

# Chapter 1

## Setting the Stage for *Validity and Validation in Social, Behavioral, and Health Sciences: Trends in Validation Practices*

Bruno D. Zumbo and Eric K.H. Chan

As witnessed in the seminal work of Messick (1989) and Kane (2006, 2013), over the last 50 years validity theories have become more expansive and complex. Prior to the 1950s, a diversity of procedures was used in validation practice and an array of names for these procedures was used when researchers reported validity evidence. Early in the history of the social and behavioral sciences, the criterion- and content-based models dominated the practice of validation (Anastasi 1986). The early practices reflected the then dominant ‘behavioral’ view in the social sciences and hence tests and measures were primarily considered predictive devices – wherein one could predict some future behavior, or was a short-hand for a more complex current behavior. With this in mind, one can see how the correlation with the criterion (i.e., the future or current behavior) was the dominant perspective in validation. Simply put, a test or measure was valid if it predicted the criterion. In 1954, the *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (the first version of the North American test standards) was published by the American Psychological Association in collaboration with the American Educational Research Association and the National Council on Measurement in Education. In this document, validity was classified into content, predictive, concurrent, and construct. A year later, Cronbach and Meehl (1955) published a seminal paper and argued that the focus should be on construct validity, emphasizing the importance of a nomological network as a form of theory building about the psychological phenomenon of interest. This signaled the change in viewing tests and measures as reflective devices (or signs) of some unobserved phenomena (i.e., one definition of a construct). This shift in emphasis to unobserved phenomena is an important landmark in the history of measurement,

---

B.D. Zumbo, Ph.D. (✉) • E.K.H. Chan

Measurement, Evaluation, and Research Methodology (MERM) Program, Department of Educational and Counseling Psychology, and Special Education, The University of British Columbia, 2125 Main Mall, Vancouver, BC V6T 1Z4, Canada  
e-mail: [bruno.zumbo@ubc.ca](mailto:bruno.zumbo@ubc.ca)

assessment, and testing. Please note, however, that the criterion view still continued but had less emphasis as the discipline of psychological theorizing began to dwell again among unobservables in response to the various forms of behaviorism that shun these unobservables.

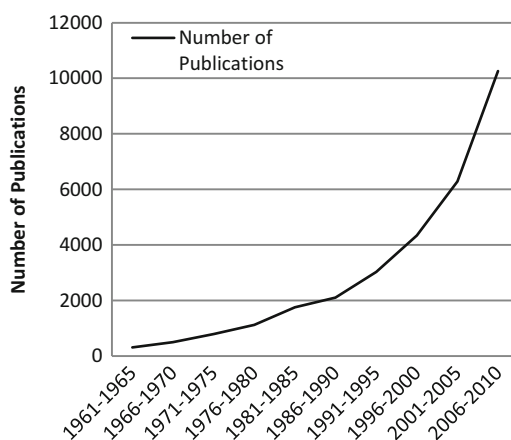
Over three decades after Cronbach and Meehl (1955), Messick (1989) published a seminal paper on the unitary view of validity. According to Messick (1989), validity is “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores” (p. 13) and is a fundamental concern in measurement. Messick’s (1989) unitary view of validity remains influential in the theoretical arena of measurement and is reflected in the *Standards for Educational and Psychological Testing* (AERA et al. 1999). According to the *Standards*, validity is “the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests” (p. 9). This perspective has given rise to the situation wherein there is no singular source of evidence sufficient to support a validity claim.

There are a series of statements about validity and validation practices that are shared and characterize a contemporary view of validity (e.g., Cronbach 1988; Hubley and Zumbo 1996, 2011, 2013; Kane 2006, 2013; Messick 1989; Zumbo 2007, 2009). Validity is not about the instrument, test, or measure but rather about the inferences, claims, or decisions that one makes based on the scores. Therefore, one does not validate a test, measure, or assessment but rather one validates the inferences. Validity does not exist as distinct types and validation should not be a piecemeal activity akin to stamp collecting – or, for that matter, collecting baseball, soccer, or hockey cards. Validation is an ongoing process in which various sources of validity evidence are accumulated and synthesized to support the construct validity of the interpretation and use of instruments. In addition to the traditional sources of evidence such as content, relations to other variables (e.g., convergent, discriminant, concurrent, and predictive validity), and internal structure (dimensionality), evidence based on consequences (intended use, and misuse), and response processes (cognitive processes during item responding or during rating) are important sources of validity evidence that should be included in validation practices. Although different validity theorists emphasize each of these to varying amounts, validation practices center around establishing a validity argument (such as Cronbach and Kane), an explanation for score variation (such as Zumbo), a theoretical framework of law-like relations that is tested against data (a nomological network, Cronbach and Meehl), sample heterogeneity and exchangeability to support inferences (Zumbo), or being guided by a progressive matrix that organizes validation practices, but centers on construct validity (Messick). As a whole, these foci capture the core perspectives on validity seen in the current literature and are meant to guide the practice of validation. It should be noted that, as expected in a vibrant scholarly discipline, elements of this contemporary view are not endorsed by all and, in fact, are challenged by some important voices in the field (e.g., Borsboom et al. 2004; Markus and Borsboom 2013).

## Trends in Validation Practices: Setting the Stage

We conducted a systematic search of validation studies published since the 1960s. Our aim was to get a snapshot of the trends in validation practices for publications that explicitly presented themselves as validation studies. Of course, a good deal of validation work is done alongside substantive studies (wherein the substantive studies are the primary objective) in psychology, education, health, and other social and behavioral sciences, however, we wished to trace the validation practices of studies for which the validation work is the primary (if not sole) purpose of the publication. We did this because we believe that focusing on studies that are explicitly cast as validation studies will give us the clearest picture of validation practices. When one is doing validation as a side project to a larger study that one considers more substantive then the validation practices will likely be described in less detail and likely also a modest or minor part of the body of work. For example, if one is interested in the mediating and moderating factors in the relation between academic self-concept and academic achievement, one may report a small-scale validation exercise along the way but certainly, by definition, that validation study will be relatively limited in scope and the details presented in the manuscript as compared to a study that has as its sole purpose the reporting of a validation study.

We were interested in documenting the general trend in publication of validation studies. For each 5-year period between 1961 and 2010 we searched the PsycInfo database for the terms ‘validity’ or ‘validation’ and the terms ‘psychometric’, ‘measurement’, ‘assessment’ or ‘test’ in the abstract of the paper. In addition, we limited our search to peer-reviewed scholarly journals. As presented in Fig. 1.1 there is clearly an increase in the number of scholarly peer-reviewed journal publications with just over 300 publications between 1961 and 1965 to over 10,200 publications between 2006 and 2010. Certainly, some of that increase can be attributed to the increase in the sheer number of journals and researchers; however, the fact is that the field of measurement validity is growing in remarkable strides.



**Fig. 1.1** Trend line depicting the pattern of publication of validation studies

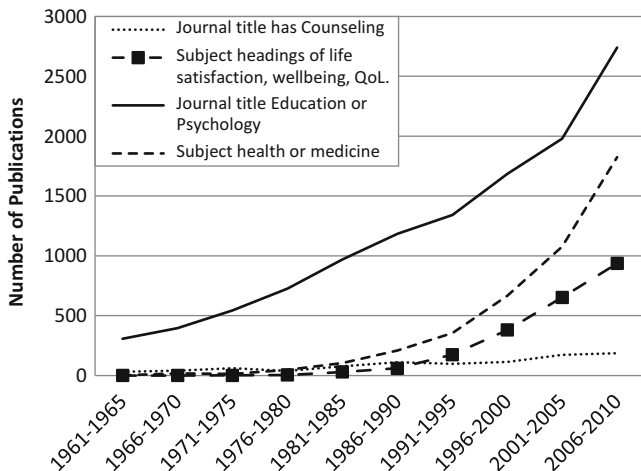
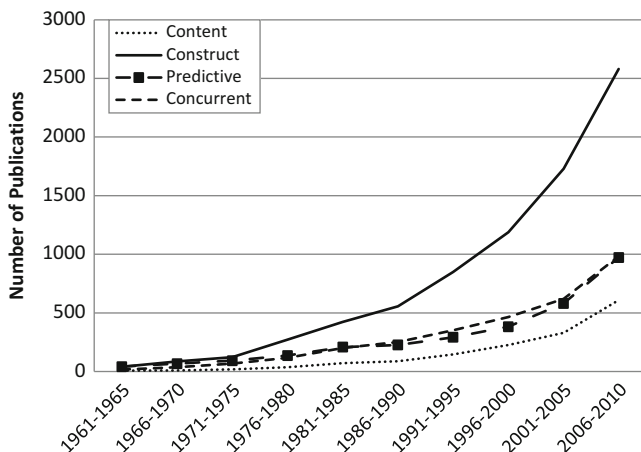


Fig. 1.2 Trend lines of publication of validation studies across disciplines

In Fig. 1.2 we documented the publication practices in four domains. Two of the trend lines represent well-established areas of measurement research that have journals dedicated to them: education or psychology, and counseling. The remaining two trend lines represent relatively emerging fields of measurement, testing, or assessment defined by terms such as ‘life satisfaction, wellbeing, or quality of life (QoL)’, and ‘health or medicine’. Again, like Fig. 1.1, we are witnessing an increase in the number of scholarly publications in these disciplines with, as expected, the greatest increase being seen in education and psychology.

Once again, in Fig. 1.3 we applied the same search strategy except that in this case we searched for various sources of validity evidence. For example, in documenting the trend in content validation studies, we searched for the terms “content validity” or “content validation” and the terms ‘psychometric’, ‘measurement’, ‘assessment’ or ‘test’ in the abstract of the papers. We continued to limit our search to peer-reviewed scholarly journals. Noting, of course, that papers can report more than one source of validity evidence, construct validity evidence is the most commonly reported followed by concurrent and predictive evidence, and finally content validity evidence.

It is important to note that in the data reported in Figs. 1.1, 1.2 and 1.3 we are looking back in time with the labels from the current *Standards*. In essence, we are looking back over our shoulders but applying today’s labels. Likewise, it is important to note that this is a “snapshot” picture that is obtained by documenting the count of words in the abstracts of the published articles and hence does not document the specifics, nor does it break it down by scholarly practices. In fact, it is this general picture that motivates the need for the studies reported in this edited volume.



**Fig. 1.3** Trend lines of publication of validation studies across sources of validity evidence

With the growing number of validation papers published in academic journals across different academic disciplines, and with the revision of the *Test Standards* scheduled to be released soon, it is timely to examine validation practices by researchers across different academic disciplines. Our focus, and the focus of this edited volume, is a study of the scholarly genre of validation reports and how this genre frames validity theory and practices.

## References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, *51*, 201–238.
- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, *37*, 1–15.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071.
- Cronbach, L. J. (1988). Five perspectives on validation argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale: Lawrence Erlbaum Associates.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. doi:[10.1037/h0040957](https://doi.org/10.1037/h0040957).
- Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology*, *123*, 207–215.
- Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research*, *103*, 219–230.

- Hubley, A. M., & Zumbo, B. D. (2013). Psychometric characteristics of assessment procedures: An overview. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology* (Vol. 1, pp. 3–19). Washington, DC: American Psychological Association Press.
- Kane, M. T. (2006). Educational measurement. In R. L. Brennan (Ed.), *Validation* (4th ed., pp. 17–64). Westport: American Council on Education/Praeger.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1–73.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York: Routledge.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.
- Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Psychometrics* (Handbook of statistics, Vol. 26, pp. 45–79). Amsterdam: Elsevier Science B.V.
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 65–82). Charlotte: IAP – Information Age Publishing.