Chapter XII

Using Messick’s Framework to Validate Assessment Tasks in Online Environments:
A Course in Writing Effectively for UNHCR

Valerie Ruhe, University of Maine, USA
Bruno D. Zumbo, University of British Columbia, Canada

Abstract

Messick (1988) maintained that technology-based delivery methods would transform both our conceptions of teaching and learning, and also our methods for evaluating student learning. Designed 100 years ago (Crocker & Algina, 1986), classical approaches to assessing test quality are unsuited to the contemporary context of technology-based learning (Messick, 1988). We will discuss evolving conceptions of validity and show how Messick’s (1989) framework is an improvement over traditional conceptions. We will then apply Messick’s framework to the evaluation data from the globally delivered, hybrid “A Course in Writing Effectively for UNHCR.” Our results will show how Messick’s framework provides a comprehensive assessment, based on evidence, values, and consequences of the merit and worth of contemporary assessment tasks.
Introduction

In his (1988) article on validity, Messick discussed how technology-based delivery methods would transform widely held conceptions of teaching, learning, and assessment. Traditional approaches to assessing the quality of assessment tasks are ill-suited to technology-based environments, which measure knowledge, skills, and strategies in adaptive, pluralistic, and dynamic ways. Messick (1988) predicted that, in these new environments, traditional evidence-based approaches to validity and validation would give way to a comprehensive approach to validation practice based not only upon evidence, but also upon values and consequences. Although Messick’s (1989) framework was developed to validate standardized paper-and-pencil tests, there is an increasing interest in its application to distance education (Bunderson, 2003; Chapelle, Jamieson, & Hegelheimer, 2003; Ruhe, 2002b, 2002c).

This chapter will be organized in the following sections. First, we will summarize Messick’s (1988) argument that technology, by its very nature, will render a classical approach to validation untenable. Next, we will give an overview of our objectives, then we will give a brief history of validity and validation practice, and show how the emergence of technology-based assessment is forcing a rethinking of conventional conceptions of validity, which were designed for pencil-and-paper environments. After that, we will present Messick’s (1989) contemporary conception of validity, followed by an adapted version for technology-based assessment tasks, and finally we will demonstrate how the adapted Messick’s framework performs when used to guide the validation practice of an assessment task in an authentic distance education course, “A Course in Writing Effectively for UNHCR.” In this exercise, we will cycle through the adapted framework and show the various aspects of merit and worth which emerge around the assessment task of writing a UNHCR field report.

An Emerging Practice

Validity refers to the merit and worth of an assessment task, and validation practice refers to the collection and analysis of evidence to assess validity (Messick, 1989). In the classical literature of educational measurement, validity has traditionally been conceptualized as a “unitary” construct, and statistical tests were used to “measure” a single aspect of validity. For example, content validity coefficients measure the “match” of the test items to the course content, criterion-related validity coefficients relate test items to authentic tasks in the “real world,” and item bias rates are the percentage of items which reflect gender, language, or cultural bias. In addition, the focus of classical validation efforts has been on large-scale, paper-and-pencil tests, and not on small-scale assessments, that is, “classroom” tests. These classical, “unitary” approaches were designed for pencil-and-paper-based environments 100 years ago (Crocker & Algina, 1986). Although our conception of validity has gradually moved from a “unitary” towards a multifaceted conception of validity, the practice of validation still tends to be based largely upon tradition, “creating a persistent “gap between psychometric theory and research practice” (Hubley & Zumbo, 1996, p. 215). Messick (1988) maintained that