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

Vahid Aryadoust[1], Priscillia Pui[2]

1. Introduction

In their article published in the second volume of the Iranian Journal of Language Assessment, Professors Pishghadam and Zabihi make a praiseworthy attempt to stress the importance of *uses* of language tests in real life situations. They explain the "life-language model of proficiency" and its application to language assessment and pedagogy. Pishghadam and Zabihi's examination of literature is fairly thorough, although several important theories which would be highly relevant are missing. In this commentary, we will focus primarily on the assessment and validation issues observed in the article, as other issues have been thoroughly discussed by other commentators in this volume. We will further show that the concerns raised by Pishghadam and Zabihi are partly reminiscent of the notions put forward by the Common European Framework of Reference and can partly be traced back to the 1980's. As a final point, we will discuss the shortcomings of their framework and validity argument in light of the current language assessment literature (Fox, 2005; Fulcher & Davidson, 2009; Bachman & Palmer, 2010).

2. Theoretical Underpinnings

To begin with, it is important to note that the term "applied ELT" in Pishghadam and Zabihi's article published in the current volume stands for applied English language *testing*, whereas in their previous works the authors take ELT as an acronym for English language *teaching*. Unfortunately, this lack of clarity would cause a great deal of confusion. The authors mix up the two acronyms in their discussion of the validity argument. We will further discuss this concern below.

Pishghadam and Zabihi state their main argument upfront: "we argue that language testers are expected to become *Educational Language Testers* who take into account not only the essential language elements but also those of other disciplines which are the relevant and vital aspects of learners' lives" (p. 2). The use of the term *Educational Language Testers* may be new but the whole concept echoes the basis for the Common European Framework of Reference (CEFR) which aims to provide "the means for educational administrators, course designers,

---

1 Centre for English Language Communication, National University of Singapore, Singapore, Email: *elcsva@nus.edu.sg*

2 Centre for English Language Communication, National University of Singapore, Singapore, Email: *elcpple@nus.edu.sg*

teachers, teacher trainers, examining bodies, etc., to reflect on their current practice, with a view to situating and co-ordinating their efforts and to ensuring that they meet the real needs of the learners for whom they are responsible (Council of Europe, 2011, p. 1). Pishghadam and Zabihi's argument for language assessors "to view language testing as a more interdisciplinary field" is again not foreign to CEFR that has similar aspirations "to promote methods of modern language teaching which will strengthen independence of thought, judgement and action, combined with social skills and responsibility" (Council of Europe CEFR, 2011, p. 4). Based on this premise, Pishghadam and Zabihi advocate the concept of "English for Life Purposes (ELP)" which comprises several types of "life skills" including: "motivation to learn, emotional intelligence, critical thinking ability or creativity, learners' anxiety, neuroticism, and depression or burnout" (p. 6).

Accordingly, language tests, as they argue, should accommodate (or engage) two dimensions: a) language elements, and b) test takers' attributes, the latter being life skills such as "creativity, critical thinking, emotional intelligence, anxiety, and so on" (p. 5). Unfortunately, Pishghadam and Zabihi do not further elaborate the relationship between the two dimensions in assessments and their reasons for integrating them. Furthermore, their framework's functionality and feasibility in real-life situations remain obscure; instead, they provide one very general example of measuring learners' use of target language and critical thinking on page eight and frequently refer readers to previous works, such as Pishghadam, Zabihi, and Norouz Kermanshahi (2012), leaving the reader to speculate on how to "measure the life issues" (p. 6) through language tests. Other relevant issues that might raise questions for a reader would be:

- What "issues from other disciplines" (p. 7) should we consider in assessments? To what disciplines are the authors referring?
- Would these "issues" not make our assessments convoluted, thereby inflicting construct-irrelevant variance on language tests?
- What "useful aspects of learners' lives" (p. 7) are we advised to incorporate into our tests? Are learners' anxiety, neuroticism, and critical thinking all "useful aspects" of life?

The relationship between test performance and test takers' demographic and psychological factors is not a new research line in language and educational assessment. Embretson (1983) in her response to Cronbach and Meehl (1955) and Messick's "unitary concept" of validity (see Messick, 1989) proposed a conceptualization of construct validity known as "nomothetic span." Her model embraces both construct representation—"identifying the theoretical mechanisms that underlie test performance" (Embretson 1983, p. 180)—and the "network of relationships" (Embretson 1983, p. 180) between a test takers' performance and *external* factors, including demographic traits and environmental circumstances (Embretson & Gorin, 2001). More recently, the environmental focus of Embretson's model led Wolfe and Smith (2007) to describe it as an "external model," and Cronbach and Meehl's model as an "internal model."

Another recommendation for Pishghadam and Zabihi is to elaborate the differences between their model and nomothetic span; articulate major validation requirements in their model; and provide a plan to investigate these requirements. More importantly, they could have discussed the dimensionality of their proposed "abilities", their *weightage* in measuring the "life-language" construct, as well as a practical model to be delineated operationally (see Fulcher & Davidson, 2009). From the current definitions they provide, we gather that tests should tap both

external variables (for example, critical thinking) and perhaps avoid inducing anxiety; however, these concerns are already well-established and heatedly debated in language and educational assessment (for example, Bachman, 1990; Bachman & Palmer, 2010; Fox, 2005; Kunnan, 2004).

Finally, it would have been useful if the authors had provided more information about the life skills that they attempt to measure in addition to language ability. The only life skill example that the authors provide in the whole article with a sample test prompt pertains to critical thinking. The argument that the prompt would "expose" test-takers' "creativity and critical thinking abilities through the medium of a second/foreign language" or "target language" (p.8) appears too simplistic and fails to take into consideration the conditions and constraints of the test-taker: a test taker who is a creative and critical thinker but lacks proficiency and vocabulary in L2 might fail miserably at the task. Since the authors' emphasis is on life skills, and one also wonders why the prompt does not focus on an authentic task but "a seemingly improbable event" (p.8).

There are many other important issues surrounding the test taker in an authentic communicative situation that the authors have not accounted for such as the kinds of target domains, tasks, conditions, themes, and interlocutors. A reader who is familiar with the CEFR might presume that Pishghadam and Zabihi are attempting to present a similar kind of general *framework* which connects language learning, teaching and assessment. The use of the term "framework" by the authors (p.7) in the following statement: "(t)his paper, therefore, seeks to expand previous models of language proficiency by offering a new framework for language use in life" seems to hint at this, although the term is later changed to "model" in other parts of the article. Perhaps, Pishghadam and Zabihi's work would also benefit from clear distinctions between the terms "model" and "framework", which are delineated by Fulcher and Davidson (2009).

3. Validation of the "Life-Language Tests"

To provide validity evidence for their model (or framework), Pishghadam and Zabihi have attempted to develop an interpretive argument (AI) (see, for example, Kane, 2012) or perhaps assessment use argument (AUA) (see Bachman, 2005; Bachman & Palmer, 2010) based on the informal reasoning / argument framework proposed by Toulmin (1958/2003). Toulmin's framework advocates that claims should be clearly articulated with relevant evidence for support. The argument also ought to be robust enough to meet the challenges of claimants, and attenuate or refute them.

There are two major issues surrounding the validity argument proposed by Pishghadam and Zabihi. First, the logic of the authors' argument is murky at best and lacks focus, as they seem to make (at least) two claims which address English language *teaching* as opposed to *testing*; that is, although the authors attempt to support English language testing (ELT), they repeatedly make reference to their previous works on English language teaching (ELT). For example, they start out by stating that "ELT has already gained an independent status and, therefore, it should not be considered a part of linguistics anymore" (p. 9). This claim would be irrelevant to "life-language *tests*" and can only be associated with language teaching since the authors rely heavily on their previous works on language teaching for support. The claim is further convoluted by a conclusion which is stated immediately: "Accordingly, the consequences of using life-language tests and of the decisions that are made based on them will be beneficial for the society in general, and for language learners and language teachers, in particular" (p.9).

Second, the authors purport that "the interpretations about language ability are claimed to be meaningful with respect to the course syllabus" (p. 9). "[C]ourse syllabus" is indeed a vague or loose use of terminology and leaves the reader wondering about the nature of the courses and syllabi mentioned by the authors.

The proposed supporting evidence includes "a few studies [that] have been done to examine the potentiality of ELT classes to enhance learners' life skills" (for example, Pishghadam, 2008). However, as earlier noted, these works seem to be irrelevant to language assessment as they address pedagogy. The loose connection between supports and claims would contribute to a faltering or weak argument for the development of their life-language test. Under this heading, another claim is made and left unsupported:

> …it seems that incorporating language learners' characteristics such as self-esteem, motivation, critical thinking, and emotional intelligence into the process of test design can strengthen the measurement procedure. (p. 10)

In addition, the authors' treatment of validity and validation is insufficient and it would be useful to reexamine Kane's (2002, 2004, 2006) works for a comprehensive description. The authors, for instance, could have discussed the differences in the terminologies employed by Kane (2002, 2012) and Bachman and Palmer (2010) as well as the process to develop validity arguments. Kane (2012, p. 4) stated that

> The argument-based framework is quite simple and involves two steps. First, specify the proposed interpretations and uses of the scores in some detail. Second, evaluate the overall plausibility of the proposed interpretations and uses.

The former stage includes developing an interpretive argument (IA) which is supported in the latter stage by validity arguments (VA).

We disagree with Pishghadam and Zabihi's view that Bachman and Palmer's (2010) validation framework is a "revised view of validation". Indeed, Bachman and Palmer's model is founded upon Messick's (1988, 1989), Kane, Crooks, and Cohen (1999), Kane's (2002), and Mislevy's (2004) models (see, for example, Bachman, 2003, 2005); however, the commonly-endorsed principles of validation, i.e., uses and interpretations of scores, as well as the inference levels proposed by Kane are all built into Bachman and Palmer's model (Bachman, 2011). Finally, warrants and backings of their proposed IA are not thoroughly discussed, leaving the readers wondering about the warrants and backing the authors are proposing and how these two concepts are different from each other.

4. Closing Remarks

We find Professors Pishghadam and Zabihi's work interesting and thought-provocative. It would appear that authors believe English language testing (ELT) has a narrow focus and their proposed model would help the language testing field evolve and take on a different dimension. Although this view would seem to have the potential to develop into a new model / framework of language assessment, it has not been grounded within a solid and well-articulated framework. We believe that their research would benefit from a more robust underlying structure, clearly stated definitions, and validated operationalization. The authors could also further strengthen their work by providing evidence to support their claims; an established methodology to validate

tests which are devised based on their model; and an articulation of the advantages that their model offers over conventional models in language and educational assessment.

References

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

Bachman, L. F. (2003). Constructing an assessment use argument and supporting claims about test taker-assessment task interactions in evidence-cantered assessment design. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 63-65.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2*(1), 1-34.

Bachman, L. F. (2011, July). *How do different language frameworks impact language assessment practice?* Plenary speech delivered at the fourth ALTE Conference, Krakow, Poland.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Council of Europe (2011). Common European Framework of Reference for languages: Learning, teaching, assessment. Retrieved from the Website of the Council of Europe. http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*(1)*,* 179-197.

Embretson, S., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement, 38*(4), 343-368.

Fox, J. (2005). Re-thinking second language (L2) admission requirements: Problems with language-residency criteria and the need for language assessment and support. *Language Assessment Quarterly, 2*(2), 85-115.

Fulcher, G., & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing 26*(1), 123-144

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21*(1), 31-41.

Kane, M. (2004). The analysis of interpretive arguments: Some observations inspired by the comments. *Measurement*: *Interdisciplinary Research and Perspectives*, 2(3), 192-200.

Kane, M. (2006). Validation. In R. L. Brennan (ed.), *Educational measurement* (4th ed.), (pp. 17-64). USA: American Council on Education, Praeger Series on Higher Education.

Kane, M. (2012). Validating score interpretations and uses*. Language Testing*, *29*(1), 3-17.

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18*(2), 5-17.

Kunnan, A. J. (2004). Test fairness. In Milanovic, M. & Weir C., (Eds.), *European language testing in a global context: Proceedings of the ALTE Barcelona Conference* (pp. 27-48). Cambridge, UK: Cambridge University Press.

Messick, S. (1988). The once and future issues of validity: assessing the meaning and consequences of measurement. In H. Wainer, & H. Braun, (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.

Messick, S. (1989). Validity. In R. L. Linn (Ed), *Educational measurement* (pp. 13-103). New

    York: ACE and Macmillan.

Mislevy, R. J. (2004). Commentaries. *Measurement: Interdisciplinary Research & Perspective, 2*(1), 33-5.

Pishghadam, R. (2008). Afzayeshe tafakore enteghadi az tarighe mobahesye adabi. *Journal of Literature*, *48*, 153-167.

Pishghadam, R. (2011). Introducing Applied ELT as a new approach in second/foreign language studies. *Iranian EFL Journal*, *7*(2), 8-14.

Pishghadam, R., Zabihi, R., & Norouz Kermanshahi, P. (2012). Educational language teaching: A new movement beyond reflective/critical teacher. *Life Science Journal*, *9*(1), 892-899.

Toulmin, S. E. (1958/2003). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Wolfe, E. W., & Smith, Jr. E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation Activities. *Journal of Applied Measurement, 8*(2)*, 204-233.