Validity and Reliability Screening Tool: Step by Step Guide

Su-Ting Li, MD, MPH and Daniel C. West, MD

Identify the Construct (What area of knowledge, skill, or behavior is being measured?)

Assess Importance: Is what is being assessed important?

____ Yes (assess validity)
____ No (stop)

Assess validity: Does it measure what it is supposed to measure? (Does it hit the target?)

Step 1: Does it have *face validity*? (at first glance, does it measure what it supposed to measure?)

____ Yes (go to step 2)
____ No (stop)

Step 2: Assess *content validity* (Is there evidence that items included in assessment match the competency to be assessed)

____ Yes (go to step 3)
____ No (stop)

Step 3: Assess validity compared with a *criterion standard* ("gold-standard")? (How does it compare with a criterion standard?)

____ Yes (valid tool, go to step 5)
____ No (not a valid tool, stop)
____ No criterion standard exists (go to step 4)

Step 4: Assess validity compared with other measures of *construct and criterion validity* (in the absence of criterion standard, use these to build a case for validity)

Step 4A: Assess *construct validity*

Does it have *convergent validity*? (Is an individual's score on different measures of the same construct (e.g. skill or behavior) the same?)

____ Yes (evidence supporting validity)
____ No

Does it have *divergent validity*? (Is an individual's score on different measures of different constructs different?)

____ Yes (evidence supporting validity)
____ No

Step 4B: Does it have *concurrent validity*? (How does this measure relate to current performance? i.e. Does a PL3 score higher than a PL1?)

____ Yes (evidence supporting validity)
____ No

Step 4C: Does it have *predictive validity*? (Does this measure predict future performance?)

____ Yes (evidence supporting validity)
____ No

Step 5: Assess *external validity*. (Was the tool tested in institutions and with trainees similar to yours?)

____ Yes (potentially useful tool)
____ No (needs more study in your population)

Assess Reliability: Is the assessment tool dependable? Does it hit the same spot on the target?

Step 1: Was the *appropriate reliability assessment* performed? (if you care about different sources of error and how best to administer the assessment tool, then G-theory should be used)

____Yes (go to step 2)
____ Reliability not assessed (reliability unknown, not a useful tool)
____ Reliability assessed, but not optimally (go to step 2)

Step 2: What was the reliability or generalizability coefficient(s)?

____ 0.8 – 1.0 [good reliability, this is where it should be]
____ 0.6 – 0.8 [marginal reliability, must be better for decision making]
____ < 0.6 [poor reliability]

Feasibility: What would it take to implement this? Would it work in your program? Consider time, training, expertise, expense, data collection, equipment, etc.

____ Yes
____ Maybe (if it would require modification, beware of reliability issues)
____ No (look for something else)

The small group exercises will utilize 2 articles:

1. Brett-Fleegler MB, Vinci RJ, Weiner DL, Harris SK, Shigh MC, Kleinman ME. A Simulator-Based Tool That Assesses Pediatric Resident Resuscitation Competency. *Pediatrics* 2008: 121:e597-603, reproduced with permission from *Pediatrics*, 121:e597-e603, Copyright © 2008 by the AAP

2. Violato C, Lockyer JM, Fidler H. Assessment of Pediatricians by a Regulatory Authority. *Pediatrics* 2006: 117:796-802, reproduced with permission from *Pediatrics*, 117:796-802, Copyright © 2006 by the AAP