

Reconsidering Consequential Validity in Diagnostic Language Assessment

Yong-Won Lee, Seoul National University, South Korea

Diagnostic language assessment (DLA) is attracting a great amount of attention from language testing researchers and practitioners. DLA is designed to identify learners' weaknesses, as well as their strengths, in a targeted domain of communicative competence. One unique feature of DLA is that it has an explicit goal of positively impacting subsequent learning by providing the learners with diagnostic feedback and (guidance for) remedial activities. One implication of such learning-inducing characteristics of DLA for validation is that the evidence for consequential validity should be carefully collected and evaluated in support of the accuracy, meaningfulness, and effectiveness of diagnosis, feedback, and remedial learning/instruction based on assessment results.

Despite such strong needs demonstrated for careful evaluations of consequences in DLA, however, there have been on-going debates in the measurement community regarding whether it is justifiable to include consequences of testing in validity frameworks for psychological and educational tests (Borsboom, 2006; Kane, 2009; Lissitz & Samuelson, 2007; Markus & Borsboom, 2013; Messick, 1989; Popham, 2007; Sheppard, 1997). Reductionists claim that the notion of validity should be confined only to the accuracy of score-based inferences, whereas expansionists argue for inclusion of the consequences of test use and score-based actions in the validity framework. In this regard, DLA seems to provide a good testing ground for refining the rationales and methods for dealing with consequences in the validity frameworks.

With these backgrounds, the major goals of the study are to: (a) re-examine the major arguments for, and against, the inclusion of consequences in the validity framework from the perspective of DLA, (b) identify some of the major issues that need to be considered in creating validity frameworks for DLA, and (c) propose and illustrate two alternative validity (or utility argument) frameworks for DLA. In the paper, I also argue that the scope of validation and evaluation in DLA should include not only the quantitative information (or scores) but also its linkage to the qualitative information that describes the nature of attributes being measured, learners' proficiency levels on the attributes, weakness-strength patterns, and referral information for recommended remedial activities.