The American Journal of Surgery (2016) 212, 1026-1027

The American Journal of Surgery®

Editorial Opinion

"Face validity" is not a legitimate type of validity evidence!

KEYWORDS:
Validity;
Research reporting;
Face validity;
Surgery;
Medical education

Abstract The *American Journal of Surgery* (AJS) is a long-standing leader among surgery publications and has published high-quality research dating back to 1926. To preserve AJS's reputation, it is necessary to identify issues that may affect the journal, and subsequently the broader field of surgery, in a negative way, and attempt to resolve those issues. To that end, I would like to address the issue of citing "face validity" as a type of validity evidence. In a review of AJS articles since 2006, 30 articles have consisted of researchers citing face validity as part of their validity evidence. This is problematic because validity theorists (scholars who dedicate their careers to studying validity and what constitutes acceptable types of validity evidence) have universally agreed there is no such thing as "face validity". Thus, the purpose of this article is to encourage surgeons to study recent, and authoritative, works on validity theory, so they can better articulate the trustworthiness of their research findings and be more informed consumers of research.

The *American Journal of Surgery* is a long-standing leader among surgery publications and has published high-quality research dating back to 1926. To preserve AJS's reputation, it is necessary to identify issues that may affect the journal, and subsequently the broader field of surgery, in a negative way, and attempt to resolve those issues. To that end, I would like to address the issue of citing "face validity" as a type of validity evidence. In a review of AJS articles since 2006, 30 articles have consisted of researchers citing face validity as part of their validity evidence. This is problematic because validity theorists (scholars who dedicate their careers to studying validity and what constitutes acceptable types of validity evidence) have universally agreed there is no such thing as "face validity". In fact, since 1947, validity theorists have condemned the term, acknowledging that the appearance of validity does not constitute scientific evidence.[1] In discussing why the usage of the term face validity is problematic in research, Downing[2(p8)] states "the term face validity is sloppy at best and fraudulent and misleading at worst."

So, what presumably is meant by face validity? Researchers who use the term generally intend for it to mean the findings look and feel right on the surface. For example, any observation that appears to be true would be purported to possess face validity. Although appearance is often an important characteristic, the fact remains that the appearance of something does not constitute legitimate scientific evidence. After all, we all know that looks can be deceiving. In the following, I will present some examples of articles appearing in AJS in which the authors have used face validity arguments to support their inferences. It is important to note that the purpose of illustrating these problems is not to indict or otherwise question the quality of those works; but rather to denote the nature of the usage of the term and to understand where the confusion originates.

First, a recent article by Jalink et al[3] cite no other form of evidence beyond "face validity" to support inferences about the utility of a video game designed for teaching basic laparoscopic skills. In the limitations section of the article, the authors state "Face validity is a subjective form of validation"[p1106] and cite an article by Schijven and Jakimowicz[4] in which the authors describe an antiquated model for understanding types of validity evidence. Another article by Dulan et al[5] described a robotic training curriculum and based their validity evidence on "face" and "content" validity. The authors essentially argued that because experts deemed 23 skills as relevant ie, provided content validity evidence and because experts agreed that exercises seemed to measure relevant skills, there was evidence of face validity. In reviewing where they obtained their definition of validity, the authors cite a text[6] from 1979. A third and final example

* Corresponding author. Tel.: +19195136100; fax: +19195136464.
E-mail address: kdroyal2@ncsu.edu
Manuscript received February 6, 2016

0002-9610/$ - see front matter © 2016 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amjsurg.2016.02.018

involves an article by Arain et al[7] in which the authors described an effort to construct an objective scoring system. The authors cite 2 works[8,9] as the source for describing construct validity but provide no citations indicating where they acquired the concept of face validity.

In each of these instances, there is a common thread. The information used as the primary (and perhaps authoritative) source on validity evidence came from another article appearing in a surgery journal or from a psychometric text (written by nonpsychometricians) that was more than 30-years old at the time it was accessed. In my view, this finding provides several excellent teaching and learning opportunities for us. First is the reminder that reviewing nonauthoritative works on a topic can lead to problems. Consider, for example, a surgeon who wants to learn the best way to give an effective lecture using active learning techniques. Although it would be easy and convenient to search the familiar surgery literature and learn from extant published works on the topic, one would be much better served to search the larger teaching and learning (education) literature. Clearly, there are education experts who devote their careers to study such issues, so one would be wise to defer to them. Surgeons make a major cognitive error when they assume their peers, who are no doubt incredibly bright and talented individuals, are experts at everything. Thus, it makes little sense to search familiar surgery literature for information about validity theory and evidence, a field of inquiry all its own.

Second, it is important to recognize that knowledge is cumulative; therefore, searching dated texts may also lead to problems. Just as surgeons remain vibrant and are continually working to advance their fields, so too are professionals working in the areas of validity theory, one subspecialty area under the broader measurement and psychometric domains. Most researchers subscribe to Samuel Messick's uniform conceptualization of validity,[10] but this is hardly an absolute. Numerous validity theorists have conceptualized validity in many different ways. In fact, just a couple years ago a new "Standards"[11] was produced by three leading national organizations that lead efforts to conceptualize validity, provide guidelines for attaining valid results and interpretations, and encourage appropriate use of results. Of course, the notion of face validity was nowhere to be found in these "Standards". This point underscores the need to not only locate authoritative information but also to locate recent information.

In closing, I challenge researchers in the field of surgery to study recent and authoritative works on validity theory. There are numerous articles available, but if I might recommend a few I would suggest Cizek,[12] Kane,[13] and the aforementioned Messick.[10] Understanding current conceptualizations of validity is critical for researchers in the field of surgery (and other fields of medicine). Some consider validity the very cornerstone of scientific investigation, thus surgeons must commit to understanding current conceptualizations so they can not only better articulate the trustworthiness of their research findings but also be more informed consumers of research.

Kenneth Royal, Ph.D., M.S.Ed.*
North Carolina State University
Clinical Sciences
1060 William Moore Dr., Raleigh
NC 27607, USA

References

1. Mosier CI. A critical examination of the concepts of face validity. Educ Psychol Meas 1947;7:191–205.
2. Downing S. Face validity of assessments: faith-based interpretations or evidence-based science? Med Educ 2006;40:7–8.
3. Jalink MB, Goris J, Heineman E, et al. Face validity of a wii U video game for training basic laparoscopic skills. Am J Surg 2015;209:1102–6.
4. Schijven M, Jakimowicz J. Validation of virtual reality simulators: key to the successful integration of a novel teaching technology into minimal access surgery. Minim Invasive Ther Allied Technol 2005;14:244–6.
5. Dulan G, Rege RV, Hogg DC, et al. Content and face validity of a comprehensive robotic skills training program for general surgery, urology, and gynecology. Am J Surg 2012;203:535–9.
6. Carmines EG, Zeller A. Reliability and Validity Assessment. Beverly Hills, CA: Sage; 1979.
7. Arain NA, Hogg DC, Gala RB, et al. Construct and face validity of the American College of Surgeons/Association of Program Directors in Surgery laparoscopic troubleshooting team training exercise. Am J Surg 2012;203:53–62.
8. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: from selection to certification. Curr Probl Surg 2002;39:573–659.
9. Scott DJ, Jones DB. Virtual reality training and teaching tools. In: Soper NJ, Swanstrom LL, Eubanks WS, eds. Mastery of Endoscopic and Laparoscopic Surgery. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2005. p. 146–69.
10. Messick S. Validity. In: Linn RL, ed. Educational Measurement. 3rd ed. New York, NY: Macmillan; 1989. p. 13–103.
11. American Educational Research AssociationAmerican Psychological AssociationNational Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.
12. Cizek G. Defining and distinguishing validity: interpretations of score meaning and justifications of test use. Psychol Methods 2012;17:31–43.
13. Kane MT. Validation. In: Brennan R, ed. Educational Measurement. 4th ed. Westport, CT: Praeger; 2006. p. 17–64.