Commentary on Establishing a Life-Language Model of Proficiency: A New Challenge for Language Testers

Janna Fox[1]

The intriguing article, *Establishing a life-language model of proficiency: A new challenge for language testers*, raises two concerns. First, the proposed construct of a "life-language model" of proficiency in the hands of "Educational Language Testers" would require the measurement of "the psychological, emotional, and social needs of learners (through careful design of *Life-Language Tests*)" (p. 97)*.* This is a very broad definition of proficiency indeed, more suited to other kinds of tests, namely diagnostic tests. Second, the proposed construct of critical thinking, which is highlighted within the discussion of the Life-Language Test, may be more of a proxy measure for intelligence testing than critical thinking which, it is argued, is embedded within disciplinary cultures. Each of these concerns is discussed below.

1) *Purposes for testing: proficiency testing versus diagnostic testing*

The article limits its discussion of tests and testing to proficiency testing. However, the first step in testing of any sort is to carefully define the purpose of the test (and whether or not a test is even necessary). In my view, a key omission in the article's consideration of construct is its failure to acknowledge different purposes for assessment. In other words, purpose shapes what and how a test measures a construct of interest. So tests of language proficiency are essentially rather crude, external, at-a-distance measures, which by their nature are not tied to specific curricular or classroom contexts (unless the purpose of the class itself is narrowly defined as test preparation). They provide a window on whether or not a person "has a sufficient command of a language for a particular purpose" (Mousavi, 2009, p. 531). They are typically used alongside other sources of evidence to identify students who have acquired a level of language that will allow them to function in a particular context, such as academic study in university. In my view, it would be inappropriate to measure other constructs such as motivation, goal setting, affect, and so forth in a proficiency test.

However, a quality of life construct might be measured by two other types of tests. For example, diagnostic tests might be developed to represent language constructs which could also take into account such 'whole person' qualities such as motivation, goals, interests, and abilities. At present diagnostic tests such as the Diagnostic English Language Needs Assessment (DELNA) (see *www.delna.auckland.ac.nz/*) are increasingly being used by English-medium

---

1 Carleton University, Ottawa, ON, Canada.
E-mail: janna_fox@carleton.ca

universities. Such tests are typically administered to all new undergraduate students (regardless of their language background) after they have been admitted to their first-year of study.

DELNA is specifically designed to measure language needs, but in some cases, institutions have added their own components to the DELNA language battery in order to measure other knowledge, skills or capabilities. For example, at my own university, where DELNA is administered to new undergraduate engineering students, we added a questionnaire which elicited information on background and experience in engineering, along with a mathematics diagnostic test. It would, of course, be possible to add measures of other quality of life constructs, such as motivation, goal setting, social adaptability, etc., by adding to the diagnostic test -- provided a justification could be made for such measures. If such a measurement were used, what would be the outcome? For example, would a counsellor meet with students who were potentially unmotivated – based on test results? Given the complexity of motivation, how could we use such information to support student learning? How would it help a student to know that he or she tested at the bottom of a motivation scale? What would be the impact of such testing? Perhaps in the contexts of diagnostic assessment such information might be used to scaffold a student's transition to university. In the context of proficiency testing, however, the un- or under-motivated might well be prevented from entering the university.

Diagnostic tests such as DELNA typically link score outcomes to specific accelerated programs or specific academic support. In other words, they are used specifically as assessment *for* learning measures. Such tests have little or no washback on classroom teaching; they diagnose strengths (for accelerated learning) and identify potential gaps in knowledge, skill or capability. They tend to have less content or face validity than other types of tests because they measure narrowly, in precise and arguably more scientific ways, and teachers could not or would not generally be inclined to appropriate them for use in their own classrooms. Results could be used by teachers to inform their teaching, but this gives rise to other issues. (For a discussion of differences in teachers' use of diagnostic assessment results within a university-level English for Academic Purposes (EAP) program see Fox, 2009).

The second type of test that comes to mind as a quality of life measure is the end of term or end of unit summative assessment. These are achievement tests which teachers, alone or in groups, are often responsible for developing. They are typically content specific and directly represent or operationalize what has been taught (and hopefully learned) in a unit, course, or term of study. Although the emphasis of such tests is assessment *of* learning (i.e., summative assessment), they can and do directly support student learning when they are linked to assessment *for* learning approaches. For example, portfolio approaches allow students to develop collections of their work to demonstrate and document their learning. Such approaches encourage the development of reflection (Fox & Hartwick, 2011; Hirvela, 1997), becoming a repository of artifacts for future reference. By virtue of being unique, individual collections of work, they also provide a place that officially recognizes the *whole person*, and could easily house quality of life measures (or be viewed as one). Such approaches require that teachers have increased assessment literacy, i.e., knowledge, skills and practices "in both normative and classroom assessment" (Fulcher, 2012, p. 126). It goes without saying that assessment literacy is a critical requirement for both effective teaching and for testing of quality of life constructs.

In sum, in my view, the issues raised by the article are not so much ones of construct definition, but rather issues of purposes for testing. Proficiency tests are global measures which

take a wide sweep, at-a-distance view of what we know and can do with a language. They cannot and do not operationalize incremental language development (of relevance in a classroom) or personal attributes. This is why a determined and hard-working student can study in an intensive 25-hour a week language class for three months, take a proficiency test before and after the course, and test at the same level. Change and learning are no doubt there, but proficiency tests detect large criterion-referenced leaps in language (e.g., a move from band 5 to 6 on IELTS). Proficiency tests are not designed to detect all the small incremental steps (or all of the personal effort a student put forth) in attempting to increase proficiency. Proficiency tests are not appropriate vehicles for so-called quality of life measures, whereas diagnostic tests may be.

2) Critical thinking as a construct

Hopefully teachers will not be inclined to use Scholastic Aptitude Test (SAT) or Graduate Record Exam (GRE) items as examples of items that could be used in a classroom to encourage the development of critical thinking (although they might engender a discussion of vocabulary meanings and relationships).   Critical reading and critical thinking are discipline and context specific. What is valued as critical thinking within one discipline is often not of value to another. Much undergraduate study is devoted to the development of thinking that is consistent with disciplinary perspectives. Whether thinking like an economist or thinking like a sociologist, the thinking is not the same. Only within English for Specific Purposes (ESP) contexts would it be possible to attempt to define a construct which arguably represented critical thinking as embedded within disciplinary language use.   Again, diagnostic tests might be useful in identifying students' understandings, skills, and capabilities for doing engineering, or chemistry, or philosophy. There are some examples at the classroom level of diagnostic assessment approaches that inform teaching within the context of a university discipline. For example, Artemeva and Fox (2010) describe the use of a diagnostic testing approach in the context of a first-year undergraduate engineering communications course.

In sum, if there are issues and challenges to construct definition in the field of proficiency testing, they may relate more to increasing recognitions of the multimodality of written and spoken communication; to recognitions that meaning is made through more than words. Rather, language use occurs in situated contexts which give rise to meanings that are realized not only through linguistic modes (spoken or written), but also through images, sound, gesture, gaze, expression, movement, and the use of space. The multimodality of language may be more relevant to issues of construct representation in proficiency testing than the life language constructs considered in the article.

References

Artemeva, N. & Fox, J. (2010). Awareness vs. production: Probing students' antecedent genre knowledge. *Journal of Business and Technical Communication*, *24*(4), 476-515.

Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 8, 26-42.

Fox, J. & Hartwick, P. (2011). Taking a diagnostic turn: Reinventing the portfolio in EAP

classrooms. In D. Tsagari & I. Csepes (Eds.), pp. 47-62). *Classroom-based language assessment.* Frankfurt: Peter Lang.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), pp. 113-132.

Hirvela, A. (1997). Disciplinary portfolios and EAP writing instruction. In L. Black, D. L. Daiker., J. Sommers, G. Stygall (Eds.), New directions in portfolio assessment (pp. 140-156). Portsmouth, NH: Boynton Cook.

Mousavi, S. A. (2009). *An encyclopedic dictionary of language testing*. Teheran, IR: Rahnama Press.