

**Course Syllabus**  
**16:300:695:01**  
**Topics in Ed Psych: Test Equating and Differential Item Functioning**  
**Fall 2013**

**Instructor:** Youngsuk Suh

**Office:** Room 323, Graduate School of Education

**Email:** [youngsuk.suh@gse.rutgers.edu](mailto:youngsuk.suh@gse.rutgers.edu)

**Phone:** (732) 932-7496 Ext. 8303

**Office hours:** Monday 2:30 - 3:30 PM, or by appointment

**Time:** Tuesday, 1:10 – 3:50 PM

**Place:** Room 347, Graduate School of Education

**Prerequisite:** Psychometric Theory I, Item Response Theory

**Course Description**

This course is designed to provide an overview of basic theories and methods in test equating and differential item functioning (DIF), which are two important applications of psychometric theories in testing practices. The course topics include fundamental concepts, models, techniques, topics, and issues in test equating and DIF.

**Course Goals**

After successfully completing this class students should

1. understand the fundamental concepts and theories of test equating and DIF;
2. be acquainted with the methods and techniques involved in conducting test equating and DIF;
3. have a familiarity with the issues and topics of research in test equating and DIF.

**Course Requirements**

1. ***Class Participation*** (30% of the Final Grade)

Students are expected to actively contribute to the class discussion, and will be graded based on their participation. Comments and questions in the class discussion should be based on the assigned reading materials. In-class and homework exercises will also be discussed.

Attendances at the meetings are expected during the semester. Please bring any necessary planned absences to my attention ahead of time.

2. **Project** (40% of the Final Grade)

An individual class project that illustrates the theories and methods of test equating or differential item functioning will be required. The project should be written up, and should be of conference or journal quality. In this project, students can choose to (1) do purely theoretical work, (2) conduct a simulation study, or (3) analyze real data. The project is due on the last day of the class.

3. **Project Presentation** (30% of the Final Grade)

Two meetings at the end of the semester will be devoted to project presentation. Students will be given approximately 30 minutes to give a presentation based on their project. The presentation will be graded based on its clarity, and the student’s ability to respond to project-related questions.

**Grading System**

Final letter grade will be assigned as follows:

Final Score	Letter Grade
90% and Above	A
80%-89.99%	B+
75%-79.99%	B
65%-74.99%	C+
60%-64.99%	C
55%-59.99%	D
Below 55%	F

**Class Schedule**

The class schedule below is subject to change if necessary. For the topic/s to be covered each week, there are reading assignments. Reading assignments must be completed prior to each class. Assigned readings for each week will be available on Sakai website at <http://sakai.rutgers.edu> under **16:300:695:01 F13 Equating and DIF**.

Date	Topic and Assigned Readings
Sep. 3	<i>Introduction and Concepts in Equating</i> <sup>1</sup> Kolen & Brennan (2004) - Chapter 1
Sep. 10	<i>Traditional Equating Methods for Random Groups and Nonequivalent Groups</i> <sup>1</sup> Kolen (1988); Kolen & Brennan (2004) - Chapters 2 and 4 <i>R lab preparation:</i> <sup>2</sup> Albano (2013)
Sep. 17	<i>R lab for traditional equating methods:</i> <sup>1</sup> Albano (2011)

Sep. 24	<i>IRT Equating Methods</i> <sup>1</sup> Cook & Eignor (1991); <sup>2</sup> Kolen & Brennan (2004) - Chapter 6 <i>R lab preparation: Battauz (2013)</i>
Oct. 1	<i>R lab for IRT equating methods; EQUATE (Baker [1993])</i> <i>Equating research</i> <sup>1</sup> <b><u>Hanson &amp; Beguin (2002)</u></b> ; <sup>3</sup> Kim (2006); <sup>2</sup> <b><u>Pang et al. (2010)</u></b>
Oct. 8	<i>Scaling, Score Scales</i> <sup>2</sup> Kolen & Brennan (2004) - Chapter 9; <sup>1</sup> Tong & Kolen (2010) <i>R lab preparation: Magis et al. (2010); Magis et al. (2013)</i>
Oct. 15	<i>Introduction and Concepts in Differential Item Functioning</i> Millsap & Everson (1993); <sup>1</sup> Zumbo (2007) <i>Observed-Score Methods : MH and Standardization</i> <sup>1</sup> Dorans (1989); <sup>2</sup> <b><u>Zwick (2012)</u></b> <i>R lab for MH and standardization</i>
Oct. 22	<i>Observed-Score Methods: Logistic Regression and SIBTEST</i> <sup>3</sup> Mazor et al. (1994); <sup>2</sup> Monahan et al. (2007); <sup>1</sup> <b><u>Swaminathan &amp; Rogers (1990)</u></b> <i>R lab for LR and Breslow-Day test</i>
Oct. 29	<i>IRT-based Methods: Likelihood Ratio Test and Lord's test</i> <sup>1</sup> Thissen et al. (1988); <sup>2</sup> <b><u>Woods et al. (2013)</u></b> <i>R lab for LRT and Lord's test; IRTLRDIF (<sup>3</sup>Thissen [2001])</i>
Nov. 5	<i>IRT-based Methods: Raju's area measures, DFIT</i> <sup>3</sup> Oshima & Morris (2008); Raju (1988; <sup>1</sup> 1990); <sup>2</sup> <b><u>Raju et al. (1995)</u></b> <i>R lab for Raju's area measures</i>
Nov. 12	<i>DIF in Polytomous Items: Log-linear, Standardization, Odds-Ratio, and LRT</i> <sup>2</sup> <b><u>Dorans et al. (1992)</u></b> ; <sup>1</sup> Green et al. (1989); <sup>3</sup> <b><u>Penfield (2008)</u></b> ; Potenza & Dorans (1995); Suh & Bolt (2011)
Nov. 19	<i>DIF in Multidimensional IRT: DFIT, Lord's test and Area measures</i> <sup>3</sup> <b><u>Bolt &amp; Johnson (2009)</u></b> ; <sup>1</sup> Oshima et al. (1997); <sup>2</sup> <b><u>Yao &amp; Li (2010)</u></b> <i>DIF review</i> <sup>4</sup> Clauser & Mazor (1998)
Nov. 26	Thanksgiving (No Class)

Dec. 3	<b>Project Presentation 1</b>
Dec. 10	<b>Project Presentation 2</b>
Dec. 17	<b>Final Project</b>

### **Complete References**

Albano, A. (2011). Statistical Equating Methods. Available from

<http://cran.r-project.org/web/packages/equate/vignettes/equatevignette.pdf>

Albano, A. (2013). <http://cran.r-project.org/web/packages/equate>

Baker, F. B. (1993). Equate 2.0: A computer program for the characteristic curve method of IRT equating. *Applied Psychological Measurement*, 17, 20.

Battauz, M. (2013). <http://cran.r-project.org/web/packages/equateIRT>

**Bolt, D. M. & Johnson, T. R. (2009)**. Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, 33, 335-352.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 37-45.

Cook, L. L. & Eignor, D. R. (1991). IRT equating methods. *Educational Measurement: Issues and Practice*, 10(3), 37-45.

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and Mantel-Haenszel Method. *Applied Measurement in Education*, 2, 217-233.

**Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1992)**. The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29, 309-319.

**Green, B. F., Crone, C. R., & Folk, V. G. (1989)**. A method for studying differential distractor functioning. *Journal of Educational Measurement*, 26, 147-160.

**Hanson, B. A. & Beguin, A. A. (2002)**. Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26, 3-24.

Kim, J.-S. (2006). Using the distractor categories of multiple-choice items to improve IRT linking. *Journal of Educational Measurement*, 43, 193-213.

Kolen, M. J. (1988). Traditional equating methodology. *Educational Measurement: Issues and*

*Practice*, 7(4), 29-36.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*. New York, NY: Springer.

Magis, D., Beland, S., & Raiche, G. (2013). <http://cran.r-project.org/web/packages/difR>

Magis, D., Beland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42, 847-862.

Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement*, 54, 284-291.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17, 297-334.

Monahan, P. O., McHorney, C. A., Stump, T. E., & Perkins, A. J. (2007). Odds Ratio, delta, ETS classification, and standardization measures of DIF magnitude for binary logistic regression. *Journal of Educational and Behavioral Statistics*, 32, 92-109.

Oshima, T. C., & Morris, S. (2008). Raju's differential functioning of items and tests (DFIT). *Educational Measurement: Issues and Practice*, 27(3), 43-50.

Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential functioning of items and tests. *Journal of Educational Measurement*, 34, 253-272.

**Pang, X., Madera, E., Radwan, N., & Zhang, S. (2010).** A comparison of four testing equating methods. Education Quality and Accountability Office Research. Available from [http://www.eqao.com/Research/pdf/E/Equating\\_Crp\\_cftem\\_ne\\_0410.pdf](http://www.eqao.com/Research/pdf/E/Equating_Crp_cftem_ne_0410.pdf)

Penfield, R. D. (2008). An odds ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement*, 45, 247-269.

Potenza, M. T., & Dorans, N. J. (1995). DIF assessment for polytomous scored items: A framework for classification and evaluation. *Applied Psychological Measurement*, 19, 23-37.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 495-502.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14, 197-207.

**Raju, N. S., van der Linden, W. J., & Fler, P. F. (1995).** IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement*, 19, 353-

368.

Suh, Y., & Bolt, D. M. (2011). A Nested logit approach for investigating distractors as causes of differential item functioning. *Journal of Educational Measurement, 48*, 188-205.

**Swaminathan, H., & Rogers, H. J. (1990).** Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370.

Thissen, D. (2001). IRTLRF v2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning [Computer program]. Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 147-169). Hillsdale NJ: Erlbaum.

Tong, Y., & Kolen, M. J. (2010). Scaling: An ITEMS Module, *Educational Measurement: Issues and Practice, 29*(4), 39-48.

**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*, 532-547.

**Yao, L., & Li, F. (2010).** A DIF detection procedure in multidimensional item response theory framework and its applications. Paper presented at the 2010 annual meeting of the National Council on Measurement in Education, Colorado, Denver.

Zumbo, B. D. (2007). Three generations of DIF analyses: considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*, 223-233.

**Zwick, R. (2012).** A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report No. RR-12-08*.

### **Policy on Academic Integrity**

Please refer to the Policy on Academic Integrity for Undergraduate and Graduate Students at <http://academicintegrity.rutgers.edu>. Clear evidence of violation of academic integrity policy may result in a grade of F for the course.