

# VALIDITY IN A DIFFERENT CONTEXT: EXPLORING RELATIONSHIP TO OTHER VARIABLES VALIDITY EVIDENCE

Gregory E. Stone<sup>a</sup>, Toni A. Sondergeld<sup>b</sup>, Jonathan Bostic<sup>c</sup>, & Gabriel Matney<sup>c</sup>

<sup>a</sup> MetriKs Amerique, <sup>b</sup> Drexel University, <sup>c</sup> Bowling Green State University

## PROBLEM AND BACKGROUND OF STUDY

Multiple forms of validity evidence should be reviewed to produce assessments with valid and reliable results (AERA, APA, NCME, 2014). Most mathematics validation studies do not, however, investigate beyond content and internal structure (Bostic, Krupa, Carney, & Shih, *in press*). The purpose of this study is to examine the less commonly reviewed validity evidence of "relationships to other variables" (RTOV) using mathematics problem-solving assessments (*PSM3-5*) as an example. RTOV explores how test scores may be related to other variables. When RTOV has been examined in mathematics validation studies, it was at the overall test level (see Bostic, Sondergeld, Folger, & Kruse, 2017 for an example). As such, the research question guiding our study is: What information is present when examining RTOV at both the overall test and individual item-levels?

## ANALYSIS AND RESULTS

*PSM* assessment items were hypothesized to be unrelated to the variables of gender and race/ethnicity. To test this hypothesis, student outcome measures were compared, using independent samples *t*-tests. Item level differences were then evaluated using Rasch differential functioning (DIF) analysis. While findings revealed few *test-level PSM* differences by gender or race/ethnicity, *specific item level differences* were found across all *PSMs*.

## DISCUSSION AND IMPLICATIONS

When *PSM* assessments demonstrated test-level differences in scores by gender and/or race/ethnicity, item level DIF better informed researchers on how to specifically modify the assessments to address biases. Without investigating RTOV in multiple ways, results from mathematics assessments risk being unintentionally biased, and may produce spurious results of problem-solvers' abilities.

## References

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bostic, J., Krupa, E., Carney, M., & Shih, J. (2019). Reflecting on the past and thinking ahead in the measurement of students' outcomes. In J. Bostic, E. Krupa, & J. Shih (Eds.), *Quantitative measures of mathematical knowledge: Researching instruments and perspectives*. New York, NY: Routledge.

Bostic, J. D., Sondergeld, T. A., Folger, T., & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement*, 18(2), 1-12.