

Some Notes on the Contemporary Views of Validity in Psychological and Educational Assessment

Estu Widodo
Universitas Muhammadiyah Malang
Estu_wid@yahoo.com

Abstract—This paper aims at providing an analysis of views of validity as an important concept in psychological and educational assessment which has undergone transformation for several decades. The revolutionary change took place along with the release of the 1999 Standards followed by the 2014 Standards in which validity is associated with interpretation and use of test scores. With the new framework, several types of theoretical and empirical evidence are required to support the validation process. The results of the analysis show that the drastic change of the validity concept and the complicated process of validation have resulted in misunderstanding, misinterpretation, and misuse. Apart from different views which challenge the new concept, the argument-based approach has provided basis for new studies in many fields. Some strategies required to deal with complicated validation processes are suggested, along with some alternative perspective associated with the limitations of the standards.

Keywords—validity, assessment, use of test score, validity evidence, interpretation, argument-based validity.

INTRODUCTION

As one of the most important aspects in psychological and educational assessment, validity concept has developed and changed several times. The marked shift happened along with the emergence of the 1999 Standards [1] followed by the 2014 standards [2]. The development of validity theorization which took several decades, starting from validity as being property of the test to the contemporary view of validity [3], albeit challenged by those who hold the instrument-based approach [4], has made a validation process very complicated. In fact, a revolutionary change took place along with the release of the 1999 Standards. Importantly, although the 2014 has also been released, many researchers still use the older or traditional views of validity for their research and publish their articles in journals for particular reasons.

TRADITIONAL AND CONTEMPORARY VIEWS OF VALIDITY

What has been known as the traditional concept of validity [5] encompasses validity as the property of instrument, the criteria of being valid or invalid, the use of commonly four types of validity, the employment of statistical analysis, and reliability as a condition for validity. Several types of validity of the traditional view have been replaced by a new view. The new concept of validity, which was already articulated as a unified concept [6] and became the culmination of the long debate in spite of its abstract nature and lack of explicit validation guidance [7], on the other hand, refers to validity as inferences of scores, validity as a continuum, construct validity as the core of validity, validity as an elaboration of theories and methods, and test consequences and uses—the traditional view excludes social consequences of test interpretation and use [8]—which contribute to the process of validation. According to the relatively new view, the validation starts from making an interpretive argument, and then building and performing an evaluation of a validity argument, using various evidence—either theoretical or and empirical [7]. Thus, the new view defines validity evidence based on various sources such as test content [9], response processes [10], internal structure [11], and testing consequences [12] to support the argument. Such types of validity evidence sources, however, have made many researchers and other standards users in doubt [3] and lead to misconceptions [14] or misinterpretations [15].

As far as the definition of validity in the 2014 Standards is concerned, in spite of its explicitness, there are problems related to clarity which lead to simplification, misunderstanding, misinterpretation, and misuse [16]. Furthermore, no specific guidance is found in the Standards [17]. A historical analysis [18] to reveal why the legal concept of validity, particularly with regard to the 1999 Standards, was ignored is good and scientific to perform. A statement dealing with the issue seems relevant, that is, "...views of validity have not penetrated all disciplines may be a reflection of a lack of impact of the modern views of validity on some disciplines..."[19].

The articulation of argument-based validity by Kane [20], [21], although it actually referred to Cronbach [22], brought about a new view of validity and triggered many important studies in various disciplines or fields such as English as a foreign language [23], [24], Medicine [25], [26], and Mathematics [9], among others. Importantly, several types of validity came to an end [27] given the unified view.

PROBLEMS IN VALIDATION PROCESSES AND PROPOSED SOLUTIONS

Pertaining to the processes of interpretation and uses of the scores which prove relatively complicated, some strategies are required. The first is performing good and accountable management in administering tests which can be used as claims. It is based on the statement [28] that a valid interpretation and use are achieved if the claim is based a strong case. Appropriate testing materials and procedures for licensure exam are cases in point. The second one is using available data for validation process and planning further improvement based on the evaluation of the previous process. It refers to an argument [28] that using many inferences and assumptions is very likely to lead to a hard validation process. Performing simple interpretation and using only supporting claims are good in the light of their intended uses. The third one is collecting the validity evidence using appropriate instrument. A study [29] reveals that learner's lower proficiency of English lead to inaccurate validity evidence in the light of construct-irrelevant variance. The fourth is utilizing available statistical analysis which yields different measures needed for building inferences in a validation process. Apart from the four strategies, alternative perspectives on any kinds standards need to be built since it is also very likely that there are many interests behind the Standards, one of which is from testing industries [30]. That is why, it is very scientific to follow debates about the intended and unintended uses and impacts of the standards either in journals, conferences, and classroom discussions. Also important to keep in mind that understanding the limitation of the standards [31] is helpful.

CONCLUSION

Validity in educational and psychological assessment is a keyword which must be understood in order that the goal of assessment, that is gathering data required for further decisions and improvement, is achieved. It is obvious from the discussion above that many of those involved in various activities of assessment have limited understanding—even misinterpretation—of validity and related concepts. Collaboration among measurement experts, lecturers, teachers, and other users to discuss the details as well as the limitations of the Standards along with the problems found during the implementation is urgent to build. Last but not least, critical thinking about the contents, purposes, uses, and limitations of the standards is absolutely required so that evaluation of any

aspects of the standard can be conducted for future improvement and development.

REFERENCES

- [1] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, "Standards for educational and psychological testing," Washington, DC: American Educational Research Association, 1999.
- [2] American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.), "Standards for educational and psychological testing," Washington, DC: American Educational Research Association, 2014.
- [3] S. Shaw & V. Crisp, "Tracing the evolution of validity in educational measurement: past issues and contemporary challenges," in *Research Matters: Issue 11*, pp. 4-19, 2011.
- [4] J. D. Hathcoat, "Validity Semantics in Educational and Psychological Assessment" in *Practical Assessment, Research & Evaluation*, Vol. 18, No. 9, pp. 1-14, 2013.
- [5] B. D. Zumbo, M. N. Gelin, & A. M. Hubley, "The Construction and Use of Psychological Tests and Measures". In *the Psychology theme of the Encyclopedia of Life Support Systems (EOLSS)*, Eolss Publishers, Oxford, pp. 1-28, 2002.
- [6] S. Messick, "Validity," in R. L. Linn (Ed.), *Educational measurement, 3rd ed.*, pp.13-103, New York: American Council on Education and Macmillan, 1989.
- [7] F. Rahimi, M. R. Esfandiari, & M. Amini, "An Overview of Studies Conducted on Washback, Impact and Validity Studies," in *Literature and Language*, Vol. 13, No. 4, 2016, pp. 6-14, DOI:10.3968/8877, 2016.
- [8] S. Messick, "Validity of Test Interpretation and Use," *Research Report. Princeton: Educational Testing Service*, 1990.
- [9] S. Sireci & M. F. Bond, "Validity evidence based on test content" in *Psicothema*, Vol. 26, No. 1, pp. 100-107, DOI:10.7334/psicothema2013.256,2014.
- [10] J. Padilla & I. Benítez, "Validity evidence based on response processes," in *Psicothema*, Vol. 26, No. 1, pp. 136-144, doi: 10.7334/psicothema2013.259, 2014.
- [11] J. Rios & C. Wells, "Validity evidence based on internal structure," in *Psicothema*, Vol. 26, No. 1, pp. 108-116, DOI: 10.7334/psicothema2013.260, 2014.
- [12] S. Lane, "Validity evidence based on testing consequences," in *Psicothema*, Vol. 26, No. 1, pp. 127-135, DOI: 10.7334/psicothema2013.258, 2014.

- [13] D. A. Cook, B. Zendejas, S. J. Hamstra, R. Hatala, & R. Brydges, "What counts as validity evidence? Examples and prevalence in a systematic review of simulation based assessment," in *Advances in Health Sciences Education*, Vol.19, No. 2, pp. 233–250, DOI: 10.1007/s10459-013-9458-4, 2013.
- [14] S. G. Sireci, "On the validity of useless tests," in *Assessment in Education: Principles, Policy & Practice*, Vol. 23, pp. 226-235, 2016.
- [15] T. Noble, A. Rosebery, C. Suarez, B. Warren & M. C. O'Connor, "Science Assessments and English Language Learners: Validity Evidence Based on Response Processes," in *Applied Measurement in Education*, Vol. 27, No. 4, pp.248-260, DOI:10.1080/08957347.2014.944309, 2014.
- [16] P. E. Newton, "Clarifying the Consensus Definition of Validity," in *Measurement Interdisciplinary Research and Perspectives*, Vol. 10, No. 1, pp. 1-29, DOI: 10.1080/15366367.2012.669666, 2012.
- [17] S. Sinharay, F. Rijmen, S. Choi, & N. J. Dorans, "The Revised Standards and Its Role in Research on Educational Measurement," in *Educational Measurement*, Vol. 33, No. 4, pp. 36-36, 2014.
- [18] P. E. Newton & S. Shaw, "Standards for Talking and Thinking About Validity," in *Psychological Methods*, Vo. 18, No. 3, pp. 301-319, DOI:10.1037/a0032969, 2013.
- [19] E. K.H. Chan, "Standards and Guidelines for Validation Practices: Development and Evaluation of Measurement Instruments," in *B.D. Zumbo and E.K.H. Chan (Eds.), Validity and Validation in Social, Behavioral, and Health Sciences, Social Indicators Research Series 54*, pp. 9-24, Springer International Publishing, Switzerland, DOI: 10.1007/978-3-319-07794-9_2, 2014.
- [20] M. T. Kane, "An argument-based approach to validation," *ACT Report Series*, Iowa City: The American College Testing Program, 1990.
- [21] M. T. Kane, "Validation", in *R. L. Brennan (Ed.), Educational measurement, 4th ed.*, Washington: American Council on Education, pp. 17–64, 2006.
- [22] L. J Cronbach, "Test Validation," in *R. Thorndike (Ed.), Educational Measurement, 2nd ed.*, Washington DC: American Council on Education, pp. 443, 1971.
- [23] T. Kumazawa, T. Shizuka, M. Mochizuki, & A. Mizumot, "Validity argument for the VELC Test score interpretations and uses," in *Language Testing in Asia*, Vol. 6, No.2, pp. 1-18, DOI: 10.1186/s40468-015-0023-3, 2016.
- [24] S. M. Alavi, S. Kaivanpanah, & A. P. Masjedlou, "Validity of the listening module of international English language testing system: multiple sources of evidence," in *Language Testing in Asia*, Vol. 8, No. 8, pp. 1-17, DOI:10.1186/s40468-018-0057-4, 2018.
- [25] D. A. Cook, A. Kuper, R. Hatala, & S. Ginsburg, "When Assessment Data Are Words: Validity Evidence for Qualitative Educational Assessments," in *Academic Medicine*, Vol. 91, No. 10, pp. 1359-1369, DOI: 10.1097/ACM.0000000000001175, 2016.
- [26] Y. S. Park, A. Hyderi, N. Heine, W. May, A. Nevins, M. Lee, G. Bordage, & R. Yudkowsky, "Validity Evidence and Scoring Guidelines for Standardized Patient Encounters and Patient Notes From a Multisite Study of Clinical Performance Examinations in Seven Medical Schools," in *Academic Medicine*, Vol. 92, Vol. 11, pp. 12–20, DOI: 10.1097/ACM.000000000000019, 2017.
- [27] M. T. Kane, "Validating Interpretive Arguments for Licensure and Certification Examinations," in *Evaluation and the Health Professions*, Vol. 17, pp. 133-159, 1994.
- [28] M. T. Kane, "Reflections on Bar Examining," in *The Bar Examiner*, November, pp. 6-20, 2009.
- [29] S. G. Sireci, Kyung T. Han & Craig S. Wells, "Methods for Evaluating the Validity of Test Scores for English Language Learners," in *Educational Assessment*, Vol.13, No. 2-3, pp. 108-131, DOI: doi.org/10.1080/10627190802394255, 2008.
- [30] B. D. Zumbo, "What Role Does, and Should, the Test Standards Play Outside of the United States of America?" in *Educational Measurement: Issues and Practice*, Vol. 33, No. 4, pp. 31-33. <https://doi.org/10.1111/emip.12052>, pp. 31-33, 2014.
- [31] A. Maul, "Justification Is Not Truth, and Testing Is Not Measurement: Understanding the Purpose and Limitations of the Standards," in *Educational Measurement: Issues and Practice*, Vol. 33, No. 4, pp. 39-41, 2014.