RMTD 530
Item Response Theory

Instructor: Ken A. Fujimoto, Assistant Professor
Classroom: Corboy Law Center—Room 710
Class time: Monday, 4:00—6:45pm
Office hour: By appointment
Email: kfujimoto@luc.edu
Office: Lewis Towers Room 1037

Course content

Educational and psychological tests are commonly used to measure latent traits (e.g., knowledge, aptitude, achievement, attitude, and personality), often with the goal of making inferences on individuals or groups of individuals. Item response theory (IRT) measurement models are useful tools in examining the quality of the data that arises from the administration of such tests, which in turn allows us to make judgements about the validity of the test scores (or latent trait estimates) and the quality of the tests (are the items confusing and/or are they biased).

This course will introduce students to commonly used IRT models for analyzing dichotomous (e.g., 0=incorrect and 1=correct) and polytomous (0=never, 1=sometimes, 2=always) test data. While much of the course will be application oriented, some technical details that facilitate the understanding of the IRT and highlight the advantages of it over classical test theory (CTT) will be provided. Knowledge of basic algebra, linear modeling, and CTT is required.

Technological Knowledge and Skills

Students will use the IRTPro software to analyze real-life (example), simulated test data, and a data set of the student's choosing. Additionally, students will use SPSS, SAS, and Excel to prepare the data and create figures that are visual representations of the IRT results.

Computer labs located on the Lakeshore and Water Tower campuses have SPSS, SAS, and Excel. IRTPro 3 is only available in the computer lab that our class will be held. Thus, it is highly recommended that students download the student version of IRTPro 3 (http://www.ssicentral.com/irt/student.html). This version is free.

Required text

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

*Additional required readings will be uploaded to Sakai.

Recommended References On

IRT:

de Ayala, R. J. (2009). Theory and practice of item response theory. Guilford Press, New York, NY.

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

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Mahwah, NJ: Lawrence Erlbaum.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). Statistical theories of mental test scores.

Rasch:

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design. Rasch Measurement.* Chicago: MESA Press.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis. Rasch Measurement.* Chicago: MESA Press.

Validity:

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Kane, M. T. (2006). Validation (4th ed.). In R. L. Brennan (Ed.), *Educational measurement* (pp. 17–64). Westport, CT: Praeger.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–741.

Classical Test Theory:

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Harcourt Brace.

Course Objectives

By the end of the course the student should understand the following:

1. the distinctions among the IRT models for dichotomously scored items (i.e., 1PL, 2PL, and 3PL)
2. the distinctions among the IRT models for polytomously scored items (i.e., PCM, GPCM, RSM, GRM, and MGRM)
3. the mathematical and theoretical assumptions underlying IRT
4. the distinction between latent trait estimates under IRT and CTT
5. the distinction between standard error of measurement under IRT and CTT
6. how the results from various IRT analyses provide evidence of validity of test scores
7. how to perform various IRT analyses using IRTPro
8. write coherent summaries and interpretations of data analyzed
9. what a "fair" item means from an IRT perspective (if time permits)
10. how to perform an IRT analysis that indicates whether an item is fair across specific subgroups within a population (if time permits)

Homework

Five homework assignments will make up the points devoted to homework. The assignments will not be equal in length. Total homework points will be converted to a percentage score, then weighted and combined with exam scores to obtain a final overall grade. You are encouraged to discuss the homework assignments with other students in the class, but each student must separately write up her or his own answers and turn in a copy by the due date.

Homework turned in late will be penalized one letter grade per day it is late unless prior arrangements have been made with the instructor. Please keep in mind that homework assignments turned in late (excused or unexcused) may not receive the same feedback or receive the same turnaround as those that are turned in on time. Late homework should still be uploaded to Sakai or turned in during business hours to the faculty mail box on the 11^{th} floor at the Lewis Tower campus.

If you would like to appeal a grade after your HW is graded, you must make the appeal in writing and submit it along with the graded HW to the instructor within two weeks of receiving the graded HW.

Presentation

You will give a presentation based on an IRT analysis of test data (either dichotomous [e.g., achievement test data] or polytomous [e.g., rating scale data]). The presentation will be about 20 minutes (10-15 minutes for the presentation and 5-10 minutes for questions and answers, and/or comments from the class and me. The presentation format will be similar to a professional research conference, such as the American Educational Research Association annual conference. The presentation should consist of an introduction, method, results, and discussion/conclusions/implications.

You are responsible for obtaining the data set on which your presentation will be based. I recommend that you find an appropriate data set as soon as possible. Do not wait until the last minute. If you are not sure whether a data set is appropriate for an IRT

analysis, please check with me. If you do not have access to a data set, you can use one of the example data sets that comes with the IRTPro software.

Quizzes
There will be no more than three pop quizzes throughout the semester. These quizzes are meant to keep students up to data with the material.

Examinations
There will be two exams in this course. The exams will be open book and open notes, and you may use a calculator and computer during the exams. However, books, notes, computers, and calculators cannot be shared or circulated during exams, so be sure to bring your own material.

Participation
Regular attendance and participation in class discussions are expected. Contact the instructor ahead of the class meeting if you cannot attend the class. If you miss a class, you are responsible for obtaining any information presented in class that day from one of your classmates.

Grades
Grades will be based on points accumulated on the homework and examinations. There will be 100 total possible points, distributed as follows:

Homework assignments	35%
Quizzes	10%
Midterm exam (scheduled time only)	20%
Final exam (scheduled time only)	20%
Presentation	15%

The grade ranges in terms of percentages are:

100.0-92.0 = A	87.9-84.0 = B+	74.9-72.0 = C+	64.9 and below = F
91.9-88.0 = A−	83.9-80.0 = B	71.9-70.0 = C	
	79.9-75.0 = B−	69.9-65.0 = C−	

IDEA Objectives
At the end of the course, you will have an opportunity to complete an Online IDEA course evaluation (http://luc.edu/idea/). The essential objectives for this course are:

1. Gaining factual knowledge (terminology, classifications, methods, trends)
2. Learning fundamental principles, generalizations, or theories
3. Learning to apply course material (to improve thinking, problem solving, and decisions)
4. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course

Learning Community at Loyola University Chicago

Conceptual framework

Our School's Conceptual Framework – ***Social Action through Education*** – guides the curricula of School of Education programs in the preparation of carrying out the mission of social justice. These dimensions of the conceptual framework also serve as the foundation to the School of Education – standards that are explicitly embedded in major benchmarks across all SOE programs. Our conceptual framework is described here: www.luc.edu/education/mission/

Diversity

The School of Education is committed to diversity including but not limited to race, gender, sexual orientation, social class, ethnicity, and ability. Through this course, students will learn how to interpret and critique fundamental research methods used in the social sciences. The course is designed to provide students with the knowledge necessary to evaluate research concerning the social dimensions mentioned above.

Syllabus addendum

Please visit http://luc.edu/education/syllabus-addendum/ for more information regarding academic honesty, accessibility, conceptual framework, EthicsLine Reporting hotline, and electronic communication policies and guideline.

Tentative schedule

*The schedule will most likely change depending on the flow of the class. Please allow for this flexibility.

**ER indicates reading from our course textbook (Embretson & Reise). Other readings will be made available through Sakai.

Slides	Dates	Topics	Readings	Due
	1/18	**Martin Luther King, Jr., Holiday**		
1 & 2	1/25	• Introduction • Comparison of item response theory (IRT) and classical test theory (CTT) • Overview of IRT	• ER: Ch1 & Ch2	
3 & 4	2/1	• 1-Parameter Logistic IRT Model (Rasch Model)—also called the 1PL • 2-Parameter Logistic IRT Models (2PL)	• ER: Ch3 (pp. 40–53) & Ch4 (pp. 65–70)	
4 & 5	2/8	• 2PL (cont'd) • 3-Parameter Logistic (3PL)	• ER: Ch4 (pp. 70–72)	
5 & 6	2/15	• 3PL (cont'd) • Model Selection	• For Model Selection: ER, Ch7 (pp. 160-163) & Ch4 (pp. 72–76)	HW 1
7, 8, & 9	2/22	• Model Identification Issues • Person Ability Estimation • Information functions and Standard Errors	• ER: Ch 3 (pp. 53–61) • ER: Ch 7 (pp. 158–171, 177–179, 183–185)	HW 2
9, 10, & 11	2/29	• Information functions and Standard Errors (cont'd) • Model Assumptions • Diagnostics: Item Fit	• ER: Ch3 (pp. 45–48) • ER: Ch 9 (pp. 231–242 • Orlando & Thissen (2000) • Chen & Thissen (1997)	
	3/7	~~~~~~ **Spring Break** ~~~~~~		

11, 12, & 13	3/14	• Diagnostics: Item Fit (cont'd) • Diagnostics: Local Item Independence/ Dependence • Score Conversion (might carry over to following week)	• ER: Ch3 (pp. 107)	HW 3
	3/21	~~~~~~ **Midterm Exam** ~~~~~~		
14	3/28	• IRT Models for Polytomous Outcomes	ER: Ch 5 (pp. 95–118)	HW 4
15	4/4	• IRT Models for Polytomous Outcomes (cont'd)		
	4/11	~~~~~~ AERA (no class) ~~~~~~		
16	4/18	• Validity • Open Topic	AERA, APA, & NCME (2014)	HW 5
	4/25	• **Student Presentations**		
	5/2	~~~~~~ **Final Exam** ~~~~~~		

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*(1), 50-64.

* (maybe) 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 and Psychological Measurement*, *73*(3), 532-547.