30 Years on—Evolution or Revolution?

Tim McNamara

To cite this article: Tim McNamara (2014) 30 Years on—Evolution or Revolution?, Language Assessment Quarterly, 11:2, 226-232, DOI: 10.1080/15434303.2014.895830

To link to this article: http://dx.doi.org/10.1080/15434303.2014.895830

Published online: 21 May 2014.
EPILOGUE

30 Years on—Evolution or Revolution?

Tim McNamara

University of Melbourne, Parkville, Victoria, Australia

This epilogue considers the set of articles included in the issue as a response to the question posed by Morrow (1979): Communicative language testing—Evolution or revolution? Whereas the other articles in the issue would suggest that we are in a phase of gradual evolution, with much continuity since 1980, this article argues that, on the contrary, we are facing a number of challenges to communicative language testing and the testing of languages for specific purposes. These challenges come from two principal sources: (a) the advances in technology that are making possible the automatic scoring of speech and writing, and the associated return to psycholinguistic, even structuralist, models of proficiency; and (b) the need to reflect in language test constructs and practice the reality of English as a lingua franca communication. The article considers these issues in the light of the influence of the Common European Framework of Reference in language testing, an institution that it seems is now considered “too big to fail.”

The provocative title of Keith Morrow’s (1979) paper, which was one of the triggers to the discussion commemorated in this issue, was “Communicative Language Testing: Evolution or Revolution?” As the articles by Alan Davies and Lynda Taylor in this issue in their different ways make clear, the year 1980 represented, if not a revolution, then at the very least a watershed moment in language testing. A number of developments and factors coincided to cause this. First, the prevailing orthodoxy of the 1970s, Oller’s (1979) psycholinguistic model of performance had been subject to effective critique, not least Alderson’s exposure of its Achilles heel: the claims of the cloze test to be superior to and different in kind from the structuralist tests of grammar and vocabulary, which it was intended to replace\(^1\). The critique set the stage for the eclipse of indirect language testing for two decades, although as we see, this is now the subject of a major issue.

\(^1\) The critique was originally published as Alderson (1979), but more usefully, see the republication of the paper together with an exchange with Oller in Alderson (1983).
comeback—quite apart from leading to the regrettable loss to the field of Oller, “a creative thinker and gifted researcher” in Davies’s words. Second, as several of the articles note, 1980 was also the year of publication of Canale and Swain’s (1980) landmark article defining communicative competence, which became a reference point for communicative language testing for three decades (Bachman, 1990; Bachman & Palmer, 1996, 2010; McNamara, 1996). Third, a shift in the centre of gravity of language testing away from the United States saw language testing catch up with the rapid and creative developments in language teaching which marked the previous decade in Britain. The 1970s had been a heady time for language teaching in the United Kingdom, with rapid advances in theory matching rapid advances in practice, particularly facilitated by the oil money that generated enormous opportunities for developments in curriculum and pedagogy. The new work in communicative testing was able to draw on the exciting work of theorists of discourse in the flourishing schools of applied linguistics in Edinburgh, Lancaster, Birmingham, and Reading. John Trim and David Wilkins had worked for a decade on communicative syllabus design for the Council of Europe, the precursor of the Common European Framework of Reference for languages (Council of Europe, 2001). In specific purpose contexts, particularly the testing of the English of international students, Brendan Carroll (1981) had designed a communicative test known as the English Language Testing System (ELTS)—the precursor of the current IELTS test—to replace the existing English Proficiency Test Battery. As Davies (this issue) points out, the most radical and creative general proficiency test of the period, the Royal Society of Arts test of the “Communicative Use of English as a Foreign Language” was a direct response to Morrow’s (1977) vision of what a communicative test should be. This brilliant boutique test sadly did not survive its migration to Cambridge, but it symbolized well the revolutionary impulse of the period.

The question inevitably arises, 30 years on: Are we in another revolutionary period, or is it a question of a more quiet evolution? The articles in this issue clearly suggest the latter. In this article I want to argue for the opposite, that we are at a moment of crisis in the current version of communicative language testing, and that it is potentially time for another revolution.

The discussion of general proficiency in Harsch’s article focuses, unsurprisingly, on the Common European Framework of Reference (CEFR), which dominates the scene not only in Europe but increasingly throughout the world. This extraordinary phenomenon could not have been anticipated in 1980, and the field still struggles to grapple with its significance. As the history of its origins makes clear, the CEFR is primarily a policy coordination and administrative initiative, acting as an accounting system and management tool whereby control is exercised by specifying the outcomes of learning independently of any specific test (or language; McNamara 2011). Policymakers need tools that serve their needs, which are for accountability, administrative ease, ease of explanation to stakeholders, “scientific” respectability, and so on. The CEFR, with its pyramidal shape (culminating in the six numbered reference levels), is such a tool. The functionality of a universal letter/number system to code the six levels is the key feature of the CEFR, which makes it attractive to administrators and policymakers.

The relation of the CEFR to the construct of proficiency is complex. There seem to be two positions. On one hand, the CEFR can be seen as a powerful, even dominating statement of a particular view of language proficiency, anchored in the European functionalist, communicative tradition of the early 1970s, to which reference will necessarily have to be made by all tests linked to it. The adequacy of the construct definition in the CEFR is the subject of recent important
critiques, which we discuss next. On the other hand, as Harsch reminds us in relation to the *Manual for Relating Exams to the CEFR* (Council of Europe, 2009),

the authors of the Manual . . . point out that different exams (claiming to have been) linked to a certain CEFR level do not necessarily have to be testing the same constructs, nor would this imply the tests are in some way equivalent (this issue, pp. 159–160).

This is presumably because of the generality and broad range of the descriptions at each level. In other words, a test of vocabulary or grammar at one level represents a test of a different construct from, say, interactional ability. But as CEFR levels are written into policy and legislation covering all sorts of aspects of life, including immigration and citizenship rights, the recognition of tests as meeting the administrative or policy requirements depends on their ability to claim “B1-ness,” “B2-ness,” and so on, and if they can do that satisfactorily, they will be recognized and used, regardless of their actual character and the relevance to the policy decision being made. This again underlines the overwhelmingly administrative and policy-oriented character of the CEFR, the very point that language testers seem least capable of engaging productively. And pace the position referred to in several articles (Harsch, Taylor), the appeals to the need for “assessment literacy” will not address the issue. The demand for assessment literacy is like asking a person to look through the “correct” end of the telescope, when what they want is not detail and close-up but a reductive and distanced overview—a number for accounting and policy implementation purposes.

In fact, of course, there is discussion of the CEFR in Europe. Harsch reports a plethora of studies on applications of the CEFR in various contexts, and issues in linking particular tests to the framework. Few of these studies are critical in any important sense. Most are overwhelmingly and unquestioningly technicist and functionalist. More interesting are studies that draw attention to weaknesses in the grounding of the CEFR in empirical research and construct thinking. As has often been pointed out, North’s (2000) study, establishing the basis of the development of the scale, used existing scale descriptors, that is, it did not involve fundamental work on the nature of language proficiency. Conceptually, the framework reflects the construct thinking of early 1970s functionalism, closely related as it is to the Council of Europe’s syllabus work attempting to define different levels of achievement at that time. Recently, however, two important lines of critique have emerged. The first is a critique from psycholinguistics, particularly by Jan Hulstijn (2007, 2011), discussed by Harsch, which raises serious questions about the construct of the scale. Hulstijn’s (2011) analysis is twofold. On one hand, he proposes a distinction between Basic and Higher Language Cognition, with L1 speakers outdoing L2 speakers in the former but not necessarily in the latter; this has implications for scales such as the CEFR, where the higher levels clearly involve Higher Language Cognition (a study of native speaker performance on the IELTS reading test some years ago pointed up a similar issue in relation to tests of English for Academic Purposes [Hamilton, Lopes, McNamara, & Sheridan, 1993]). Further, as Harsch points out, Hulstijn’s (2011) model distinguishes “a core of linguistic knowledge and the speed with which this can be processed” from “non-linguistic components at the periphery (strategic and metalinguistic competences)” (this issue, p. 155). This takes the construct in a conservative direction, a tendency already well entrenched in the English Profile Programme (http://www. englishprofile.org/), a Cambridge project based on the Cambridge Learner Corpus and related corpora, with its focus on vocabulary, grammar, and individual language functions.
The second critique is from sociolinguistics, represented by the criticism of the CEFR from those working in the field of English as a lingua franca (Jenkins, 2007; Seidlhofer, 2003, 2011). This critique has potentially more far-reaching implications, in that it disputes the equivalence of English and other languages, given the special place that English now has in global communication, and so questions one of the assumptions of the entire Council of Europe initiative since the early 1970s, that is, the functional equivalence of languages. It also draws attention to the ideology of the privilege of the native speaker in the wording of levels, which reaches its wonderful and absurd apotheosis in the following characterization of ability to hold a conversation at B2 level: “Can sustain relationships with native speakers without unintentionally amusing or irritating them or requiring them to behave other than they would with a native speaker” (Council of Europe, 2001, p. 76). Significantly, and astonishingly, there is not a single reference to English as a Lingua Franca (ELF) or the critiques of the CEFR (and language testing more generally) from an ELF perspective in any of the articles in this special issue,² apart from a single remark of Lynda Taylor’s: “Concepts such as English as a Lingua Franca (ELF) have begun to challenge the classical, dualistic L1 versus L2 paradigm, though the implications of this for language testing and assessment still need to be carefully thought through” (this issue, p. 145). Indeed.

The logical outcome of the critique of Hulstijn would be a revised CEFR. That is, its focus on the construct of the CEFR does not engage with the impact of the use of the scale given its policy-friendly character but simply suggests the necessity of a reworking of its terms. The critique from the point of view of English as a lingua franca is potentially more damaging, as it implies a separate scale for English as a lingua franca with levels worded differently from the levels for other languages, unless they were conceived in terms of lingua franca communication too, an interesting but relatively unexplored possibility. And the reconceptualization of the scales required is radical. But it seems that for most language testers in Europe, the CEFR is “too big to fail”—therein lying precisely one of the problems with it. The discussion of the CEFR in the articles in this special issue is marked by the characteristic absence among language testers in Europe (with some notable exceptions, e.g., Fulcher 2004, 2010) of the kind of critical discussion on the policy role of the CEFR represented, for example, by papers in the forum at the Language Testing Research Colloquium in Princeton in 2012 (Shohamy & McNamara, 2012). Significantly, perhaps, this latter event did not take place in Europe, and there was only a single European presenter (Barni, 2012). The lack of engagement with the larger question of the role and function of the CEFR is reflective of a general conservatism in the field of language testing. Despite the acknowledgment that has gradually been made of the necessity to engage with the political and social character of language tests, few language testers do so at any depth, and there is little discussion of the conceptual tools which would allow us to understand the issues and options more clearly.

Another theme of the anniversary conference, the testing of languages for specific purposes, produced a consensus in both papers addressing this topic (Emery, Brunfaut) that the issues confronting the field as articulated in 1980 remain the same 30 years later. A case study of the

²In fairness, Harding has elsewhere been one of the few to take seriously the challenge of ELF (Harding, 2012), and there are traces of his position in his article in this issue, where he speaks of “shuttling between different varieties” and “the development and validation of language tests which specifically assess a test-taker’s ability to deal with diverse, and potentially unfamiliar, varieties of English” (this issue, p. 194). But if this is also meant to refer to ELF communication, that reference remains implicit, and the issues specific to ELF as distinct from native varieties are not raised.
testing of aviation English is presented in useful detail by Emery to illustrate this argument. On the contrary, I would suggest that it is precisely the case of aviation English that challenges current practice in language testing for specific purposes, and language testing more generally. These issues are exemplified in a recent Ph.D. study (Kim, 2012; see also Kim & Elder, 2009) which questions the very policy of the International Civil Aviation Organization (ICAO) that dictates English language testing of non-native English speaking air traffic controllers and pilots (ICAO, 2010). Kim (2012) gained access to recordings of aviation communication from the control tower at Incheon International Airport in Seoul, Korea, and focused on a number of instances of miscommunication between the Korean air traffic controllers and foreign pilots, the latter a mix of native and non-native speakers of English. In some cases the miscommunication led to potentially dangerous situations, though fortunately they were resolved in time. She then invited focus groups of Korean air traffic controllers and pilots to interpret the transcripts of the events for her—What had led to the miscommunication? The findings showed that poor language proficiency was seldom the issue. She identified two main factors. First, there was a failure to cooperate communicatively, and in particular a failure to follow the policies of ICAO, requiring the use of set phrases and plain English; native speakers were often the source of the problem, when they insisted on their right to use wordy and idiomatic spontaneous English, when the policy specifically does not allow such use. Second, there was in some cases a failure of professional knowledge or competence. The Korean aviation personnel were highly critical of the ICAO policy. They argued that the policy had got it wrong about the causes of miscommunication. On one hand, it failed to acknowledge the role of experience and professional competence in guaranteeing safety, an issue all the more serious as it was the older, more experienced personnel who were most at risk under the policy, as their levels of English were generally somewhat lower. Moreover the policy focused exclusively on non-native speakers of English (native speakers are exempt from aviation English proficiency testing) when in fact there was a joint responsibility for successful communication. The opposition of aviation personnel in Korea in fact resulted in the effective undermining of the policy in Korea (and something similar happened in Japan). The study raises two fundamental issues for the testing of languages for specific purposes (LSP). First, the issue of the role of content knowledge in successful communication is highlighted; successful communication depends on and requires professional competence. The attempt to separate this in LSP testing will necessarily fail. This fact lends support to the argument for “strong” as against “weak” performance tests (McNamara, 1996) in LSP testing, especially as in cases such as this where the stakes are high. But as Harding points out in his discussion of communicative language testing, the models of communicative competence referred to in communicative language testing, particularly the influential model of Canale and Swain (1980) and its successors, specifically excludes what Hymes (1972) called “ability for use,” which includes nonlinguistic cognitive factors—in this case, professional knowledge and experience. Recent work on indigenous assessment criteria in ESP testing in clinical settings has suggested some ways forward here, by incorporating criteria that can be shown to matter to experienced health professionals in the criteria for assessment (Elder et al., 2013; Elder et al., 2012). Second, the argument proposed by Seidlhofer and others (see earlier) about the nature of communication in professional settings in which English is used as a lingua franca—and this is one—is clearly borne out in this study. Not all native speakers are good at ELF communication, as the study clearly shows, but current policy exempts them from any testing. Emery did not discuss the policy
underlying the tests he describes; this is symptomatic, as I have already argued, of a wider failure in language testing to engage with the policy and administrative context in which language tests are located.

Far from complacency, then, about communicative language testing now being the unremarked norm, it seems to me that communicative language testing is at a point of fundamental change. On one hand, there are the technological advances that allow the automatic scoring of oral proficiency, not by looking at the content or coherence of what is said but by looking at fine details of production, as, for example, in the Versant suite of tests, methods now incorporated in the Pearson Test of English (Academic), a rival to IELTS and TOEFL iBT. Ironically, this takes us back to Oller, who proposed a psycholinguistic model of performance testing that was displaced by the sociolinguistic model represented for example by Canale and Swain (1980). The emphasis on mastery of the linguistic system is echoed in the psycholinguistic work of Hulstijn (2007, 2011), as discussed earlier; this potentially takes us further back, to the structuralist focus on grammar and lexis which Oller (1979) had attempted to displace. The economy and efficiency of these new test methods are likely to prove irresistible to administrative authorities and to test takers alike.

On the other hand, the growing awareness of the nature of English as a lingua franca communication overturns all the givens of the communicative movement as it has developed over the last 30 or 40 years. The distinction between native and non-native speaker competence, which lies at the heart of the movement, can no longer be sustained; we need a radical reconceptualization of the construct of successful communication that does not depend on this distinction. Whether the inertia of the CEFR monolith continues to hamper any such development, as seems currently to be the case, remains to be seen. But clearly, echoing Morrow (1979), in my view, we should not be content with evolution in communicative language testing. Viva la revolución!

REFERENCES


