

References

- Braun, H. I., & Mislevy, R. (2005). Intuitive Test Theory. *Phi Delta Kappan*, 86(7), 488-497.
- Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890-1930*. New York, NY: New York University Press.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.
- Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621-694). Washington, DC: American Council on Education.
- Dee, T. S., & Jacob, B. A. (2010). The Impact of No Child Left Behind on Students, Teachers, and Schools. *Brookings Papers on Economic Activity* (pp. 149-194). Washington, DC: Brookings Institution Press.
- Diamond, J. B. & Cooper, K. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago. In P. A. Moss (Ed.), *Evidence and decision making* (pp. 241-263). Malden, MA: National Society for the Study of Education.
- Guskey, T. R. (Ed.) (1994). *High Stakes Performance Assessment: Perspectives on Kentucky's Educational Reform*. Thousand Oaks, CA: Corwin Press.
- Haertel, E. H. (2013). How is testing supposed to improve schooling? *Measurement: Interdisciplinary Research and Perspectives*, 11, 1-18.
- Haertel, E., & Ho, A. (2016). Fairness using derived scores. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational measurement and assessment* (pp. 217-237). New York: Routledge.
- Haertel, E. H., & Lorié, W. A. (2004). Validating standards-based test score interpretations. *Measurement: Interdisciplinary Research and Perspectives*, 2(2), 61-103.
- Ho, A. D. (2008). The problem with "proficiency." *Educational Researcher*, 37(6), 351-360.
- Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education/Praeger.
- Kane, M. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, MA: Harvard University Press.
- Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990-2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, 34(2), 209-231.
- McQuillan, P. J., & Salomon-Fernandez, Y. (2008). The impact of state intervention on "underperforming schools" in Massachusetts: Implications for policy and practice. *Education Policy Analysis Archives*, 16(18). Downloaded from <http://epaa.asu.edu/ojs/article/view/43/169>.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education and Macmillan.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington, DC: The National Academies Press. doi: 10.17226/23409.
- No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

- Reardon, S.F., Greenberg, E.H., Kalogrides, D., Shores, K.A., & Valentino, R.A. (2013). Left Behind? The Effect of No Child Left Behind on Academic Achievement Gaps. (downloaded from <https://cepa.stanford.edu/content/left-behind-effect-no-child-left-behind-academic-achievement-gaps>, July 10, 2017)
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.G. Gifford & M.C. O'Conner (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: political rhetoric and measurement reality *Educational Researcher*, 21(4), 22-27.
- Walker, D. F., & Schaffarzick, J. (1974). Comparing curricula. *Review of Educational Research*, 44, 83-111.
- Washburne, C. W. (1925). A program of individualization. In C. W. Washburne (Chairman), *Adapting the schools to individual differences* (Twenty-fourth yearbook of the National Society for the Study of Education, Part II, pp. 257-272). Bloomington, IL: Public School Publishing Co.
- Wei, X. (2012). Are more stringent NCLB state accountability systems associated with better student outcomes? An analysis of NAEP results across states. *Educational Policy*, 26(2), 268-308.
- Wiliam, D. (1998, July). *The validity of teachers' assessments*. (Paper presented to Working Group 6 [Research on the Psychology of Mathematics Teacher Development] of the 22nd annual conference of the International Group for the Psychology of Mathematics Education. Stellenbosch, South Africa). (Retrieved from http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/PME%2022%20WG6%20paper.doc)
- Wong, M., Cook, T. D., & Steiner, P. M. (2011). No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series (WP-09-11) *Institute for Policy Research Working Paper Series*. Evanston, IL: Northwestern University.