link TOEFL and TOEIC to the CEF in February 2004.

within which they live. In language testing, as in other policy areas, it is therefore critical to be aware of harmonization that may lead to further political unification by stealth, irrespective of whether the framework is a suitable tool for this purpose or not.

However, this does not mean that the CEF may not play a useful role language testing in Europe or elsewhere. Alderson (1991) differentiated between three types of scales: constructor-oriented, assessor-oriented and user-oriented. User-oriented scales are developed to provide a useful, understandable, reporting tool for score users. We have seen that the CEF is proving difficult to implement in other settings (Alderson et al., 2004), and should not be used to compare tests by mediation, but scaled can-do statements are ideal for reporting a generalizable meaning of test scores to users, in terms of what a test taker with a particular score on a given test may typically be able to do. Indeed, North (1992) indicated that this was always intended to be a use of the CEF:

> Scales vary in their purpose and audience and the degree of detail which they give. ... It is with user-oriented scales that we are concerned in connection with the European Language Portfolio. (p. 17)

Continually updating and refining the CEF (which North, 2000, p. 574, recommends) would also send the message that it is neither set in stone nor represents external truth. Rather, it is a tool, and like any tool can only be used for the limited job for which it is suitable.

This article is written in the spirit of promoting discussion and debate of frameworks and their limitations in the context of political mandates that promote harmonization of the kind being experienced within Europe, rather than guidelines for professional practice (AERA, 1999), for

> It is only in the earlier stages that any stand can be successfully made against the encroachment. The demand that all other people shall resemble ourselves, grows by what it feeds on. If resistance waits till life is reduced nearly to one uniform type, all deviations from that type will come to be considered ... contrary to nature. (Mill, 1859, p. 82)

REFERENCES

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.

Alderson, J. C. (1991). Bands and scores. In J. C. Alderson & B. North (Eds.), *Language testing in the 1990s* (pp. 71–86). London: Modern English Publications and the British Council.

Alderson, J. C. (2004). *Waystage* and *threshold*. Or does the Emperor have any clothes? Mimeo.

Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2004, February). *Specifications for item development and classification within the CEF: Reading and listening (English, French and German): The Dutch CEF Construct Project.* Paper presented at a workshop on research into and with the CEFR, University of Amsterdam.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford, UK: Oxford University Press.

Carroll, B. J., & West, R. (1989). *ESU framework: Performance scales for English language examinations.* London: Longman and the English Speaking Union.

Chalhoub-Deville, M. (1997). Theoretical models, assessment frameworks and test construction. *Language Testing 14*, 3–22.

Chalhoub-Deville, M., & Fulcher, G. (2003). The oral proficiency interview: A research agenda. *Foreign Language Annals 36*, 498–506.

Council of Europe. (1949). *Statute of the Council of Europe, European Treaty Series 001.* Retrieved May 26, 2004 from http://conventions.coe.int/Treaty/EN/cadreprincipal.htm

Council of Europe. (1996). *Common European framework of reference for language learning and teaching: Language learning for European citizenship.* Strasbourg, France: Author.

Council of Europe. (2001). *Common European framework of reference for language learning and teaching.* Cambridge, UK: Cambridge University Press.

Council of Europe. (2003). *Relating language examinations to the common European framework of reference for languages: Learning, teaching, assessment.* Strasbourg, France: Author.

Davidson, F., & Lynch, B. K. (2002). *Testcraft.* New Haven, CT: Yale University Press.

ECL. (2004). European Consortium for the Certificate of Attainment in Modern Languages. Retrieved April 9, 2004 from http://www.cambridge-efl.org/rs_notes/rs_nts15.pdf

El Gazette Examinations Supplement. (2004). Testing, testing, 1 2 3. p. 1.

Fulcher, G. (2003a, April 17). Few ills cured by setting scores. *Education Guardian* [online edition]. Retrieved May 28, 2004 from http://education.guardian.co.uk/tefl/story/0,5500,937988,00.html

Fulcher, G. (2003b). *Testing second language speaking.* London: Longman/Pearson Education.

Fulcher, G. (2004, March 18). Are Europe's tests being built on an unsafe framework? *Education Guardian* [online edition]. Retrieved April 22, 2004 from http://education.guardian.co.uk/tefl/story/0,5500,1170569,00.html

Kelly, M. (2004). Engaging with Europe. *Languages, Linguistics and Area Studies, 5*, 1.

Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education 16*, 83–102.

Liskin-Gasparro, J. E. (1984). The ACTFL proficiency guidelines: A historical perspective. In T. V. Higgs (Ed), *1984 Teaching for proficiency, the organizing principle* (pp. 11–42). Lincolnwood, IL: National Textbook Company.

Liskin-Gasparro, J. E. (2003). The ACTFL proficiency guidelines and the oral proficiency interview: A brief history and analysis of their survival. *Foreign Language Annals, 36*, 483–490.

Lowe, P. (1986). Proficiency: Panacea, framework, process? A reply to Kramsch, Schulz, and particularly Bachman and Savignon. *Modern Language Journal, 70*, 391–397.

Milanovic, M. (2002). *Common European framework of reference for languages: Learning, teaching, assessment: Language examining and test development.* Strasbourg, France: Council of Europe Language Policy Division.

Mill, J. S. (1859). *On liberty.* In J. Gray (Ed.), *John Stuart Mill on liberty and other essays.* Oxford, UK: Oxford University Press.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods and prospects.* Princeton, NJ: Educational Testing Service Policy Information Center.

Morrissey, M. (2003). *Proficiency assessment benchmarks.* Retrieved May 28, 2004 from http://www.mdmorrissey.com/profassess.htm

North, B. (1992). Options for scales of proficiency for a European Language Framework. In R. Schärer & B. North (Eds.), *Toward a common European framework for reporting language competency* (pp. 9–26). Washington, DC: National Foreign Language Center.

North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.

North, B. (2004, April 15). Europe's framework promotes language discussion, not directives. *Education Guardian* [online edition]. Retrieved May 22, 2004 from http://education.guardian.co.uk/tefl/story/0,5500,1191130,00.html

North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing, 15*, 217–263.

Ruebeling, H. (2003, Autumn/Winter). Test Deutsch alfa–A German as a foreign language test for people who cannot read or write. *ALTE News*, 2–3.

Schärer, R. (1992). A European language portfolio. In R. Schärer & B. North (Eds.), *Toward a Common European Framework for Reporting Language Competency* (pp. 3–7). Washington, DC: National Foreign Language Center.

Schärer, R., & North, B. (1992). *Toward a common European framework for reporting language competency*. Washington, DC: National Foreign Language Center.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman/Pearson Education.

Tannenbaum, R. J., & Wylie, E. C. (in press). *Mapping test scores onto the common European framework: Setting standards of language proficiency on the Test of English as a Foreign Language (TOEFL), the Test of Spoken English (TSE), the Test of Written English (TWE), and the Test of English for International Communication (TOEIC)*. Princeton, NJ: Educational Testing Service.

Taylor, L. (2004). *Issues of test comparability*. Research Notes 15. Cambridge, UK: UCLES, 2–5. Retrieved September 20, 2004 from http://www.cambridge-efl.org/rs_notes/rs_nts15.pdf

van Ek, J. A. (1975). *The Threshold level in a European unit/credit system for modern language learning by adults*. Strasbourg, France: Council of Europe.

van Ek, J. A., & Trim, J. L. M. (1990a). *Threshold level 1990: Modern languages*. Strasbourg, France: Council of Europe.

Van Ek, J. A., & Trim, J. L. M. (1990b). *Waystage*. Cambridge, UK: Cambridge University Press.

Van Ek, J. A., & Trim, J. L. M. (2001). *Vantage level*. Cambridge, UK: Cambridge University Press.

Weir, C. J. (2004). *Limitations of the Council of Europe's Framework (CEF) in developing comparable examinations and tests*. Mimeo.