

## References

- Allen, N.L., Carlson, J.E., & Zalanak, C.A. (1999). *The NAEP 1996 technical report*. Washington, DC: National Center for Education Statistics.
- American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.) *Testing structural equation models*. Pp. 136-162. Beverly Hills, CA: Sage.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 292-334.
- Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified parallel tests. *Educational and Psychological Measurement*, *25*, 291-312.
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In *Differential item functioning*, edited by Paul W. Holland & Howard Wainer. Hillsdale, NJ: Lawrence Erlbaum.
- Feldt, L. S., & Brennan, R. L. (1989) *Reliability*. In Linn, R. L. (ed.), *Educational measurement*. New York:Macmillan.
- Holland, P. W., & Thayer, D. T. (1988). "Differential Item Performance and the Mantel-Haenszel Procedure." In *Test Validity*, edited by Howard Wainer and Henry I. Braun. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria in fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55.
- Kim, S., & Kolen, M. J. (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models*. Iowa City, IA: Iowa Testing Programs, The University of Iowa.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading Massachusetts: Addison-Wesley Publishing Company.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*, 5-11.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological measurement*, *16*, 159-176.
- No Child Left Behind Act of 2001, 20 U.S.C. 6301 et seq (2001) (PL 107-110).
- Spray, J. A., Ackerman, T. A., Reckase, M. D., & Carlson, J. E. (1989). Effect of the medium of item presentation on examinee performance and item characteristics. *Journal of Educational measurement*, *26*, 261-271.
- Suen, H. K. (1990). *Principles of test theories*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Thissen, D., Chen, W-H., & Bock, R. D. (2003). *MUTILOG for Windows, Version 7* [Computer Software]. Lincolnwood, IL: Scientific Software International.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [today's date], from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>