

Validity and Assessment

16:300:651:01

3 Credits

Instructor: Drew Gitomer	drew.gitomer@gse.rutgers.edu
Phone Number: 848-932-0642	10 Seminary Pl Rm 9
Office Hours: by appointment	Prerequisites or other limitations: none
Mode of Instruction: <input type="checkbox"/> Lecture <input checked="" type="checkbox"/> Seminar <input type="checkbox"/> Hybrid <input type="checkbox"/> Online <input type="checkbox"/> Other	Permission required: <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes Directions about where to get permission numbers: from the instructor

Learning goals

1. To understand the fundamental concepts underlying assessment, including assessment design, reliability and validity
2. To understand the evolution of conceptions of validity
3. To be able to develop, evaluate and critique validity arguments for educational assessments
4. To understand the design and use of assessments within the social and policy context of education

Course catalogue description

Assessment in education is ubiquitous and is increasingly consequential for decisions about students, teachers, administrators, and schools, including higher education. Researchers frequently use assessments to obtain measures relevant to their particular studies. Validity provides a conceptual framework for how to consider the quality of inferences drawn from assessments.

Users of assessments can get overwhelmed with all of the technical processes and information associated with tests, but these technical aspects are only tools that help us address these fundamental questions of validity. This course explores validity at a conceptual level and does not require strong quantitative skills.

The course will explore evolving conceptions of validity and focus on aspects of validity including fairness, test design and development, reliability, scoring and interpretation, and consequences of testing. We will also explore applications of validity to current assessment initiatives in the educational landscape.

In addition to the content of the course, I also want to help develop your skills in becoming a sophisticated and critical reader, interpreter and discussant of research. These are not skills easily learned and they only develop through deliberate practice, for which you'll have plenty of opportunity.

Class materials:

A set of papers from the Reading List that will be posted on eCollege or are available through web links provided in the citation.

Included in the course readings will be chapters from the Knowing What Students Know (KWSK) volume:

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

A second valuable resource is:

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

(note: a revised version of this has been in the works for years and should have been published in 2013. It may come out imminently and I would recommend getting the new edition)

Reading List

Week 1 (January 28): Course Overview – What is Validity

I will give an overview of the course and attempt to provide an organizing structure for the course that you can use to engage with the content of the course. We will explore basic ideas such as the nature of assessment and assessment scores.

*KWSK Chapters 1 and 2: Chapters can be retrieved from www.nap.edu/catalog/10019.html

Week 2 (February 4): Current Conceptions of Validity and Assessment

In this class, we review and analyze the more recent, and now commonly accepted view of validity as an argument to be made and developed with different sources of evidence. We also examine the introductory chapters from KWSK to begin to provide a sense of more current thinking about assessment.

*Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement* 50(1) 1-73.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). New York, NY: American Council on Education, Macmillan Publishing. (this was the elaborated version from which Kane (2013) was developed – this is more detailed but I think the later version is clearer).

Week 3 (February 11): Early Views of Validity

This class will focus on early conceptions of validity theory, including the role played by contemporary psychological theory and methods.

Anastasi, A. (1950). The concept of validity in the interpretation of test scores. *Educational and Psychological Measurement* 10, 67–78.

*Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. Retrieved from http://apsychoserver.psych.arizona.edu/JJBAREprints/PSYC621/Campbell,Fiske_Convergent_and_discriminant_validation_by_the_MMM_1959.pdf

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.

*Cronbach, L. J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. Retrieved from http://learnpsychology.com/courses/neuroassmt/Unit 1 History Introduction Reliability Validity/Cronbach_Construct_Validity.pdf

Week 4 (February 18): Reliability

This class will focus on the various ways of thinking about reliability and its importance in making inferences from assessments. Issues of reliability are foundational for all discussions about assessments.

Brennan, R. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, 38, 295–317. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.2001.tb01129.x/pdf>

*Carmines, E., and Zeller, R., 1979. *Reliability and Validity Assessment*. Sage Publications, Beverly Hills, California.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher*, 23, 5–12.

Traub, R. E., & Rowley, G. L. (1991). Understanding reliability. *Instructional Topics in Educational Measurement*, 10 (1), 37–45. Retrieved from <http://ncme.org/linkservid/65F3B451-1320-5CAE-6E5A1C4257CFDA23/showMeta/0/>

Week 5 (February 25): Assessment Design and Development

This class will focus on processes associated with assessment design and development and research on different assessment item formats, scoring, and measurement issues.

Lane, S., & Stone, C. (2006). Performance assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 387–431). Washington, DC: American Council on Education.

*Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 307–353). Washington, DC: American Council on Education.

KWSK Chapter 3: Chapter can be retrieved from www.nap.edu/catalog/10019.html

Week 6 (March 4): Scoring and Reporting

This class will focus on different models used for scoring and reporting information from tests. We will also focus on the validity of different kinds of claims made in score reports.

Mislevy, R. J., Wilson, M. R., Ercikan, K., & Chudowsky, N. (2001). Psychometric principles in student assessment. In D. Stufflebeam & T. Kellaghan (Eds.), *International handbook of educational evaluation*. Dordrecht, The Netherlands: Kluwer Academic Press. Retrieved from <http://www.education.umd.edu/EDMS/mislevy/papers/principles.pdf>

Holland, P. W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics*, 27, 3–17.

*Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17(2), 145–220.

KWSK Chapter 4: Chapter can be retrieved from www.nap.edu/catalog/10019.html

Week 7 (March 11): Bias

This class will focus on the meaning of test bias, how it is operationalized, and the legal and moral dimensions of test bias.

*Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Washington, DC: The American Council on Education and the National Council on Measurement in Education.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721–735.

March 18 – Spring Break No Class

Week 8 (March 25): Unitary Models of Validity

Whereas earlier conceptions of validity focused on different types of validity (construct, content, and criterion), new conceptions emerged that treat validity as a unitary concept that needs to be considered as an integrated and complex set of judgments.

*Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–100). Washington, DC: American Council on Education.

Shepard, L. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405–450. Retrieved from <http://nepc.colorado.edu/files/EvaluatingTestValidity.pdf>

Week 9 (April 1): Validity and Special Populations

In this class we investigate issues around the validity of inferences for students from a range of populations, including English language learners and students with special needs.

*Abedi, J. (2005). Issues and consequences for English language learners. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement: The 104th yearbook of the National Society for the Study of Education, Part 2* (pp. 175–198). Malden, MA: Blackwell Publishing.

*Pullin, D. (2005). When one size does not fit all: The special challenges of accountability testing for students with disabilities. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement: The 104th yearbook of the National Society for the Study of Education, Part 2* (pp. 199–222). Malden, MA: Blackwell Publishing.

Week 10 (April 8): New Directions In Assessment Design

A number of recent efforts have reconceptualized validity as developing coherent assessment designs that take into account models of student learning and performance and building assessment designs, task structures, and scoring models to model learning and performance as faithfully as possible. In this class we introduce seminal work of Mislevy, Wilson, and colleagues.

Mislevy, R. J. (2008). Validity from the perspective of model-based reasoning. In R.L. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications*. Charlotte, NC: Information Age Publishing.

*Mislevy, R. J. (2011). *Evidence-centered design for simulation-based assessment* (CRESST Report 800). Los Angeles, CA: The National Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*(2), 181–208.

KWSK Chapter 5: Chapter can be retrieved from www.nap.edu/catalog/10019.html

Week 11 (April 15): Consequences of Assessment

Messick argued strongly that assessments are not only value-laden, but they have both intended and unintended consequences that must be considered in the overall evaluative judgment of validity. In this class we review some of the critiques of high-stakes assessments in terms of the consequences they have for educational practice as well as for students and teachers.

*Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*, 1012–1027.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

*Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 85–106). New York, NY: Century Foundation.

KWSK Chapter 6: Chapter can be retrieved from www.nap.edu/catalog/10019.html

Week 12 (April 22): Post-Modern Views of Validity

Most of the validity work has evolved from psychological science that is grounded in empiricism and experimentation. Some theorists have challenged these models and brought in ideas from sociocultural and other perspectives. This class will introduce some of these ideas and implications for assessment practice and claims about validity.

*Moss, P. A., Girard, B., & Haniford, L. (2006). Validity in educational assessment. *Review of Research in Education*, 30, 109–162.

Week 13 (April 29): Validity Arguments in the Current Assessment Environment

In this class we explore how researchers are evaluating the validity of assessments in the context of ubiquitous high-stakes testing of students and teachers.

Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Developing and Selecting Assessments of Student Growth for Use in Teacher Evaluation Systems*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from http://datause.cse.ucla.edu/DOCS/shortTermGrowthMeasures_v8.pdf

*Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. Retrieved from <http://roundtheinkwell.files.wordpress.com/2011/10/a-validity-argument-approach-to-evaluation-teacher-vam-scores1.pdf>

*Marion, S. F., & Pellegrino, J. W. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice*, 25(4), 47–57. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.2006.00078.x/pdf>

KWSK Chapter 8: Chapter can be retrieved from www.nap.edu/catalog/10019.html

Week 14 (May 6): Final Project Presentations

Students will make brief presentations of their final project to the class.

Suggested Extra Reading

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Grading and Activities

Students will be evaluated based on their participation in class (20%), responses to weekly prompts (10%), and three papers of increasing complexity (70%). Each paper will require a student to identify and analyze related published research studies that address one of the central issues explored in class. The first two assignments are similar because I have found this is not something easily developed in one assignment.

Paper 1 (15%) - Assigned February 11, Due March 11

Analysis of two research studies that highlight a core validity issue in assessment. Students will be asked to summarize, interpret, and contrast the studies in light of the selected validity issue. Students may find studies independently or draw from a list of potential studies provided by the instructor.

Paper 2 (20%) - Assigned March 11, Due April 15

Analysis of two research studies that highlight a different core issue in assessment. Students will be asked to summarize, interpret, and contrast the studies in light of the selected validity issue. Students may find studies independently or draw from a list of potential studies provided by the instructor.

Paper 3 (35%) - Assigned April 15, Due May 9

Consider a large-scale assessment used in any assessment context (K–12, teaching, admissions, employment, etc.). Identify technical papers and/or research studies that attempt to establish the validity argument for the assessment. What are the central claims the assessment is intended to support, and what is the nature and quality of the validity evidence to support those claims? How could the validity argument be improved with other research and/or development efforts? This will also include a final presentation to be delivered during the last week of class.

Description of Activities

The basic structure will include both a lecture and a discussion in which a particular article or two will be explored in depth. Each week there is a set of available readings. I will expect that you read thoroughly one, or at most two of those readings. The others you can peruse or have for future reference. I will cover many of those papers in the lecture.

I expect that for the paper I ask you to read thoroughly, you will be prepared to contribute fully to a discussion. These papers are marked by an asterisk in the syllabus. I want you to make your best effort to understand the paper and to bring questions about the paper to the discussion. In fact, each week you should come to a class with a question about the paper – it could be that you had difficulty understanding

something (and you can assume if you had difficulty, you are not alone) or it could be a substantive or methodological issue that the paper raises for you.

Prior to each class, students will be asked to prepare and submit a brief response to an orienting question based on that week's reading. Students will also be asked to establish connections between the readings and contemporary assessment issues that arise from their own experiences, current policy issues, and/or other research studies in an online discussion following each class.

In addition to the readings, you will get some experiences in exploring some of the processes involved in test development and analysis that all contribute to the validity of an assessment argument.

We will use eCollege as the course management system. Assignments and readings (as available) will be posted on the system. All student submissions should be made through the system as well.