

Ed Psy 8-222
Advanced Measurement: Theory and Applications
and
Psy 5-865
Advanced Psychological and Educational Measurement

Spring 2008
Lectures: 11:15 - 12:30 Tuesday & Thursday: 125 Burton Hall
Lab: Wednesday 2:00 - 4:00: 325 Peik Hall

Part 2: Item Response Theory

*Note: There is a Displayed Materials packets available at the Coffman bookstore:
This packet is essential since a considerable amount of visual material is used in the lectures.*

The Instructor for Part 2 (March 13 through May 8) is Professor David J. Weiss, N660 Elliott Hall, phone 625-0342, email djweiss@umn.edu. The teaching assistant and lab instructor is Ben Babcock, N667 Elliott Hall, phone 5-5381, e-mail babco062@umn.edu; Ben's office hours during Part 2 are from *** to *** on **days.

Grades for the course will be based on the following:

1. A paper based on Part 1.
2. Seven graded lab assignments for Part 2. Each lab assignment will account for 10% of your Part 2 grade. The lab assignments will be due one week after they were assigned. Lab assignments that are submitted late will be subject to a 20% penalty for each day they are late. The take-home exam will account for the remaining 30% of the Part 2 grade. Your course grade will be based on the sum of equally weighted Part 1 and Part 2 grades.
3. A take-home examination. The exam will include two questions from Part 1, two questions from Part 2, and two questions that integrate the material from Parts 1 and 2. The exam must be typed; single-spaced is acceptable. The take-home exam is due on May 16, following the guidelines described in the Part 1 syllabus. It should be submitted electronically (as a Word-compatible file) to djweiss@umn.edu.

Late paper/exam policy: The instructors have as much time to read the paper as the paper is late. That means that if you submit your paper three months after the course is over, we have three months to read it (see further conditions below). If your paper is two years late, we have the option of taking two years to read it (we *might* not exercise that option, but reserve that right). Furthermore, we do *not* read papers during the summer break, nor do we read them during sabbatical or other leaves. Therefore, if you will need your grade in this course for a specific purpose (e.g., to take a prelim oral or to graduate) be sure to submit your papers/exam with sufficient time for us to read it under the policies stated above.

Part 2: Item Response Theory and Methods

Text: Embretson, S.E.. & Reise, S. P. (2000). Item response theory for psychologists. Mahwah NJ: Erlbaum.

- March 12 (lab, 1 hr.) **Basic concepts of IRT: Models and item response functions for dichotomously scored data: Rasch, 2- and 3-parameter models, normal ogive and logistic models**
- Mar. 13 **A brief history of IRT. IRT versus classical test theory (Guest Lecturer:)**
Readings: Bock; E&R chs. 1-3, 9; Weiss & Yoes
- Mar.17 - 24 **No classes: Spring Break**
- Mar. 25 - Mar. 27 **Rationale(s), parameters, uses, and transformations of the IRF.**
Readings: E&R ch. 4, 6, 7; Andrich; Lord (1986); Trabin & Weiss
Lab 1: Baker's Basics of IRT; propensity distributions and IRFs
- Apr. 1 – Apr. 3 **Person parameter estimation. The person response function. Item parameter estimation: Joint ML (LOGIST), marginal ML estimation (BILOG and XCALIBRE)**
Readings: E&R ch. 8,13; Baker (1987); Mislevy & Stocking; Baker (1988); Yoes
Lab 2: Person parameter estimation.
- Apr. 8 - Apr. 10 **Information: item, test, test score, response pattern. Efficiency and relative efficiency. Applications of information. Relationships with reliability and precision.**
Readings: H & S, ch. 6; Lord (1974); Samejima
Lab 3: Item parameter estimation .
- Apr. 15 - Apr. 17 **Monte carlo simulation as a research tool. Linking and equating.**
Readings: Harwell et al.; Vale; Davey, et al.
Lab 4: Item fit; information
- Apr. 22 - Apr. 24 **Differential item functioning. Person fit.**
Readings: Millsap & Everson (pp. 307-334); Meijer & Sijtsma; Reise & Flannery;
Lab 5: Monte carlo simulation; linking
- Apr. 29 – May 1 **Adaptive testing; Polytomous models**
Readings: Weiss & Kingsbury; Weiss; E&R ch. 15; De Ayala
Lab 6: DIF; person fit
- May 6 - May 8 **Multidimensional and other models; limitations of IRT.**
Readings: Reckase (1985, 1991); Ackerman (1996); E&R ch.11
Lab 7: adaptive testing; polytomous models

Part 2: Assigned Readings

- Ackerman, T. A. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement, 20*, 311-329.
- Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological Methodology* (pp. 33-80). San Francisco: Jossey-Bass
- Baker, F. B. (1987). Methodology review: Item parameter estimation under the one-, two- and three-parameter logistic models. *Applied Psychological Measurement, 11*, 111-141.
- Baker, F. B. (1988). The item log-likelihood surface for two- and three-parameter item characteristic curve models. *Applied Psychological Measurement, 12*, 387-395.
- Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice, 16* (4), 21-33.
- De Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development, 25*, 172-189.
- Davey, T, Oshima, T. C., & Lee, K. (1996). Linking multidimensional item calibrations. *Applied Psychological Measurement, 20*, 405-416.
- Harwell, M., Stone, C. A., Hsu, T.-C., & Kirisci, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement, 20*, 101-125.
- Lord, F. M. (1974). Quick estimates of the relative efficiency of two tests as a function of ability level. *Journal of Educational Measurement, 11*, 247-254.
- Lord, F. M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement, 23*, 157-162.
- Meijer, R. R. & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*, 107-135.
- Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17*, 297-334.
- Mislevy, R. J., & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement, 13*, 57-75.
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9*, 401-412.
- Reckase, M. D., & McKinley, R. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 15*, 361-373.
- Reise, S. J., & Flannery, W. P. Assessing person-fit on measures of typical performance. *Applied Measurement in Education, 9*, 9-26.

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18*, 229-244.

Trabin, T. E., & Weiss, D. J. (1983). The person response curve: Fit of individuals to item response theory models. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.

Vale, C. D. (1986). Linking item parameters on a common scale. *Applied Psychological Measurement, 10*, 333-344.

Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Clinical and Consulting Psychology, 53*, 774-789.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*, 361-375.

Weiss, D. J., & Yoes, M. E. (1991). Item response theory. In R. K. Hambleton and J. Zaal (eds.), *Advances in educational and psychological testing*. Boston: Kluwer Academic Publishers.

Yoes, M. E. (1995, April). *An updated comparison of microcomputer-based item parameter estimation procedures with the 3-parameter IRT model*. St. Paul MN: Assessment Systems Corporation.

Part 2: Resource Materials

Adaptive Testing

Special Issue: Computerized adaptive testing. *Applied Psychological Measurement*, Volume 23 Number 3, September 1999. (Available from Sage Publications.)

Computerized Adaptive Testing: Theory and Practice. (2000). Edited by: Wim J. Van der Linden and Cees A. W. Glas. Norwell MA: Kluwer.

Wainer, H. et al. (2000). *Computerized adaptive testing: A Primer* (second edition). Mahwah NJ: Erlbaum.

CAT Central Web site: www.psych.umn.edu/psylabs/CATCentral/

Automated Test Construction

Special Issue: Optimal test assembly. *Applied Psychological Measurement*, Volume 22 Number 3, September 1998. (Available from Sage Publications.)

Multidimensional IRT

Special Issue: Developments in multidimensional IRT. *Applied Psychological Measurement*, Volume 20 Number 4, December 1996, plus two articles in the Volume 21 Number 1, March 1997, issue. (Available from APM Inc., N660 Elliott Hall, University of Minnesota.)

Nonparametric IRT

Special Issue: Nonparametric item response theory, *Applied Psychological Measurement*, Volume 25 Number 3, September 2001. (Available from Sage Publications.)

Parameter Estimation for Dichotomous and Polytomous Models

Baker, F. B. & Kim, S.H. (2004). *Item response theory: Parameter estimation techniques (Second Edition)*. New York: Dekker.

Polytomous IRT

Special Issue: Polytomous IRT. *Applied Psychological Measurement*, Volume 19 Number 1, March 1995. (Available from APM Inc., N660 Elliott Hall, University of Minnesota.)

Ostini, R. & Nering, M. (2005). Polytomous item response theory models. *Quantitative Applications in the Social Sciences, Volume 144*. Thousand Oaks CA: Sage Publications.

Special Purpose IRT Models and Applications

Van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer

IRT Software (available from Assessment Systems Corporation, St. Paul MN 55114. www.assess.com, unless otherwise noted)

Rasch Model Analysis: RASCAL, Quest, ConQuest, RUMFOLD/ss/pp, RSP, BIGSTEPS (available from MESA Laboratory, University of Chicago)

Two- and Three-Parameter Model Analysis: XCALIBRE, BILOG-MG, PARSCALE, TESTFACT

Polytomous Models: MULTILOG, PARSCALE

Nonparametric Models: MSP, PARELLA

Dimensionality Assessment: MicroFACT, TESTFACT, LISREL, NOHARM (available from Prof. Roderick P. McDonald, University of Illinois)

Multidimensional Models: TESTFACT, NOHARM (see above)

Test Development Utilities: PARDSIM, TESTINFO, SCOREALL, POSTSIM