

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 292–334.
- ETS (2002). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service.
- ETS (October, 5, 2009). *Considerations for Setting New Target Test Characteristic Curves for the Maryland High School Assessments*. Princeton, NJ: Educational Testing Service.
- ETS (October, 29, 2010). *Study of the Comparability of Online and Paper Forms of the May 2010 Maryland High School Assessments*. Princeton, NJ: Educational Testing Service.
- Haertel, E. H. (2006). Reliability. In R. H. Brennan (Ed.) *Educational measurement* (4th edition, pp. 64–110). Westport, CT: Praeger.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.
- Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer-Verlag.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classification based on test scores. *Journal of Educational Measurement*, *32*, 179–197.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.) *Educational measurement* (3rd edition, pp. 13–103). New York: American Council on Education and Macmillan.
- Muraki, E., & Bock, D. (1995). Proprietary version of PARSCALE [Computer software]. Chicago: Scientific Software International.
- Muthén, B.O. (1998–2004). Mplus Technical Appendices. Los Angeles: Muthén & Muthén.
- Muthén, B., du Toit, S.H.C. & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished manuscript.
- Muthén B. O., & Muthén, L. K. (2007). Mplus 5 [Computer program]. Los Angeles: Muthén & Muthén.
- Stocking, M. L., & Lord, F. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*, 201–210.

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement*, *21*, 93–111.

Yen, W. M., & Fitzpatrick, A. R. (2006). *Item response theory*. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111–153). Westport, CN: Greenwood Publishing Group.