

## Reading List for EPRS9350 and EPRS9360

<Caution> References below are for your information only. Do not copy and paste these references directly into your APA style document. The format below may not be in the correct APA style.

### 1. Books:

- Baker, F. B. (1992). *Item Response Theory: Parameter estimation techniques*. New York: M. Dekker.  
(On reserve through Net Library:the library does not own a physical copy of this book)
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Reinhart, and Winston.
- de Ayala, R. J. (2008). *The theory and practice of item response theory* . New York, NY: Guilford.
- Embretson, S.E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ. L. Erlbaum Associates.
- Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Elbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA.: Addison-Wesley.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.
- van der Linden, W. J. & Hambleton, R. (1997). *Handbook of modern item response theory*. New York: Springer.
- Warm, T. A. (1978). *A primer of item response theory*. Technical Report 941078. Oklahoma City: U.S. Coast Guard Institute.

### 2. Major Journals in which you can find IRT articles

- Journal of Educational Measurement (JEM)
- Applied Psychological Measurement (APM)
- Applied Measurement in Education (AME)
- Educational Measurement: Issues and Practice (EMIP)
- Educational and Psychological Measurement (EPM)

### 3. Journal articles: Note: The asterisk \* indicates reading materials for EPRS9360.

- Ackerman, T., Gierl, M. & Walker, C.M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *EPIP*, *22*(3), 37-54. (ITEM - PDF)
- \*Banh, M., Crane, P., Rhew, I., Gudmundsen, G. Stone, A. Lyon, A, & McCauley, E. (2012). Measurement equivalence across racial/ethnic groups of the mood and feelings questionnaire for childhood depression. *Journal of Abnormal Child Psychology*, *40*(3), 353-367.
- Banks, K. (2013). A synthesis of the peer-reviewed differential bundle functioning research. *Educational Measurement: Issues & Practice*, *32*(1), 43-55.
- Camilli, G. (1994). Origin of the scaling constant  $d = 1.7$  in item response theory. *Journal of Educational and Behavioral Statistics*, *19*, 293-295.
- Cappelleri, J. C., Lundy, J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, *36*(5), 648-662.
- Clauser, B.E. & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *EMIP*, *17*(1), 31-44. (ITEM - PDF)
- Cohen, A. S, & Kim, S. H. (1992). Detecting calculator effects on item performance. *AME*, *5*, 303-320.
- \* Croudace, T., Ploubidis, G., & Abbott, R. (2005). Statistical software review. *British Journal Of Mathematical & Statistical Psychology*, *58*(1), 193-195.
- De Ayala, R. J. (1993). An introduction to polytomous item response theory. *Measurement & Evaluation in Counseling & Development*, *25*, 172-190.
- De Ayala, R. J. (1993). Fundamentals of item response theory, by Ronald K. Hambleton, H. Swaminathan, H. Jane Rogers. (Book Review) *JEM*, *30*, 84-87.
- DeMars C. An analytic comparison of effect sizes for differential item functioning. *Applied Measurement In Education* [serial online]. July 2011;24(3):189-209.
- Edelen, M., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality Of Life Research*, 165-18.
- Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *JEM*, *45*, 225-245.
- Flowers, C. P., Oshima, T. C., & Raju, N. S. (1999). A description and demonstration of polytomous-DFIT framework. *Applied Psychological Measurement*, *23*, 309-326. (PDF)
- Forbey, J. D., & Ben-Porath, Y. S. (2007). Computerized adaptive personality testing: A Review and illustration with the MMPI—2 computerized adaptive version. *Psychological Assessment*, *19*(1),

14-24.

- \*French, R. F., & Finch, W. H. (2010). Hierarchical logistic regression: Accounting for multilevel data in DIF detection. *Journal of educational measurement*, 47(3), 299-317.
- Galli, S., Chiesi, F., & Primi, C. (2011). Measuring mathematical ability needed for “non-mathematical” majors: The construction of a scale applying IRT and differential item functioning across educational contexts. *Learning & Individual Differences*, 21(4), 392-402.
- Gierl, M. J., Bisanz, J., Bisanz, G., & Boughton, K. A. (2001). Illustrating the utility of differential bundle functioning analyses to identify and interpret group differences on achievement tests. *EMIP*, 20(2), 26-36. (ITEM -PDF)
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *EMIP*, 12(3), 38-47. (ITEM - PDF)
- Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *EMIP*, 8(1), 35-41. (ITEM - PDF)
- \*Kim, J., & Oshima, T. C. (2013). Effect of multiple testing adjustment in DIF detection. *Educational and Psychological Measurement*, 73, 458-470.
- Lee, W., & Ban, J. (2010). A comparison of IRT linking procedures. *Applied Measurement In Education*, 23(1), 23-48.
- Linn, R. L., & Drasgow, F. (1987). Implications of the Golden Rule settlement for test construction. *Educational Measurement: Issues and Practice*, 6, 13-17.
- Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49(4), 305-328.
- \*McCarty, F., Oshima, T. C., & Raju, N. S. (2007). Identifying possible sources of differential functioning using differential bundle functioning with polytomously scored data. *Applied Measurement in Education*, 20, 205-225.
- \*Meriac, J., Woehr, D. & Banister, C. (2010). Generational differences in work ethic. An examination of measurement equivalence across three cohorts. *Journal of Business & Psychology*, 25(2), 315-324.
- \*Nanda, A. O., Oshima, T. C., & Gagne, P. (2006). DIFCUT: An SAS program for conducting significance tests for differential functioning of items and tests (DFIT). *Applied Psychological Measurement*, 30, 150-151.
- Millsap, R. E., & Everson, H. T. (1993). Methodological Review: Statistical approaches for assessing measurement bias. *APM*, 17, 297-334.

- Neel, J. H. (1992). TESTINFO: A graphics-oriented program for selecting items for target test information and standard error of measurement functions. *APM*, *16*, 260.
- Neel, J. H. (1994). GRAPHDIF. *APM*.
- Oshima, T. C. (1994). The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement*, *31*, 200-219.
- Oshima, T. C., & Morris, S. B. (2008). An NCME instructional module on Raju's differential functioning of items and tests (DFIT). *Educational Measurement: Issues and Practice*, *27*, 43-50.
- Oshima, T. C., Raju, N. S., Flowers, C. P., & Slinde, J. A. (1998). Differential bundle functioning using the DFIT framework: Procedures for identifying possible sources of differential functioning. *AME*, *11*, 353-369. (PDF)
- Paek, I., & Han, K. T. (2013). IRTPRO 2.1 for Windows (Item response theory for patient-reported outcomes). *Applied Psychological Measurement*, *37*(3), 242-252.
- Pastor, D. A. (2003). The use of multilevel item response theory modeling in applied research: An illustration. *AME*, *16*, 223-244.
- Penfield, R. (2014). An NCME instructional module on polytomous item response theory models. *Educational Measurement: Issues & Practice*, *33*(1), 36-48.
- Pohl, S., Gräfe, L., & Rose, N. (2014). Dealing with omitted and not-reached items in competence tests: Evaluating approaches accounting for missing responses in item response theory models. *Educational & Psychological Measurement*, *74*(3), 423-452.
- Qi, D., & van der Linden, W. J. (2013). Integrating test-form formatting into automated test assembly. *Applied Psychological Measurement*, *37*(5), 361-374.
- Raju, N. S., van der Linden, W. J., & Fler, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *APM*, *19*, 353-368.
- Raju, N. S., & Oshima, T. C. (2005). Two prophecy formulas for assessing the reliability of item response theory - based ability estimates. *Educational and Psychological Measurement*, *65*, 361-375.
- Rebollo, P., Castejón, I., Cuervo, J., Villa, G., García-Cueto, E., Díaz-Cuervo, H., & ... Alonso, J. (2010). Validation of a computer-adaptive test to evaluate generic health-related quality of life. *Health & Quality Of Life Outcomes*, 8147-154. doi:10.1186/1477-7525-8-147
- Reckase M. D. (2000, April). Computerized testing -The adolescent years: Juvenile delinquent or positive role model? Paper presented at the annual meeting of the National council on

Measurement in Education, New Orleans.

- Reid, C. A., Kolakowsky-Hayner, S. A., Lewis, A. N., & Armstrong, A. J. (2007). Modern psychometric methodology: Applications of item response theory. *Rehabilitation Counseling Bulletin, 50*(3), 177-188.
- Roberts, J. S. (2008). Modified likelihood-based item fit statistics for the generalized graded unfolding model. *Applied Psychological Measurement, 32*(5), 407-423.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*(1), 3-32.
- Rupp, A. A. (2003). Item response modeling with BILOG-MG and MULTILOG for Windows. *International Journal of Testing, 3*(4), 365-384.
- \*Sangwon, K., Kim, Seock-Ho, & Kamphaus, R. W. (2010). Is aggression the same for boys and girls? Assessing measurement equivalence with confirmatory factor analysis and item response theory. *School Psychology Quarterly, 25*(1), 45-61.
- Seybert, J., & Stark, S. (2012). Iterative linking with the differential functioning of items and tests (DFIT) method: Comparison of testwide and item parameter replication (IPR) critical values. *Applied Psychological Measurement, 36*(6), 494-515.
- \*Seybert, J., Stark, S., & Chernyshenko, O. S. (2014). Detecting DIF with ideal point models: A comparison of area and parameter difference methods. *Applied Psychological Measurement, 38*(2), 151-165.
- Shaftel, J., Xiangdong, Y., Glasnapp, D., & Poggio, J. (2005). Improving assessment validity for students with disabilities in large-scale assessment programs. *Educational Assessment, 10*(4), 357-375.
- Sinharay, S., & Haberman, S. J. (2014). How often is the misfit of item response theory models practically significant?. *Educational Measurement: Issues & Practice, 33*(1), 23-35.
- Smith, L. L. (2002). On the usefulness of item bias analysis to personality psychology. *Personality and Social Psychology Bulletin, 28*, 754-763.
- Stone, C. A., & Bo, Z. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *Journal of Educational Measurement, 40*(4), 331-352.
- Sun, K., Chen, Y., Tsai, S., & Cheng, C. (2008). Creating IRT-based parallel test forms using the genetic algorithm method. *Applied Measurement In Education, 21*(2), 141-161.
- Stone, C. A., & Zhang B. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *JEM, 40*, 331-353.
- \*Teresi, J. A., Ocepek-Welikson, K., Ramirez, M., Kleinman, M., Ornstein, K. & Siu, A. (2015).

Evaluation of measurement equivalence of the Family Satisfaction with the End-of-Life Care in an ethnically diverse cohort: Tests of differential item functioning. *Palliative Medicine*, 29(1), 83-96.

Toland, M. D. (2014). Practical guide to conducting an item response theory analysis. *Journal of Early Adolescence*, 34(1), 120-151

Traub, R. E., & Rowley, G. L. (1991). Understanding reliability, EMIP, 10(1), 37 – 45. (ITEM - PDF)

Weiss, D. J. (2004). Computer adaptive testing for effective and efficient measurement in counseling and education. *Measurement & Evaluation in Counseling & Development*, 37, 70-85.

\*Wiberg, M. (2007). Measuring and detecting differential item functioning in criterion referenced licensing test: A theoretic comparison of methods (Department of Educational Measurement Research Report EM No. 60). Sweden: UMEA University.

\*Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-Improved Wald Test for DIF Testing With Multiple Groups: Evaluation and Comparison to Two-Group IRT. *Educational & Psychological Measurement*, 73(3), 532-547.

\*Wright, K., & Oshima, T. C. (2015). An effect size measure for Raju's Differential Functioning for Items and Tests. *Educational and Psychological Measurement*. April, 29, 2014, DOI: 10.1177/0013164414532944

You can obtain ITEMS (ITEMS: The Instructional Topics in Educational Measurement Series) article by visiting: <http://ncme.org/publications/items/>