

S-061A2: Statistical and Psychometric Methods for Educational Measurement
Harvard Graduate School of Education
October 19 – December 14, 2016

Class meets Mondays and Wednesdays, 10:10-12:00, in Longfellow 229

Course website (same as S-061A1): <https://canvas.harvard.edu/courses/18969/>

Instructor: Andrew Ho

455 Gutman Library

Andrew_Ho@gse.harvard.edu

TF: Maria (Masha) Bertling

mbertling@g.harvard.edu

Instructor Office Hours: After class or email Wendy_Angus@gse.harvard.edu for appointments

Description

This is the second of two sequential modules on quantitative methods for educational measurement. Students will continue their training in psychometric and statistical methods of measurement, with greater emphasis on understanding and critiquing recent research, as well as the development of an individual research proposal that has promise for advancing the field. Training will continue in reliability, generalizability theory, validation, differential item functioning, item response theory, scaling, linking, standard setting, and adjustments for measurement error. Contexts of assessment include small-scale educational and psychological assessments for targeted research studies as well as large-scale district, state, and national assessments for formative, summative, and evaluative purposes. Students are strongly encouraged to bring their own score-level or item-level data although this is not required. In the preceding module, S-061A1, which is a prerequisite, students will have learned and applied methods in class and through completion of data analytic assignments. In this second module, students will complete assignments, participate in class discussions on current research in educational measurement, and develop their own research proposal, which they will present at the end of the semester.

Prerequisite: S-061A1 in the same semester or a previous version of S-061A2. This course complements S-043 and S-090, and students may enroll in these courses in any order. Students who do not meet the prerequisite may enroll instead in S-011, which provides a nontechnical introduction to educational measurement.

Grading

The requirements of the course include regular attendance (10%), active participation in class and in out-of-class discussions (25%), satisfactory completion of preliminary proposal (15%), and satisfactory completion and in-person presentation of a full research proposal or project (50%). These weights are approximate—the final course grade may factor in improvement over time and exemplary performance on one or more dimensions.

This course is letter-grade-only; students may not take this course on a Satisfactory/No Credit basis. Registered students must submit a course evaluation form at the end of the semester in order to fulfill the requirements of the course. In-person auditing of this course is not allowed—all attendees must be registered students.

Support

Our TF, Masha Bertling, will hold weekly office hours by appointment. Occasionally, Masha may also offer optional discussion sections as she sees necessary. Participation is strongly recommended. Office hours with the instructor are typically available after class meetings and are also available by appointment, Contact my assistant, at Wendy_Angus@gse.harvard.edu to schedule appointments. Occasional one-on-one check-ins with the instructor are strongly recommended.

October 2016						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
16	17 – A1 Class 12	18	19 – Class 1 (Review)	20 – A1 final out: 9AM	21 – A1 final due: 5PM	22
23	24 – Class 2	25	26 – Class 3	27	28 – 1-page proposal due	29

November 2016						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
30	31 – Class 4	1	2 – Class 5	3	4	5
6	7 – Class 6	8	9 – Class 7	10	11	12
13	14 – Class 8	15	16 – Class 9	17	18 – Prelim proposal due	19
20	21 – Class 10	22	23 – Holiday	24 – Thanksgiving	25	26
27	28 – Class 11	29	30 – Class 12	1	2- Full drafts due	3

December 2016						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
4	5 – Presentation Session 1	6 – Presentation Session 2	7	8	9	10
11	12	13	14 – Final Projects Due (5PM)	15	16	17

Course Project

To complete the course, you must develop an original research project, present it to other class members, and submit a final written product. In conducting this research project, you may collaborate with a partner in the class. While my vision of what constitutes a viable project is somewhat contextual and subject to negotiation, there are two broad possibilities:

1. A complete and extended outline of a research paper, including results from data analysis that uses the methods introduced in this course. I will provide more detail on what is meant by “complete and extended outline”, but in essence this will require you to write down the underlying structure of the paper from start to finish. This includes (in order): 1) Introduction / motivation; 2) Background / literature review; 3) Data and sample; 4) Methods; 5) Results; and 6) Discussion and implications for policy and/or practice. To be clear, while this is less effort than a completed final paper, it also means that you must produce the main tables and figures that you intend to include in a final version of your paper.
2. In the event that producing results is not possible, I will require you to write a “pre-analysis plan”. This is very similar to #1 above, except that the “Results” section will instead require a full list of models that you wish to estimate, a detailed discussion of what you expect to find, and what you might learn from your study (regardless of how the analysis turns out). This option is available primarily for students who are working on a project that has scholarly promise, but also logistical headaches that are beyond your control (for example, if you are collecting data or waiting on the acquisition of a restricted-use data set).

Your grade on the course project will depend on the contribution of the project is and how well you implement it. In addition to the scheduled deadlines on the syllabus, I encourage you to keep me informed on the progress of your project throughout the semester.

In order to ensure that you make an early start and continue to make progress on your project, you must submit an initial one-page, single-spaced overview of your proposed project, or a short list of concise research possibilities to me by 5PM on Friday, October 28, a draft by 5PM, November 18, make a presentation of your project to me and other class members on December 5 or 6, and submit a final version by 5PM, Wednesday, December 14, that reflects your responses to the feedback that we provide on your initial draft. The maximum length of the extended outline or pre-analysis plan (not counting references, tables, and figures) is 20 pages of double-spaced 12 point type with one-inch margins. During the course, we will provide explicit guidance and support tailored to each research project. I expect that your course project will eventually (after the course is complete) result in a published research paper that can appear on your Curriculum Vitae.

To help you prepare the presentation of your course project, we will devote class time to discussing the components of a good scholarly talk. We may also be available to help you practice a “dry run” of your presentation. You (and perhaps your research partner) should make

an appointment with me to discuss the feedback I provide on your draft. You should also see Masha or me for ongoing support for your project. Prior to any meetings with either of us, please prepare a brief memo describing your agenda for the meeting and email it the day before our meeting. This will help us to help you.

Readings –
THESE ARE SUBJECT TO CHANGE
IN RESPONSE TO STUDENT RESPONSE FROM STUDENTS IN S-061A1

I require students to respond to online, Google Doc, discussion questions to central readings by 10PM the night before each class. A good response will demonstrate that the student has read and carefully considered the central reading and spent time considering what the discussion question is asking. Central readings that are very technical need not be mastered in detail, but do pay attention to notation and the underlying motivation of derivations. I intend noncentral readings as additional context and citations for future reference. Links will work on campus and, if you are off campus, if you have a VPN connection. Other readings are available via the [iPa©](#) tab on the Canvas website

There are two required textbooks that are available for purchase:

- 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. I recommend purchasing this via [APA/AERA/NCME](#), where members can get a discount, and you'll also get an electronic text (very useful). The Coop is also an option.
- Koretz, D. (2009). *Measuring up: What educational testing really tells us*. Cambridge: Harvard University Press. See [Amazon](#) or the Coop.

1. Reliability and validation of large-scale assessments

AERA/APA/NCME Standards, Chapters 1, 2, 12, & 13. (required text)

MCAS Technical Report: [link](#)

California (CST) Tech Report,

<http://www.cde.ca.gov/ta/tg/sr/documents/cst14techrpt.pdf>

2. Reliability and validation of psychological and sociological instruments

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-S). *Journal of Personality Assessment*, 91(2), 166-174. [link](#)

National Center for Education Statistics. (2012). *Improving the measurement of socioeconomic status for the national assessment of educational progress: A theoretical foundation--recommendations to the national center for education statistics*

https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf

3. Generalizability Analyses

- Ho, A.D. & Kane, T.J. (2013). *The reliability of classroom observations by school personnel*. Bill & Melinda Gates Foundation. [link](#)
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56-64. [link](#)
- Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, 57(3), 373-399. [link](#)

4. Item Response Theory

- Mislevy, R. J., Johnson, E. G., & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics*, 17(2), Special Issue: National Assessment of Educational Progress), 131-154. [link](#)
- Lockwood, J. R., & Castellano, K. E. (2015). Alternative statistical frameworks for student growth percentile estimation. *Statistics and Public Policy*, 2(1), e962718-1-10. [link](#)
- Ho, A. D., & Yu, C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, 75(3), 365-388. [link](#)

5. Scaling and Value-Added Models

- Briggs, D. C., & Domingue, B. (2013). The gains from vertical scaling. *Journal of Educational and Behavioral Statistics*, 38(6), 551-576. [link](#)
- Castellano, K. E., & Ho, A. D. (2015). Practical differences among aggregate-level conditional status metrics: From median student growth percentiles to value-added models. *Journal of Educational and Behavioral Statistics*, 40(1), 35-68. [link](#)

6. Standard Setting

- McClarty, K., Way, W., Porter, A., Beimers, J., & Miles, J. (2013). Evidence-based standard setting: Establishing a validity framework for cut scores. *Educational Researcher*, 42(2), 78-88. [link](#)
- Haertel, E. H. (2002). Standard setting as a participatory process: Implications for validation of Standards-Based accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 16-22. [link](#)
- Ho, Andrew. (2012). Off track: Problems with “on track” inferences in empirical and predictive standard setting. Working Paper. Harvard Graduate School of Education. [link](#)

7. Bias, Differential Item Functioning, and Accommodations

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment*, 14(3-4), 160-179. [link](#)

Hamilton, L. S. (1999). Detecting gender- based differential item functioning on a constructed- response science test. *Applied Measurement in Education*, 12(3), 211-235. [link](#)

Varni, J. W., Thissen, D., Stucky, B. D., Liu, Y., Magnus, B., Quinn, H., et al. (2014). PROMIS® parent proxy report scales for children ages 5-7 years: An item response theory analysis of differential item functioning across age groups. *Quality of Life Research*, 23(1), 349-361. [link](#)

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English Language Learners: Implications for policy-based empirical research. *Review of Educational Research*, 74, 1-28. [link](#)

8. “Scale-invariant,” Ordinal, Nonparametric Distributional Recovery

Ho, A. D. (2009). A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics*, 34(2), 201-228. [link](#)

Quinn, D. M. (2015). Black-white summer learning gaps: Interpreting the variability of estimates across representations. *Educational Evaluation and Policy Analysis*, 37(1), 50-69. [link](#)

Reardon, S. F., Shear, B. R., Castellano, K. E., & Ho, A. D. (in press). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. *Journal of Educational and Behavioral Statistics*. <https://cepa.stanford.edu/content/using-heteroskedastic-ordered-probit-models-recover-moments-continuous-test-score-distributions-coarsened-data>

Statistical and psychometric computing

Statistical computing is an integral part of S-061. I will be using Stata this year, and you will require Stata 14 to use Item Response Theory methods.

I do not teach programming during class time, although code is threaded through the lecture slides. We provide resources to help you learn how to program on your own at your own pace. John may also cover coding issues in their sections.

There are two ways you can access Stata. The least expensive option is to use one of the networked workstations available in the Learning Technology Center (LTC) on the 3rd floor of Gutman Library and elsewhere on the HGSE campus (e.g., on the 2nd and 4th floors of Gutman Library). For students who would like to use Stata on their own PCs, you may purchase Stata following this link: <http://www.stata.com/order/new/edu/gradplans/student-pricing/>. Stata/IC, which will be sufficient for this course, is available for \$75 for a 6-month license and \$198 for a perpetual license.

Collaboration and study groups

When collaborating across groups, **students must turn in work as pairs or individuals where specified above, not group work. Papers should be written in the pair's own words—your text should reflect your own understanding of the material.**

Accommodations

Students needing accommodations in instruction or testing must notify the instructor early in the semester, and HGSE's policies must be followed. Late requests for accommodations will not be honored unless there is a pressing reason, such as a recent injury.

A Note on Plagiarism

Please read the School's policy on plagiarism in the HGSE Student Handbook, which includes the statement, "Students who submit work either not their own or without clear attribution to the original source, for whatever reason, ordinarily will be dismissed from the Harvard Graduate School of Education." Attention to this policy is particularly important in a course like S-061, in which collaboration with other students is often required and generally encouraged. If you work closely with other students or partnerships—a process that I encourage and fully support—recognize the other students' contributions explicitly in your written account (a footnote is fine for this purpose). This helps avoid the natural questions that arise when similarities are detected at grading. **If you have any questions about what constitutes appropriate collaboration, or how to define what constitutes your own work, please see me or a Teaching Fellow.**

I cannot overemphasize the need for all students to monitor their own behavior. Assignments are structured such that you can receive feedback on your and your partner's understanding of the material. The consequences for plagiarism are appropriately severe.

Other Writing Resources

- HGSE Academic Writing Services: Gutman Library
- APA Online Tutorial: http://isites.harvard.edu/icb/icb.do?keyword=apa_exposed
- Writing Resources (including *Writing Like an Educator* Course and Reference Materials): <http://isites.harvard.edu/icb/icb.do?keyword=awrs&pageid=icb.page48297>
- Sign-up for Individual Sessions at the Writing Center: <http://www.appointmentquest.com/provider/2030159020>