



EDUC 8804: Advanced Topics in Measurement Spring 2008

Tue 9:00-11:30 EDUC 330

Instructor

Derek Briggs

Office: EDUC 211 Tel: (303) 492-6320 E-mail: Derek.Briggs@colorado.edu

Office Hours: Tue 1:00-2:00 or by appointment

Course Overview

The focus of this course is on psychometric models for measurement and their applications in educational testing. An emphasis is placed on understanding and evaluating the utility of models from item response theory (IRT). This course is especially appropriate for students expecting to do research in which the score from a test instrument is the basis for evaluative conclusions about learning.

This course takes the approach that the best way to gain an understanding about measurement models is to apply and compare them in the context of simulated or empirical data sets. To this end, readings on various aspects of measurement models will first be presented and discussed in class. Next, the use of these models will be demonstrated by the instructor. Finally, students will be expected to apply the models using data sets provided by the instructor.

The focus of the first half of the semester is foundational topics necessary to help you understand IRT models at an acceptable level of depth.

1. IRT as a Statistical Model
2. Modeling Dichotomous Items (BILOG)
3. Modeling Polytomous Items (MULTILOG)
4. Interpreting the IRT Measurement Scale
5. Estimating the Parameters of an IRT Model
6. Evaluating Model Fit

The focus on the second half of the course will be on extensions and applications of IRT. Potential topics include:

1. Dealing with Violations of Local Independence: Testlet Response Theory
2. Dealing with Violations of Unidimensionality: Multidimensional Item Response Theory
3. Evaluating Test Fairness: Differential Item Functioning
4. Making Tests Comparable: Test Equating, Scaling & Linking
5. Making Test Scores Interpretable: Setting Performance Standards and Reporting Test Scores to the Public

6. Making Tests Formatively Useful: Developing and Modeling Diagnostic Test Items
7. Making Tests Efficient: Computer Adaptive Testing

While these (and other) extensions and applications are both interesting and important, it will only be possible to discuss three of them formally as a whole class given our time constraints. This decision will be made collectively, and will depend upon the interests of enrolled students. (Note: there also be an opportunity to address one of these topics as part of an independent project.)

Students enrolling in this course are expected to have previously taken EDUC 8804: Measurement in Survey Research, in which students are given an introduction to a variety of foundational measurement topics (e.g., survey instrument design, item formats, item analysis, reliability theory and validity theory).

Course Readings

Required Textbook

Embretson, S. E., and Reise, S. P. (2000) *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Highly Recommended Textbooks

Brennan, R. L., ed. (2006) *Educational Measurement*, 4th ed. Westport, CT: American Council on Education/Praeger.

Thissen, D., & Wainer, H., eds. (2001) *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.

Other Recommended Books

* indicates technical material requiring a strong background in statistics

Baker, Frank (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. Available online for free at: <http://edres.org/irt/>

Baker, F. B. & Kim, S-H. (2004) *Item response theory: Parameter Estimation Techniques*. 2nd Edition. New York, NY: Marcel Dekker, Inc.*

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: John Wiley.* [out of print]

- De Boeck, P. & Wilson, M., Eds. (2004) *Explanatory item response models: a generalized linear and nonlinear approach*. New York: Springer.*
- Fischer, G. & Molenaar, I., Eds. (1995) *Rasch models: foundations, recent developments and applications*. New York: Springer.*
- Hambleton, R. K. & Swaminathan, H. (1985) *Item response theory: principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991) *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications.
- Kolen, M. J. & Brennan, R.L. (2004) *Test Equating, Scaling and Linking*. 2nd Edition. New York: Springer-Verlag.
- Lord, F. M. (1980) *Applications of Item response theory to Practical Testing Problems*. Mahwah, NJ: Lawrence Erlbaum Associates.*
- Linn, R.L., Ed. (1989) *Educational Measurement*. 3rd edition. Washington, D.C.: American Council on Education.
- Lord, F. M & Novick, M. R. (1968) *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.* [out of print, high in demand!]
- Smith, E. V., & Smith, R. M., eds. (2004) *Introduction to Rasch Measurement*. Maple Grove, MN: JAM Press.
- Van der Linden, W. J., & Hambleton, R. K. (1997) *Handbook of modern item response theory*. New York, NY: Springer-Verlag.*
- Wright, B. D., & Masters, G. M. (1982). *Rating scale analysis*. Chicago, IL: Mesa Press.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: Mesa Press.

Recommended Journals with a Focus on Educational Measurement

(* indicates a journal that focuses primarily on technical modeling issues in measurement)

Applied Measurement in Education
 Applied Psychological Measurement*
 Educational Assessment
 Educational Measurement: Issues and Practice
 Educational and Psychological Measurement
 Journal of Applied Measurement
 Journal of Educational Measurement*
 Measurement: Interdisciplinary Research and Perspectives
 Psychometrika*

Psychological Bulletin

Internet Resources

National Council for Measurement in Education www.ncme.org

A wonderful feature at this site is the NCME instructional modules, some of which we will be using in this class: <http://www.ncme.org/pubs/items.cfm>

Rasch Measurement <http://www.rasch.org/rmt/index.htm>

Lots of short, informative articles about the application of the Rasch Model.

The Psychometric Society <http://www.psychometrika.org/>

A Visual Guide to IRT by Ivailo Partchev

<http://www.metheval.uni-jena.de/irt/VisualIRT.pdf>

This is a very nice article, with lots of visually appealing applets. We may make use of this in class.

IRT Tutorial from the University of Illinois at Urbana-Champaign

<http://work.psych.uiuc.edu/irt/tutorial.asp>

IRT at Wikipedia http://en.wikipedia.org/wiki/Item_response_theory

This is actually a pretty good summary, believe it or not.

Course Schedule

There will be no class meeting on the following date:

March 25 (Spring Break)

READINGS FOR FOUNDATIONAL TOPICS

All required readings available at course web site: <http://culearn.colorado.edu>

TOPIC 1: Overview and Historical Context of Measurement Models

Required Readings:

Embretson & Reise, Chapter 1

Traub, R. E. (1997) Classical test theory in historical perspective. *Educational Measurement: Issues and Practice*. December 1997, 8-14.

Brennan, R. (1997) A perspective on the history of generalizability theory. *Educational Measurement: Issues and Practice*. December 1997, 14-20.

Bock, R. D. (1997) A brief history of item response theory. *Educational Measurement: Issues and Practice*. December 1997, 21-32.

Wright, B. D (1997) A history of social science measurement. *Educational Measurement: Issues and Practice*. December 1997, 33-45.

- Lord, F. M & Novick, M. R. (1968) *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company. [Ch. 1, 13-26]
- Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press. [Preface and introduction, xx-xxiii, 3-12]

TOPIC 2: IRT as a Statistical Model

Required Readings:

- Embretson & Reise, Chapter 3
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991) *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications. [Ch. 2]
- Hambleton, R., & Jones, R. (1993) An NCME instructional module: Comparison of Classical Test Theory and Item Response Theory and their applications to test development. *Educational Measurement: Issues and Practice* 12 (3), 38–47.

Going Deeper

- Lord, F. M and Novick, M. R. (1968) *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.* [Ch. 16]
- De Boeck, P. and Wilson, M., eds. (2004) *Explanatory item response models: a generalized linear and nonlinear approach*. New York: Springer.* [Ch. 1-2]
- Van der Linden, W. J., and Hambleton, R. K. (1997) *Handbook of modern item response theory*. New York, NY: Springer-Verlag.* [Preface, 1-28]

TOPIC 3: Modeling Dichotomous Items

Required Readings

- Embretson & Reise, Chapter 4.
- Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice* 8 (1), 35–41.
- BILOG User's Manual

Going Deeper:

- Thissen, D., & Orlando, M. Ch. 3 (73-98) in D. Thissen and H. Wainer, eds. (2001) *Test Scoring* Mahwah, NJ: Lawrence Erlbaum Associates.
- Lord, F. M and Novick, M. R. (1968) *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.* [Ch. 17]

TOPIC 4: Modeling Polytomous Items

Required Readings

- Embretson & Reise, Chapter 5.
- MULTILOG User's Manual

Going Deeper

Thissen et. al, Ch. 4 (141-186) in D. Thissen and H. Wainer, eds. (2001) *Test Scoring*
Mahwah, NJ: Lawrence Erlbaum Associates.

De Boeck, P. and Wilson, M., eds. (2004) *Explanatory item response models: a
generalized linear and nonlinear approach*. New York: Springer.* [Ch. 3]

TOPIC 5: Interpreting the IRT Measurement Scale

Required Readings

Embretson & Reise, Chapter 6.

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991) *Fundamentals of item
response theory*. Newbury Park, CA: SAGE Publications. [Ch. 5]

Going Deeper

Kolen, M. and Brennan, R.L. (2004) *Test Equating, Scaling and Linking*. 2nd Edition.
New York: Springer-Verlag. [Ch. 6, 155-230]

Kolen, M (2006) Scaling and norming. In R. Brennan, (ed.) *Educational Measurement*,
4th ed, Westport, CT: American Council on Education/Praeger. [Ch. 5, 155-186]

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests*.
Chicago: University of Chicago Press. [Ch. 5-6, 62-108]

TOPIC 6: Estimating the Parameters of IRT Models

Required Readings

Embretson & Reise, Chapter 7-8

Mislevy, R. & Stocking, M. (1989) A consumer's guide to LOGIST and BILOG.
Applied Psychological Measurement, 13(1), 57-75.

Going Deeper

Thissen & Orlando, Ch. 3 (98-137) in *Test Scoring*.

Thissen et. al, Ch. 4 (149-173) in *Test Scoring*.

If You Really Want to be an Expert

Baker, F. B. and Kim, S-H. (2004) *Item response theory: Parameter Estimation
Techniques*. 2nd Edition. New York, NY: Marcel Dekker, Inc.*

TOPIC 7: Evaluating Model Fit

Required Readings

Embretson & Reise, Chapter 9

Hambleton, R. K. & Swaminathan, H. (1985) *Item response theory: principles and
applications*. Boston: Kluwer-Nijhoff Publishing. [Chapters 8-9, 151-196]

Yen, W. (2006) Item response theory. In R. Brennan, (ed.) *Educational Measurement*,
4th ed, Westport, CT: American Council on Education/Praeger. [Ch. 4, 138-143]

BONUS TOPIC: Generalizability Theory

Required Readings:

Brennan, R. (1992) NCME Instructional Module on Generalizability Theory.

Thompson, B. (2003) A brief introduction to generalizability theory. In *Score Reliability*, B. Thompson, ed. Thousand Oaks, CA: SAGE Publications.

Going Deeper

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A Primer*. Newbury Park: Sage Publications.

If You Really Want to be an Expert

Brennan, R. (2001). *Generalizability Theory*. New York: Springer-Verlag.*

Briggs, D. C., & Wilson, M. (2007) Generalizability in item response modeling. *Journal of Educational Measurement*, 44(2), 131-155.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: John Wiley.* [out of print]

Topic Discussions

You will be expected to team with one or two or your classmates on a regular basis to lead classroom discussions or activities on selected topics. Each student team will meet with me to discuss their plan of action.

Data Analyses

You will be given two data analysis assignments in which you will typically apply and/or compare measurement models in the context of simulated or empirical data sets that I will provide. (Note: you are welcome to use your own data sets, provided they meet certain requirements in terms of number of items and respondents.) The assignment writeup should be between 5 and 10 double-spaced pages, including figures (but not including references) You are required to turn in these analyses for feedback. This feedback will include suggestions for improving your analysis. It will be up to you to decide when you wish to turn in these assignments, but keep in that I will need a minimum of one week turnaround time to give you feedback on your work.

Assignments:

Data Analysis #1: Modeling dichotomous test items using the software BILOG

Data Analysis #2: Modeling polytomous test items using the software MULTILOG

Details on these specific assignments and accompanying data sets will be forthcoming as separate handouts.

Independent Project

As will become clear, there are many, many applications and extensions of IRT, especially in the context of large-scale standardized testing. For this project you will be given the flexibility to pick your own application/extension to pursue in detail.

Here are just a few examples of topics you could choose:

- Checking for multidimensionality in a test instrument.
- Checking for differential item functioning in a test instrument.
- Comparing standard setting approaches for tests scaled using IRT.
- Applying an IRT Model Tailored for Diagnostic Assessment

A good place to look for ideas for your independent project would be Chapters 10-12 in the Embretson & Reise textbook. I expect you to tailor this project to your specific interests in educational measurement.

- Turn in a 1-2 page proposal for your independent project by March 4th.
- At that point I will meet with each of you to make a plan for what readings you should consult and data you should analyze as part of your project.
- The page limit for this project is 20 pages, double-spaced.
- The final draft of your project will be due by 10:00 on May 3rd.

Software

We will be using software for two programs as part of this course.

1. BILOG-MG 3.0. Item response models for dichotomous items.
2. MULTILOG 7. Item response models for dichotomous and polytomous items.

At the following web site <http://estore.e-academy.com/index.cfm?loc=IRT/main> you can try (for 15 days), or rent (for 6 months, \$40), the programs above.

Neither of the programs above are especially intuitive. Figuring out how to use them effectively will be something we will work on together in class.

Student Assessment

Grades for this course will be based on the following criteria

Leading Topic Discussions & Class Participation	20%
Data Analyses	50%
Independent Project	30%

Reasonable Accommodation

If you qualify for accommodations because of a disability, please submit to me a letter from Disability Services in a timely manner so that your needs may be addressed. Disability Services determines accommodations based on documented disabilities. Contact: 303-492-8671, Willard 322, and www.Colorado.EDU/disabilityservices

Disability Services' letters for students with disabilities indicate legally mandated reasonable accommodations. The syllabus statements and answers to Frequently Asked Questions can be found at www.colorado.edu/disabilityservices

Religious Observances

I will make every effort to accommodate all students who, because of religious obligations, have conflicts with scheduled exams, assignments, or other required attendance, provided advance notification of the conflict is given. Whenever possible, students should give at least two weeks advance notice to request special accommodation. For additional information on this policy, see http://www.colorado.edu/policies/fac_relig.html

Classroom Behavior

Students and faculty each have responsibility for maintaining an appropriate learning environment. Students who fail to adhere to such behavioral standards may be subject to discipline. Faculty have the professional responsibility to treat all students with understanding, dignity and respect, to guide classroom discussion and to set reasonable limits on the manner in which they and their students express opinions. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender variance, and nationalities. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. See policies at <http://www.colorado.edu/policies/classbehavior.html> and at http://www.colorado.edu/studentaffairs/judicialaffairs/code.html#student_code

The University of Colorado at Boulder policy on Discrimination and Harassment (<http://www.colorado.edu/policies/discrimination.html>), the University of Colorado policy on Sexual Harassment and the University of Colorado policy on Amorous Relationships applies to all students, staff and faculty. Any student, staff or faculty member who believes s/he has been the subject of discrimination or harassment based upon race, color, national origin, sex, age, disability, religion, sexual orientation, or veteran status should contact the Office of Discrimination and Harassment (ODH) at 303-492-2127 or the Office of Judicial Affairs at 303-492-5550. Information about the ODH and the campus resources available to assist individuals regarding discrimination or harassment can be obtained at <http://www.colorado.edu/odh>

Student Honor Code

All students of the University of Colorado at Boulder are responsible for knowing and adhering to the academic integrity policy of this institution. Violations of this policy may include: cheating, plagiarism, aid of academic dishonesty, fabrication, lying, bribery, and threatening behavior. All incidents of academic misconduct shall be reported to the Honor Code Council (honor@colorado.edu; 303-725-2273). Students who are found to be in violation of the academic integrity policy will be subject to both academic sanctions from the faculty member and non-academic sanctions (including but not limited to university probation, suspension, or expulsion). Other information on the Honor Code can be found at <http://www.colorado.edu/policies/honor.html> and at <http://www.colorado.edu/academics/honorcode/>