**REFERENCES**

Board, C., & Whitney, D. R. (1972). The effect of selected poor item writing practices on test difficulty, reliability, and validity. *Journal of Educational Measurement*, *9*, 225-233.

Camilli, G., Shepart, L.A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

Dorans, N.J., & Holland, P.W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland & H. Wainer (Eds.). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dudycha, A.L., & Carpenter, J. (1973). Effects of item format on item discrimination and difficulty. *Journal of Applied Psychology, 58*, 116-121.

Green, K. (1984). Effects of item characteristics on multiple-choice item difficulty. *Educational and Psychological Measurement, 44*(3), 551-561.

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

Henrysson, S. (1971). Gathering, analyzing, and using data on test items. In R.L. Thorndike (Ed.) *Educational measurement* (2nd Ed., pp. 130-159). Washington, DC: Council on Education.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher,* 23, 13-23.

Mislevy, R.J. (1988). Exploiting auxiliary information about items in the estimation of Rasch item difficulty parameters. *Applied Psychological Measurement, 12*(3), 281-296.