

# Bibliography

American Educational Research Association, American Psychological Association, National Council on Measurement in Education,

- Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*. New York: Springer.
- Lee, W. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement, 47*, 1-17.
- Linacre, J. M. (2018). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com.
- Lohr, S. L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Duxbury Press.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5–11.
- O'Malley, K., Keng, L., & Miles, J. (2012). Using validity evidence to set performance standards. In G.J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations* (2<sup>nd</sup> ed., pp. 301–322). New York: Routledge.
- Petersen, N. S. (1982). *Setting Standards, Setting Goals* (1<sup>st</sup> ed.). Englewood Cliffs, NJ: Prentice-Hall.

Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-24.

Torgerson, W. S. (1958). *Theory and Methods of Scaling*