



Unpacking the Process of L2 Test Validation for Certification Decisions

James E. Purpura
Fulbright Scholar at the University for Foreigners of Siena
and
Teachers College, Columbia University, NYC

Società Dante Alighieri, Rome
Monday, June 5, 2017, 16:00 to 18:30

A naïve assumption often made by the public is that if a language test, or any other test for that matter, has been created by a reputable institution and if their test is widely used to make decisions about examinees, then the exam offered to the public constitutes, by attribution, a valid instrument for use in decision-making. If further probed, the public would assume that the creators of the exam would have an empirical basis for making such claims, and that if needed, they would be able to provide scientific evidence to justify not only score interpretations resulting from the exam, but also the real-life decisions made on the basis of these interpretations. While no accrediting body exists to vouch for validity of language exams, as one might expect with other products in society that affect people's lives, the public generally trusts the quality of exams since any evidence to the contrary would present considerable risk not only to the reputation of the exam, but also to the brand of the institution producing the exam.

Although the public might have blind faith in the quality of assessments and the evidence supporting validity claims, language assessment researchers are only satisfied when claims made about assessment instruments for high-stakes decision-making can be justified by rational and empirical argumentation. This concern has prompted considerable research and theory in the area of language test validation. In fact, over the years, several language assessment researchers have provided comprehensive reflective reviews of the current thinking and best practices in the methods used to provide evidence to justify test score interpretations and score-based decision-making (e.g., Palmer, Groot, & Trostler, 1981; Anivan, 1991; Douglas & Chapelle, 1993; Alderson, 1991, 1994; Cumming, 1996; Bachman & Eignor, 1997; Hamp-Lyons & Lynch, 1998; Kunnan, 1998, 1999; Chapelle, 1999; Bachman, 2000; Alderson & Banerjee, 2001, 2002;

Lumley & Brown, 2005; Chalhoub-Deville & Deville, 2005, 2008; Xi, 2008; Purpura, 2010). These reviews have generally described validation methods used in published research, explained ongoing refinements to the methods along with examples of their applications, and revealed advances in both validation theory and educational measurement, with reflections of how new research methods might be applied to questions in language assessment and practice. These reviews have also discussed the types of validity evidence needed to support or refute claims of particular meanings or uses of language assessments. In this regard, as theories of validation and understandings of language constructs have evolved, so have the methods and associated technologies used to provide evidence of validity claims.

In this presentation, I provide a brief overview of the test validation frameworks that have guided validation theory and practice in language assessment. Then, drawing on Kane (2013), Bachman (2005), and Chapelle, Enright and Jamieson (2008), I demonstrate how a validation framework can be used to support (or question) multiple claims in making certification decisions.

James E. Purpura is Professor of linguistics and education in the Applied Linguistics and TESOL Program at Teachers College, Columbia University, where he teaches L2 assessment and L2 research methods. Jim's scholarly interests involve grammar assessment, the assessment of meaning, the cognitive underpinning of language test performance, learning-oriented and scenario-based language assessment, and the use of validation frameworks in L2 assessment practice. Jim is currently co-authoring a book entitled *Learning-oriented assessment in language classrooms: Using assessments to gauge and promote language learning* (Routledge). Jim is the Co-Editor-in-Chief of *Language Assessment Quarterly*, the series co-editor of both *New Perspectives on Language Assessment* (with A. J. Kunnan) (Routledge) and *Language Assessment at ETS: Innovation and Validation* (with J. Norris, S. Ross, & X. Xi) (Routledge). He was the President of the International Language Testing Association (2007-2008), and is an expert consultant for the European Association of Language Testing and Assessment. He serves on the Defense Language Testing Advisory Panel in Washington, D.C. and is currently a Fulbright Scholar at the University for Foreigners of Siena.