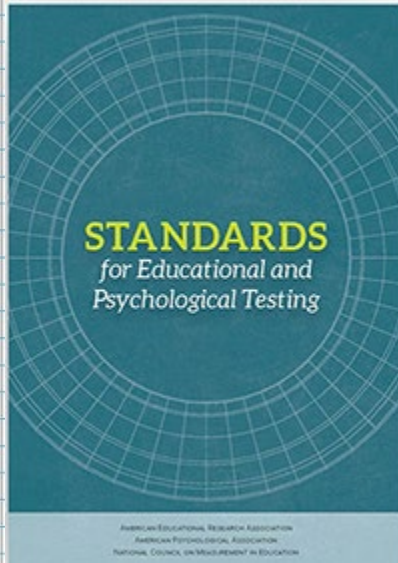


Expanding Opportunities: Facilitating Discussions to Improve Measurement Practices in Statistics Education Research

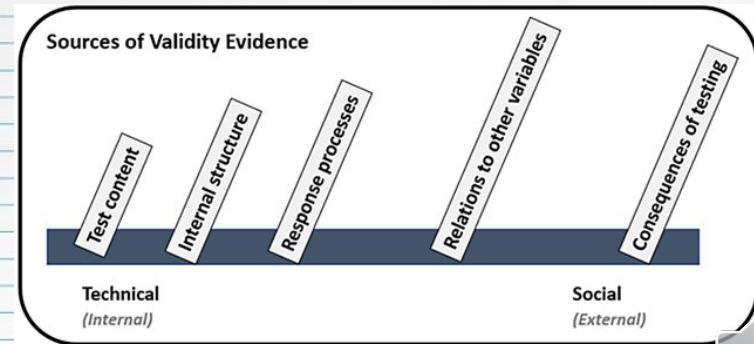
Leigh Harrell-Williams, University of Memphis
Corinne Huggins-Manley, University of Florida
Hartono Tjoe, Penn State University
Stephanie Casey, Eastern Michigan University
Christopher Engledowl, University of New Mexico
Douglas Whitaker, Mount Saint Vincent University
Charlotte Bolch, Midwestern University

Why talk about validity and validity evidence?



Validity is “the most fundamental consideration in developing tests and evaluating tests” (2014 AERA/APA/NCME Standards, p. 11)

“Validity is a unitary concept. It is the **degree** to which **all** the **accumulated evidence supports** the **intended interpretation of test scores** for the **proposed use**. Like the 1999 Standards, this edition refers to types of validity evidence, rather than distinct types of validity.” (p. 14)



Source of figure: Anunciacao & Portugal (2020)

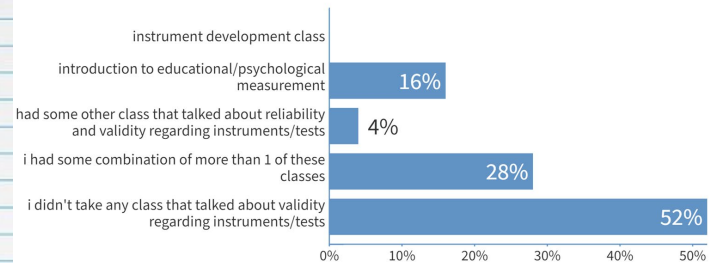
It has long been recognized that applied measurement in social science research is widely misunderstood.

For all of the advances in the measurement field, measurement theory is not regularly or appropriately incorporated into such research (i.e., Flake & Fried, 2020).

Additionally, measurement training remains de-emphasized in graduate program curricula (Aiken et al., 2008; Childs & Eyde, 2002).

Illustrating this point, Harrell-Williams and Whitaker recording the following participant responses during their 2019 USCOTS breakout session on validity evidence.

Did you ever take a class that talked about reliability and validity regarding instruments/tests?



When asked to provide an answer to the open-ended question, “What do you think of when you hear the word “validity evidence”, only 3 (27%) of 11 responses would be coded as appropriate responses.

Validity Evidence for Measurement in Mathematics Education (V-M²Ed)

PIs: Erin Krupa (NC State Univ.) & Jonathan Bostic (Bowling Green State Univ.)

Project Goals:

- Identify measures used in mathematics and statistics education.
- Compile validity evidence for these measures using a 2014 Standards-influenced framework.
- Create searchable repository for measures and validity evidence.

Synthesis Groups:

Elementary (K-6) Math

Secondary (7-12) Math

Undergrad/Grad Math

Teacher Education

Instruments

Teacher Education Tests

Statistics Education K-20



Statistics Education Synthesis Group

Completed Work (Identify Instruments):

- Database searches to find articles in Stat Ed journals and ICOTS proceedings papers to initially identify instruments since 2000
- Focused search for articles citing those instruments
- Classification of articles as containing validity evidence and usage of instrument (or not)

Group Members:

Charlotte Bolch, Stephanie Casey,
Christopher Engledowl,
Leigh Harrell-Williams, Taylor Mulé,
Justine Pointek, Hartono Tjoe,
Douglas Whitaker

Search Results: 111 instruments (50 student attitudes/beliefs; 45 student knowledge; 16 teacher-focused)

Ongoing Work for Repository: Apply validity evidence framework to classify validity evidence found in articles for these 111 instruments



Encouraging Best Practices in Measurement in Stat Ed Research

Bandalos (2018), *Measurement Theory & Applications for the Social Sciences*
“...This is why **test validation is best thought of as a program of research** in which one attempts to obtain a **body of evidence** that, taken as a whole, would **support the intended uses of and inferences from the test scores.**”

Flake & Fried (2020), *Measurement Schmeasurement: Questionable Measurement Practices*

pg. 458: “We define **questionable measurement practices (QMPs)** as decisions that researchers make that raise doubts about the validity of measures used in a study and ultimately the final conclusion.”

pg. 460: “There are **two levels** to consider: decisions made by **individual researchers** conducting individual studies and **systematic questionable practices** that appear again and again in the literature.”



Questionable measurement practices discovered in our work:

- Use of instruments for populations for which no validity evidence exists
- Combining some items from multiple instruments without providing validity evidence for “new” instrument
- Only providing a measure of internal consistency and factor analysis

Things to remember:

- Validity is not a property of the test itself. (It is a property of scores for a specific use.)
- Most articles you read use QMPs rather than best practices in measurement theory.

What you can do to encourage best practices...

- Take a measurement course and/or do a little reading
 - AERA/APA/NCME (2014) Standards
 - Bandalos (2018), Chapter 11
 - Flake & Fried (2020)
- Whether you are using a new or existing instrument, include gathering of validity evidence in your research plan
- Be explicit about why you chose a specific instrument (i.e. match to construct of interest, population of interest) or had to create one



USCOTS (and beyond) TO-DO LIST:

1) Complete Our Survey During USCOTS 2021

Use the survey link or the QR code to share your measurement-related experiences with us...

Link:

https://pennstate.qualtrics.com/jfe/form/SV_4HCvIRgoizkqayi



QR code:

2) Drop by Our Live USCOTS 2021 Poster Session to Chat with Us & View Survey Results

Wed. June 30th, 2:45-3:30 EST
(Poster & Beyond #2)

3) Look for us at the IASE 2021 Satellite Conference, 8/30 – 9/4

4) Keep Up with Us:

<https://www.researchgate.net/project/Statistics-Education-Synthesis-Group-Validity-Evidence-for-Measurement-in-Mathematics-Education-Project>



References

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, 63(1), 32–50. <https://doi.org/10.1037/0003-066X.63.1.32>

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. <https://www.testingstandards.net/open-access-files.html>

Anuniação, L., & Portugal, A. C. (2020). A Case Study on Strengthening the Link Between Psychometrics, Assessment, and Intervention in Autism Spectrum Disorder (ASD). In A. Singh, M. Viner, & C. J. Yeh (Eds.), *Special Education Design and Development Tools for School Rehabilitation Professionals*: IGI Global. <https://doi.org/10.4018/978-1-7998-1431-3>

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Press.

Childs, R. A., & Eyde, L. D. (2002). Assessment Training in Clinical Psychology Doctoral Programs: What Should We Teach? What Do We Teach? *Journal of Personality Assessment*, 78(1), 130–144. https://doi.org/10.1207/S15327752JPA7801_08

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465. <https://doi.org/10.1177/2515245920952393>

Harrell-Williams, L.M. & Whitaker, D. (2019, May). *Evaluating validity evidence for instruments in statistics education*. Breakout session for the United States Conference on Teaching Statistics (USCOTS) 2019, State College, PA.

