



HHS Public Access

Author manuscript

Int J Nurs Stud. Author manuscript; available in PMC 2022 July 25.

Published in final edited form as:

Int J Nurs Stud. 2009 September ; 46(9): 1284–1287. doi:10.1016/j.ijnurstu.2009.05.016.

A valid step in the process: A commentary on Beckstead (2009)

Allison Squires*

Center for Health Outcomes & Policy Research, University of Pennsylvania – CHOPR, 418 Curie Blvd., 3R-393, Philadelphia, PA 19104, United States

Keywords

Content validity indexing; Instrument development; Validity; Reliability

I would like to begin this commentary by reviewing the concept of validity, as defined by psychometric theory. Nunally and Bernstein (1994) address content validity in their broader chapter about validity and instrument development, the content they deem the most important to understanding the instrument development process. “Validity” itself they describe as denoting “the scientific utility of a measuring instrument, broadly statable in terms of how well it measures what it purports to measure” (p. 83). All forms of validity involve some kind of scientific generalization within the design of the instrument, making the instrument development process a key part of ensuring validity (DeVellis, 2003; Nunally and Bernstein, 1994; Furr and Bacharach, 2007). Each subsequent use of the instrument adds to the empirical evidence that it is a valid tool for measurement with different subjects; therefore, the validation process requires empirical investigations that are ongoing and constantly revisit the instrument for its applicability in different contexts and periods (DeVellis, 2003; Nunally and Bernstein, 1994; Furr and Bacharach, 2007).

Three major categories of validity exist –construct, predictive, and content – each with its own meaning, each serves as a key part of the instrument development process. Content validity, the main issue of the author purporting it is for naught, selects the best way to evaluate concepts or skills through different ways of asking the same questions to the target audience. “Expert” raters evaluate a large series of questions addressing similar content and determine the best questions for evaluating the desired construct or concept (Nunnally and Bernstein, 1994).

During the instrument development processes, it is not uncommon for researchers to develop their instrument first and test the validity later. According to the author, this is one reason why content validity is not a worthwhile step for researchers to use in the instrument development process because statistics will resolve the issue of instrument validity. Yet in a time of increasingly constrained research resources, researchers cannot afford to discover that their newly developed instrument, funded with many thousands of research dollars, is

*Tel.: +1 215 898 6485; fax: +1 215 573 2642. asq@nursing.upenn.edu.

Conflict of interest statement

The author has no conflicts of interest that may have inappropriately influenced this work.

not valid in the end. Evaluating the content validity of a survey through content validity indexing (CVI) processes improves the chances of developing a valid instrument on the first try. It is a method that is more rigorous than simple face validity.

CVI testing, the main methodological point the author took issue with in the paper and seems to misunderstand conceptually, is a *step* in the overall instrument validation process. It is a part of the overall plan for instrument development and a step that occurs before the researcher tests the instrument. CVI can save a researcher time, effort, and dollars by ensuring that the questions an instrument asks, the way the questions are asked, are appropriate for a specific population. The chance correction method developed by Polit et al. (2007) and critiqued by the author offers researchers a way to further ensure the validity of the questions they will ask their subjects. CVI testing with chance correction adds rigor to the instrument development process and facilitates the scientific generalization process that is inherent in instrument development. That being said, like any methodology, content validation has its limitations which may or may not be well addressed in published research.

Two of the methodological limitations that can arise with CVI testing relate to the concept of the “expert” rater and as the author cited, the use of the kappa statistic for evaluating inter rater reliability. With CVI testing, I have noticed that researchers rarely define “who” served as the expert raters. Often, it appears that subject area experts with high levels of education serve as the raters during the CVI testing process. Yet rater identity can affect perceptions of content validity. For example, if a researcher developed an instrument to measure patient satisfaction with nursing care, who should be the expert raters: patients or nurses? What about an instrument to evaluate the nursing work environment? Should staff nurses working in hospitals serve as the raters or doctorally prepared researchers familiar with the nuances of measuring organizational environments? If using a mix of raters to integrate a variety of perspectives, what is the right proportion of raters to ensure that the perspectives of both sets of raters are adequately represented? These questions are not adequately addressed methodologically in most published studies utilizing CVI testing to evaluate their instruments. The perceptual differences inherent in rater identity differences may affect the rating process and subsequent reliability of the use of kappa to correct for chance.

The issues the author raised about the kappa calculation appeared to have some merit. So, using my own CVI rating data, I adjusted the calculations based on the author’s recommendations. I found no significant differences in scores between 10 raters in my data, even when adjusting the order and placement of numbers. The manuscripts highlighted by the author in this study all state that the lower the number the raters, the greater the chance for error or changes in the numbers. Perhaps the main point of the author is that use of fewer raters creates a greater likelihood for an invalid result from the CVI process, even when working with chance correction. Therefore, researchers who want to use CVI should use as close to the maximum numbers of raters as possible.

In summary, I believe the author errs when stating that an evaluation of content validity is for naught in the instrument development process. Overall the author fails to demonstrate adequate understanding of the CVI testing process and its role in instrument development.

The point about the merits of using the kappa statistic was probably the most valuable contribution of the author to the debate about content validity in instrument development. That alone should serve as a cautionary reminder to researchers that maximizing the use of raters during the CVI testing process will increase the likelihood of developing a reliable and valid instrument on the first try.

Acknowledgements

Dr. Squires would like to acknowledge support for her post-doctoral fellowship from the National Institute for Nursing Research, NIH award "Advanced Training in Nursing Outcomes Research" (T32-NR-007104, Linda Aiken, PI).

References

- DeVellis RF, 2003. *Scale Development: Theory and Applications*, 2nd ed. Sage, Thousand Oaks, CA.
- Furr RM, Bacharach VR, 2007. *Psychometrics: An Introduction*. Sage, Thousand Oaks, CA.
- Nunnally JC, Bernstein IH, 1994. *Psychometric Theory*, 3rd ed. McGraw-Hill, New York.
- Polit DF, Beck CT, Owen SV, 2007. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health* 30, 459–467. [PubMed: 17654487]