REFERENCES

AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington, D.C.: Author.

Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). *The NAEP 1998 technical report* (Technical Report). Washington, DC: National Center for Educational Statistics.

Andrich, A. (1988). *Rasch models for measurement*. Newbury Park, CA: SAGE Publications, Inc.

Andrich, A. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath, & H. H. Lovibond (Eds.) *Mathematical and theoretical systems*. North-Holland: Elsevier Science Publisher B.V.

Andrich, A., & Luo, G. (2004). *Modern measurement and analysis in social science*. Murdoch University, Perth, Western Australia.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: SAGE Publications.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt Rinehart Wilson.

CTB/McGraw-Hill (2004, August). *The Maryland standard setting technical report*. (Technical Report). Monterey, CA: CTB/McGraw-Hill.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed-response and differential item functioning: A pragmatic approach* (ETS Research Report No. 91-49). Princeton, NJ: Educational Testing Service.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists.* New Jersey: Lawrence Erlbaum Associates, Publishers.

Haertel, E. H. (1996). *Estimating the decision consistency from a single administration of a performance assessment battery. A report on the National Board of Professional Teaching Standards McGEN Assessment*. Palo Alto, CA: Stanford University.

Harcourt, Inc. (2008, January). *Maryland School Assessment-Reading and Mathematics: Test administration and coordination manual*. San Antonio, TX: Harcourt Inc.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications, Inc.

Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practice,* 10, 181-189.

Huynh, H., Meyer III, J. P., & Barton, K. (2000). *Technical documentation for the 1999 Palmetto achievement challenge tests of English language arts and mathematics, grades three through eight* (Technical Report). Columbia: South Carolina Department of Education.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8 & PRELIS 2: User's reference guide*. Chicago: Scientific Software International.

Linacre, J. M., & Wright, B. D. (2000). *A user's guide to WINSTEPS: Rasch-model computer program.* Chicago, IL: MESA Press.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32,* 179-197.

Loehlin, J. C. (1987). *Latent variable models.* NJ: Lawrence Erlbaum Associates, Publishers.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement, 8,* 452-461.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58,* 690-700.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrica, 47,* 149-174.

Maryland State Department of Education. (2008). *2008 Maryland School Assessment- Grades 3 through 8 Reading and Mathematics: Test Administration and Coordination Manual*. Baltimore: Maryland State Department of Education.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18,* 5-11.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards* (pp. 249-282). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Orlando, M. (2004, June). *Critical issues to address when applying item response theory (IRT) models*. Paper presented at the The Drug Information Association, Bethesda, MD.

Qualls, A. L. (1995). Estimating the reliability of a test containing multiple item formats, *Applied Measurement in Education,* 8, 111-120.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: University of Chicago Press.

Ryan, J. P. (1983). Introduction to latent trait analysis and item response theory. In W. E. Hathaway (Ed.), *Testing in the schools. New directions for testing and measurement, 19,* San Francisco: Jossey-Bass.

South Carolina Department of Education. (2001). *Technical documentation for the 2000 Palmetto achievement challenge tests of English language arts and mathematics* (Technical Report). Columbia: South Carolina Department of Education.

Suen, H. K. (1990). *Principles of test theories*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrica, 51,* 567-577.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis,* MESA PRESS, Chicago.

Young, M. J., & Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment.* (CSE Technical Report 475). Center

for the Study of Evaluation, Standards, and Student Testing. Los Angeles, CA: University of California, Los Angeles.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30,* 233-251.