Validity Inquiries in Second Language Assessment

Soo Hyoung Joo
Teachers College, Columbia University

Bachman and Purpura (2008) emphasized the influence that tests can have on individuals, acting as either "door-openers or gatekeepers" (p. 456). Considering the potential positive or negative impact of tests, it is crucial to validate test scores through a rigorous process to justify their interpretations and uses. While validity has remained a fundamental concept in language testing, the way in which validity has been conceptualized has evolved throughout the years.

As an attempt to provide a historical overview of the conceptualization of validity, Pieter Lauwaert traces how the approach to validation studies has undergone significant changes. He mentions that the focus of validation studies has shifted from evaluating distinct validity components to constructing a logical argument for the use and interpretations of test scores. Based on this review, Pieter highlights that scholars should employ the argument-based approach of test validity (e.g., Kane, 2006; Chapelle, 2021), which allows for defining the important inferences required for interpreting and utilizing test scores. Scholars can gather empirical evidence for this pre-defined chain of inferences pertaining to domain description, scoring, generalization, explanation, extrapolation, and utilization. This approach allows us to substantiate the assumptions underlying each of those inferences.

Daniel Eskin focuses on one of the inferences in this chain, the scoring inference, which connects a sample of observations collected through a language assessment to an observed score. The scoring inference assumes that the language assessment elicited construct-relevant performances and that these performances have been converted into scores through an appropriate, consistent, accurate, and bias-free procedure (Purpura, 2011). Daniel conducts a Many-facet Rasch measurement analysis on a writing section of an in-house placement test to provide empirical backing for the warrants underlying the scoring inference. He investigates the main effects of examinee ability, rater severity, task difficulty, rubric scale difficulty, and scale functionality, and he investigates potential biases across different groups of L1 backgrounds regarding rater severity, task difficulty, and scales.

In conclusion, the conceptualization of validity has been refined to allow for a systematic examination of the inferences required for understanding and utilizing test scores. By employing an argument-based view of test validity, researchers can gather empirical evidence to support the assumptions underlying each inference in the validation chain. Such meticulous investigations enhance the overall credibility and utility of language assessments, ensuring that tests act as door-openers in individuals' educational and professional journeys.

REFERENCES

Bachman, L., & Purpura, J. (2008). Language assessments: Gate-keepers or door-openers? In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 456-468). Blackwell Publishing.

Chapelle, C.A. (2021). *Argument-based validation in testing and assessment*. Sage Publishing.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement*, (4th ed., pp. 17-64). American Council on Education.

Purpura, J. E. (2011). Quantitative research methods in assessment and testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning, Vol. II* (pp. 730-751). Routledge.