Validity of Assessment: Conceptual Framework and Evidence in Simulation Education

Presented by Kuan Xing, PhD, Chelsea Renfro, PharmD, CHSE, Teresa Britt, MSN, RN, CHSE, & Chad Epps†, MD, CHSE, FSSH

University of Tennessee Health Science Center

†Dr. Chad Epps passed away unexpectedly in December 2020. We dedicate this work to him.

Financial disclosure: We have no disclosures to report.

Welcome

- Our team members: Kuan Xing, Chelsea Renfro, and Teresa Britt

OBJECTIVES

Course objectives:

- Understand the need for research on validity and various validity evidence from the Messick's unified validity framework;
- Identify the key validity components from various simulation scenarios and interpret the results from validity studies; learn from the real-world simulation example;
- Discuss lessons learned from implementing the validity framework in simulation projects and forging collaboration opportunities.

INTRODUCTION

- Validity is an argument

Can I defend the use of the **scores** from this assessment → To make a **decision** for a given **purpose**?

For instance, you are going to a court and will defend an appropriate use of one specific assessment tool – That is validity!

Definition

- Is it reasonable to take action, for a particular purpose, based on the results of this assessment?
- Validity refers to “the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of the test” *Standards* (AERA, APA, &NCME 2014)
- Reliability is a necessary, but insufficient condition for validity

- Messick's Unified Validity Framework

Validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment.

- Messick, 1989

Samuel J. Messick III
(1931-1998)

Unified Validity Framework

All validity is construct validity

Five Sources of Validity Evidence

1. Content
2. Response process
3. Internal structure
4. Relations to other variables
5. Consequences of testing

Evidence Based on Content

Assessment content: The themes, wording, and format of the items, tasks, or questions (2014 *Standards*)

- Match between content sampled and domain blueprint

	Domain #1	Domain #2	Domain #3
Task 1	x	x	
Task 2		x	x
Task 3	x		x

- Item quality review: content, bias,..
- Independent expert review: e.g., Delphi method

Evidence Based on Response Process

Response process: Cognitive process engaged by the learners/raters etc.

- Did examinees/learners understand questions?
- Were the responses recorded properly by examinees?
- Were the responses scored properly by the rater (e.g., SP)? – Interrater reliability

Evidence Based on Internal structure

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

- Degree to which individual items fit the underlying construct of interest

- Related to the statistical or psychometric characteristics of the instrument

  - ✓Reliability of scores
  - ✓Factors

Relations to other variables

- Convergent validity
  - Results correlate with other measures of the same construct
  - Results correlate with measures of related constructs

- Divergent validity
  - Results do not correlate with measures of unrelated constructs

Consequences of Assessment

- Impact on learners
  - Pass/fail rates
  - Different passing rates for different groups
- Impact on curriculum
  - Learning from and for assessment
  - Teaching
- Impact on community/society
  - False-positive/False-negative impact
  - Fairness

Questions: Sources of Validity Evidence (1)

- A test developer met with content experts and created a test specification (blueprint) that proportionally matched the number of questions in an assessment.

  CONTENT

- Test-retest reliability of a multiple-choice test was .65.

  INTERNAL STRUCTURE

- Experts and novices were administered the same assessment. Experts scored 20% points higher than novices.

  RELATIONS TO OTHER VARIABLES

- Checklist items from an OSCE station were reviewed by content experts. 30% of the items were determined to be irrelevant to the curriculum and therefore removed from the assessment.

  CONTENT

Questions: Sources of Validity Evidence (2)

- The correlation between the end-of-course pathology exam and the national licensing examination was .70.

RELATIONS TO OTHER VARIABLES

- Faculty reviewed and revised multiple-choice items on an assessment, checking for flaws, including grammatical errors and unclear expressions.

CONTENT

- Standardized patients scoring the patient encounter were trained using standardized videos for calibration.

RESPONSE PROCESS

- The pass rates for a licensure examination increased by 20% after students were given extensive remediation, following their performance on a graduation competency examination.

CONSEQUENCES

Application examples: #1

- Example 1: I want to apply an assessment tool in my simulation course/event. What validity evidence should/could I collect?
  - New or existing one?
  - Does it involve raters? Response process
  - How about the construct which items/checklists/response sheets assess: Internal structure validity
  - Formative vs. summative? Content/Consequences

Application examples: #2

- Example 2: I want to conduct validity study on an assessment system in my simulation curriculum for research. Where to start?
  - Start documenting everything
  - High stakes vs. low stakes? The higher the stakes, the greater the requirement for collecting relevant validity evidence
  - Instrument/tool (internal) structure *
  - Examinees/learners and raters: response process
  - New system? Any comparison? Relations to other var.
  - Predictive? Consequences (e.g., sim vs. real-world)

Section Summary

- "Validity" is contextual, relative, does not eliminate measurement error
- 5 sources (ALL construct): content, response process, internal structure, relations to other variables, consequences of assessment
- May include multiple indices for one aspect of evidence resources; may require routine evaluation/re-evaluation
- Validity is an argument; instead of calling it "tool validation", you justify your interpretation and/or intended use of the tool

References

AERA, APA, NCME. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association;2014. 230 p.

Messick S. *Educational measurement*. 3rd ed. New York: Macmillan;1989. Chapter 2, Validity; p. 13-103.

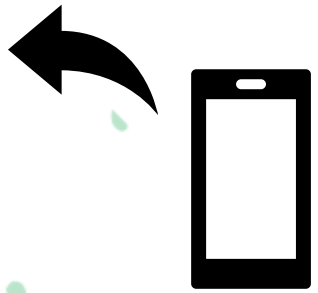
Yudkowsky R, Park YS, Downing SM. *Assessment in health professions education* (2nd ed). New York: Routledge;2020. 321 p.

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Examples for Validating Assessment Tools in Pharmacy Education

Example 1: Validity Evidence for a Vaccine Hesitancy Assessment Tool

- Program developed to train community pharmacists on how to make a presumptive recommendation to overcome vaccine hesitancy
- Needed to develop an assessment tool to guide feedback for learners



Take a picture to see more

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Quality Recommendations in Vaccine Hesitancy Assessment Tool

- Tool had six items
  - Presumptive recommendation – establishes presumptive recommendation for vaccine
  - Pharmacists' patient care process – collect and assess information pertaining to vaccine hesitancy
  - Jargon remains simple
  - Use of open-ended questions
  - Applies S.E.L.L. (sincerely encourages by loving and listening)
  - Professionalism
- Evaluated on a 5-point behavior-anchored scale

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

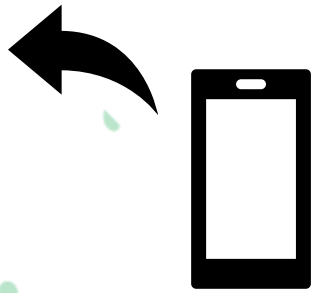
Validity Evidence

- For scenario 1, it was acceptable (.74), and was excellent in scenario 2 (.90). Overall, the interrater reliability in this study was excellent (Cicchetti, 1994).
- The lowest and highest ICC in scenario 1 were on use of open-ended questions (.51) and professionalism (.81) and in scenario 2 were on establishing presumptive recommendations for vaccination (.65) and professionalism (.86).
- The Generalizability coefficient was .42. The three largest variance components were from:
  - Person x scenario (26.1%)
  - Person x scenario x item (20.6%)
  - Person x item (8.4%)

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Example 2: Use of a Diagnostic Feedback Approach for OSCE Assessment

- Assess the validity evidence of the medication history and patient counseling rubrics used during a pharmacy OSCE
- Determine how effectively each rubric measures tasks pharmacy students should perform as part of an OSCE



Take a picture to see more

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Use of Factor Analysis to Provide Feedback

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

A 4-factor model (21 items) was obtained for the medication history checklist with 75% of variance explained.

Factor Name	Factor Name	Items Included
F1	Medication Review	Student gathered the medication name, dose, strength, route of administration, and frequency for each prescription medication.
F2	Medication Adherence	Student asks patient about medication adherence and when they took the last dose of their medications.
F3	Allergies and Adverse Drug Reactions	Student asks about any allergies and adverse drug reactions including the type of reaction and when it occurred.
F4	Medication Access, Payment, and Affordability	Student gathers information if the patient has had any problems or concerns with medication, how they pay for their medication, if they have had any issues affording medication.

Use of Factor Analysis to Provide Feedback

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

A 3-factor model (22 items) was obtained for the patient counseling rubric with 54% of variance explained.

Factor Name	Factor Name	Items Included
F1	Medication Administration Technique	Student provides education to the patient regarding the administration technique for enoxaparin injection
F2	3 Prime Questions	The student uses the 3 Prime Questions to assess current understanding of medication: 1) What did your doctor tell you this medication is for?, 2) How did your doctor tell you to take this medication?, and 3) What did your doctor tell you to expect with this medication?
F3	Medication Dosing	The student provides education to the patient regarding the medication dosing frequency, what to do if they miss a dose, storage instructions, and disposal instructions.

Lessons Learned

During Simulation Educational Design- Collaboration must be intentional, and start early in the instructional design process

- We are working to identify how to connect Faculty to our CHIPS Team for assessment design support during the early planning stages

- Transparency of communication

Simulation Research Impact

Simulation Research is important

- Expands our body of knowledge
- Catalyst to prove how simulation can cause positive change
- Exposes the power that simulation has in both healthcare education and patient safety

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Thank you!
Questions?

- Kuan Xing: kxing1@uthsc.edu
- Chelsea Renfro: crenfro@uthsc.edu
- Teresa Britt: tbritt2@uthsc.edu