S-061: Statistical and Psychometric Methods for Educational Measurement
Syllabus, Harvard Graduate School of Education – January 3-13, 2023

Class meets every weekday, January 3-13, from 10am-12pm
Course website: https://canvas.harvard.edu/courses/113091/
Instructor: Andrew Ho, Andrew_Ho@gse.harvard.edu
Teaching Fellows: Jeannette Garcia Coppersmith and Jonathan Seiden,
jgarciacoppersmith@g.harvard.edu, jseiden@g.harvard.edu
Drop-in times with Andrew: To be scheduled based on availability.
Drop-in times with Jonathan and Jeannette: To be scheduled based on availability.

Description

This 2-unit module covers statistical and psychometric methods for improving the design and use of educational and psychological measurements. Students will learn selected classical and modern measurement techniques, including reliability, validation, item response theory, test linking and equating, and adjustments for measurement error. Contexts of assessments include small-scale educational and psychological measures for targeted research studies as well as large-scale district, state, and national assessments for formative, summative, and evaluative purposes. Students have the option to develop a research proposal that has promise for advancing the field.

Prerequisite: S-052 or at least two semesters of applied statistics that includes estimation and interpretation of logistic regression coefficients. Experience with multilevel models is encouraged but not required. 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 synchronous engagement (10%), active and effortful participation in asynchronous activities ("learning checks" and "learning submissions," 25%), satisfactory completion of 3 assignments (35%), an either an in-person "conversational celebration of learning" (CCL, 30%), or satisfactory completion of a full research outline or proposal (30%). These weights are approximate—the final course grade may factor in improvement over time and exemplary performance on one or more dimensions. Students must complete the three assignments in partnerships. The CCL or project is due on Friday, January 13.

You may take this course either for a letter-grade or on a Satisfactory/No Credit basis. Registered students must submit a course evaluation form at the end of the semester to fulfill the requirements of the course. Auditors may not attend this course synchronously—all synchronous attendees must be registered students.

Additional Meetings with Andrew, Jeannette, and Jonathan

Jeannette, Jonathan, and I will be available daily after class meetings at times to be determined.

The Daily Rhythm of S-061A

S-061A has a daily rhythm as follows:

- At night or from 8am-10am in the morning, please watch the 60-90 minutes of "asynchronous modules" and complete the "learning checks," short questions requiring brief answers. Solutions are available immediately thereafter.
- 10am-11am: The first half of in-person class will be an "emphasis session," a focused review of the asynchronous materials with time for questions. There will be a short break from 11am-11:05am.
- 11:05am-12:00pm The second half of in-person class will be a "review session," a relatively more relaxed session with more time for interaction and questions in preparation for the day's assignment.
- In the afternoon, students will complete the day's "learning submission" or "assignment." Learning submissions are moderate and ungraded, whereas assignments are more extended. Learning submission solutions are available immediately upon submission, whereas assignment feedback is provided on the following day. Learning submissions must be submitted individually, although collaboration with any fellow student is welcome. Assignments must be submitted in pairs, and co-writing is only permitted within the partnership. Daily submissions are due at 9pm daily.
- After submitting the daily submission, students may watch the asynchronous modules or save them for the morning, when the daily rhythm begins again.

Timing and Purpose of S-061 Assessments

Assessment Type	Length	Completion Time	Collaboration	Feedback	Purpose
Learning Checks	Brief	Between asynchronous modules	Collaboration with anyone is encouraged. Ultimately, individual submission.	Correct answers provided immediately.	Formative
Learning Submissions	Moderate	Due Daily on Day 2, 4, 6, and 8.	Collaboration with anyone is encouraged. Ultimately, individual submission.	Correct answers provided immediately.	Formative
Assignments	Extended	Due Daily on Day 3, 5, and 7	Collaboration only with partners. Partnered submission required.	More detailed, provided the next day.	Formative and summative.

REFERENCES

I provide central references as resources and supplemental references as placeholders for additional context and citations for future reference. Links will work on campus and, if you are off campus, if you have a VPN connection. Most readings (LR tag) are available via the library reserves tab here (by day #): https://canvas.harvard.edu/courses/113091/external_tools/33439

There are two required texts that are both available online, free of charge.:

1. 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. The standards are publicly available here: https://www.testingstandards.net/open-access-files.html

If you wish to purchase a paper text, you can visit APA/AERA/NCME, where AERA, NCME, or APA members can get a discount ($49.95). The Coop is also an option.

2. Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge: Harvard University Press. Available online via Library Reserves. For a paper text, visit Amazon or the Coop.

Day 1 (Tuesday, 1/ 3): Measurement and Validation

Central:

1) Koretz (2008), Chapter 2, pp. 16-34. (required text (LR)
2) AERA/APA/NCME Standards, Chapter 1, pp. 11-31. (required text) (LR)
3) Lyons, S., Johnson, M., & Hinds, B. F. (2021). *A Call to Action: Confronting Inequity in Assessment*. Lyons Assessment Consulting. (link) (LR)

Supplemental:

4) Koretz (2008), Chapter 9, pp. 215-234. (required text) (LR)
5) Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. (LR)
6) Randall, J. (2021). "Color-neutral" is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*. https://doi.org/10.1111/emip.12429 (LR)

Day 2 (Wednesday, 1/4): Reliability and Classical Test Theory (LS)

Central:

1) AERA/APA/NCME Standards, Chapter 2, pp. 33-47. (required text) (LR)
2) 2017 MCAS Technical Report Section 3.7: (link) (LR)
3) Traub, R. E., & Rowley, G. L. (1991). Understanding reliability. *Educational Measurement: Issues and Practice, 10*(1), 37-45. (LR)

Supplemental:

4) Koretz (2008), Chapter 7, pp. 143-178. (required text) (LR)
5) Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth Group. Chapter 6: Reliability and the classical true score model pp. 105-113, 122-124; Chapter 7: Procedures for estimating reliability pp. 131-150. (LR)
6) Haertel, E. (2006). In R. Brennan (Ed.) *Educational measurement*. Westport, CT: Praeger. Chapter 3, 65-87. (LR)

Day 3 (Thursday, 1/5): Introduction to Item Response Theory

Central:

1) Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice, 8*(1), 35-41. (LR)
2) 2017 MCAS Technical Report Section 3.6: link (LR)
3) Stata Manual (version 17), pp. 1-12. http://www.stata.com/manuals/irt.pdf (LR)

Supplemental:

4) Stevens, S. S. (1946). On the theory of scales of measurement. *Science (New York, N.Y.), 103*(2684), 677-680. (LR)
5) AERA/APA/NCME Standards, Chapter 5, pp. 95-109, focus on scores, scales, and norms. (required text) (LR)

Day 4 (Friday, 1/6): Item Response Theory

Central:
1) Yen, W. M., & Fitzpatrick, A. R. (2006). In R. Brennan (Ed.), *Educational measurement*. Westport, CT: Praeger. pp. 111-118. Sections 1-2.4 (LR))
2) Stata Manual (version 17), pp. 28-58. http://www.stata.com/manuals/irt.pdf (LR)

Supplemental:
3) 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. (LR)

Day 5 (Monday, 1/9): Practical Validation Using Polytomous IRT

Central:
1) Stata Manual (version 17), pp. 12-18; 77-92. http://www.stata.com/manuals/irt.pdf (LR)
2) Jessen, A., Ho, A. D., Corrales, C. E., Yueh, B., & Shin, J. J. (2018). Improving measurement efficiency of the Inner EAR scale with Item Response Theory. *Otolaryngology-Head and Neck Surgery, 158*, 1093-1100. (LR)

Supplemental:
3) Samejima, F. (1969). *Estimation of Latent Ability Using a Response Pattern of Graded Scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved from https://www.psychometricsociety.org/sites/main/files/file-attachments/mn17.pdf?1576606975) (LR)

Day 6 (Tuesday, 1/10): Test Linking and Equating: Using IRT for Fair Comparisons

Central:
AERA/APA/NCME Standards, Chapter 5, pp. 95-109, focus on linking. (required text) (LR)

Supplemental:
1) Reardon, S. F., Kalogrides, D., & Ho, A. D. (2021). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics*. (LR)
2) Holland, P. W., & Dorans, N. J. (2006). In R. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: American Council on Education, Praeger Publishers. pp. 197-201. (LR)

Day 7 (Wednesday, 1/11): Improving Use of Sociodemographic Variables, Bias, Differential Item Functioning

Central:

1) AERA/APA/NCME Standards, Chapter 3, pp. 49-72. (required text) (LR)
2) Stata Manual (version 17), pp. 234-247. http://www.stata.com/manuals/irt.pdf (LR)

Supplemental:

3) Koretz (2008), Chapter 11, pp. 260-280. (required text) (LR)
4) 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. (LR)

Day 8 (Thursday, 1/12): Measurement Methods for Educational Accountability

Central:

1) AERA/APA/NCME Standards, Chapter 5. (required text) (LR)
2) Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under no child left behind. *Educational Researcher*, *37*(6), 351-360. (LR)
3) Kolen, M. J. (2006). In R. Brennan (Ed.), Educational measurement (4th ed.). Westport, CT: Praeger Publishers. pp. 171-180. (LR)

Supplemental:

4) Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement*, *15*(4), 237-261.(LR)
5) McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-based standard setting. *Educational Researcher*, *42*(2), 78-88. (LR)
6) Koretz (2008), Chapter 8. (required text) (LR)
7) Briggs, D. C. (2013). Measuring growth with vertical scales. *Journal of Educational Measurement, 50*(2), 204-226. (LR)
8) Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models*. Council of Chief State School Officers. (LR)
9) Pearl, J. (2014). Lord's paradox revisited - (Oh Lord! Kumbaya!). University of California. Technical Report R-36 (October). (LR)
10) Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*, 241-253. (LR)

Day 9 (Friday, 1/13): Review for Project Proposals and the Conversational Celebration of Learning

Option A: Conversational Celebration of Learning (CCL)

I will hold a 20-minute conversational "celebration of learning" with each of you who choose this option on Friday, January 13, at time windows to be scheduled. The goal is to encourage you to become fluent enough in the language of measurement and psychometrics that you can speak it actively and accurately in conversation. Questions will be based on the assignments you will have already completed in class. You can hear a bit about the reasoning behind this conversation here: https://vimeo.com/497738911

Option B: Course Project

An alternative to the CCL, for those who have data, is to outline an original research project and submit a final written product. In developing 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. An 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 "extended outline", but 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 outline the main tables and figures that you intend to include in a final version of your paper.

2. If producing results is not possible, you may 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. The maximum length of the extended outline or pre-analysis plan (not counting references, tables, and figures) is 10 pages of double-spaced 12-point type with one-inch margins. I hope that your course project could eventually (after the course is complete) lead to a published research paper that can appear on your Curriculum Vitae.

If you choose this option, please try to develop the following by the end of the first week:

- Proposed title
- Key references (2-5)
- Research Questions
- Data & sketch of analytic plan (include "8 displays" from Day 2 if data are available)
- Questions/roadblocks

As we evaluate your project, we will ask the following five questions:

1) Have you asked at least one compelling research question motivated and framed by existing literature?
2) Have you demonstrated that you can answer at least one interesting research question in a cogent and thoughtful manner using real or simulated data?
3) Have you articulated your methods clearly enough that readers can replicate them?
4) Have you demonstrated mastery of a measurement concept or technique, including accurate interpretation of analytic results?

5) Have you demonstrated sufficient momentum toward a publishable paper?

Statistical and psychometric computing

Statistical computing is an integral part of S-061. I will be using Stata this year. I assume that everyone is comfortable using a computer to perform basic statistical analysis, although I do not necessarily assume that you have used Stata. All Harvard students have free access to Stata.

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. Jeannette and Jonathan may also cover coding issues in their informal sessions.

I also allow students to complete problem sets and projects in R. For students with equal proficiency and interest in both programming languages, Stata will be the more convenient choice. However, students complete the course in R every year, and we will do our best to support this additional language.

Slack, collaboration, and study groups

We will be using Slack for collaboration and communication in this course. Please access the S-061 Slack channel on the Canvas site here: https://canvas.harvard.edu/courses/113091/external_tools/77742. Standard Slack norms are here:

- Be good citizens of Slack: answer questions, assume best intent, try not to sidetrack conversations.
- Understand public and private channels, and try to communicate with the proper audience.
- Use threads (reply) to organize smaller group discussions around specific topics.
- Format messages so that they're easier to read (e.g., shift-return creates a line break).
- **Use emoji reactions** liberally to acknowledge messages while keeping channels from becoming inundated with reaction messages.
- Use the search function to try to find an answer, reducing duplicate questions.
- Manage **notifications** to reduce information overload and maintain focus.
- Remember that there are situations when it's best to move the conversation off of Slack messaging. Students can meet face-to-face using Zoom or **Slack calls.**

Many people learn in deeper ways when working in a group, and I encourage collaborative learning. To mimic statistical and psychometric work in the real world and to provide a chance for you to use this language actively, I require completion of assignments in pairs starting from Day 3.

We mandate collaboration for at least three reasons. First, learning statistical and psychometric methods is like learning a language. To learn it, one must "speak" it actively and in a genuine context with other individuals. Second, collaborative quantitative analysis is the norm and individual work is the exception in the world of practice. Third, my experience has been that, on average, students who work in pairs and groups both perform better and enjoy themselves more

than students who work individually. Statistical and psychometric collaboration is a case where the whole is greater than the sum of its parts.

Accommodations

We encourage students needing accommodations in instruction or evaluation to notify us early in the semester. If you have a disability or health concern that may have some impact on your work in this class and for which you may require adjustments or accommodations, please contact the Access and Disability Services (ADS) administrator in Gutman 124. No accommodations can be given without authorization from ADS, or without notice. If you already have a Faculty Contact Form for this course from ADS, please provide us with that information privately in our offices so that we can make those adjustments in a timely manner. All inquiries and discussions about accommodations will remain confidential.

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, Jeannette, or Jonathan.**

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