

S-061: Statistical and Psychometric Methods for Educational Measurement Syllabus, Harvard Graduate School of Education – Fall, 2020

Class meets Mondays and Wednesdays, 10:30-11:45 Eastern Time on Zoom

Course website: <https://canvas.harvard.edu/courses/80499>

Instructor: Andrew Ho

Andrew_Ho@gse.harvard.edu

Teaching Fellows: Sophie Barnes and Emma Klugman

spb702@g.harvard.edu emma.m.klugman@gmail.com

Weekly sections with Andrew: Alternating weeks. To be scheduled.

Weekly sections with Sophie and Emma: As necessary. To be scheduled.

“Please-Drop-By-My-Office” Hours: Wednesdays, 4pm ET, on Zoom.

Description

This course covers statistical and psychometric methods for improving the design and use of educational and psychological measurements. Students will learn classical and modern measurement techniques, including reliability, generalizability theory, validation, differential item functioning, item response theory, scaling, linking, standard setting, 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. In the first two thirds of the course, students will learn and apply methods in class and complete data analytic assignments. In the remainder of the course, students will read and critique recent research in educational measurement. Students will also develop and present 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 where time zones permit (5%), active and effortful participation in synchronous and asynchronous activities (15%), satisfactory completion of 4 assignments (35%), an in-person “conversational celebration” (15%), and satisfactory completion and in-person presentation of a full research outline/project (30%). These weights are approximate—the final course grade may factor in improvement over time and exemplary performance on one or more dimensions. Students may complete the first assignment individually or in pairs and the remaining three assignments in pairs. The “conversational celebration” will be held via Zoom with me and Sophie and Emma.

This course is letter-grade-only. If students would strongly prefer to take this course on a Satisfactory/No Credit basis, I am happy to discuss this possibility with them on a case-by-case 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.

Support

Our TFs, Sophie and Emma, will hold weekly office hours by appointment. Occasionally, they may also offer optional discussion sections as they deem necessary; these may be most common around assignment due dates. Attendance at these sessions is strongly recommended. Meetings with me are typically available after class meetings and are also available by appointment. Contact me or my assistant, Wendy_Angus@gse.harvard.edu, to schedule appointments. Occasional one-on-one check-ins with me are welcome and strongly recommended. To facilitate this, I also hold, “please drop by my office” hours on Wednesdays at 4pm. No pre-set agenda or questions are necessary. Just drop by!

CALENDAR

(I expect this calendar to change slightly as I accommodate the pace and preferences of the class)

September 2020						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
30	31	1	2 – Class 1 Validation	3	4 – Class 2 Reliability	5
6	7 – HOLIDAY	8	9 – Class 3 CTest Theory	10 – Asgn 1 Released	11	12
13	14 – Class 4 Int. G Theory	15	16 – Class 5 G Theory	17	18 – Asgn 1 Due 5PM	19
20	21 – Class 6 GTheory	22	23 – Class 7 GTheory Ext.	24 – Asgn 2 Released	25	26
27	28 – Class 8 IRT Intro	29	30 – Class 9 IRT	1	2 – Asgn 2 Due 5PM	3

October 2020						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
4	5 – Class 10 IRT Infor	6	7 – Class 11 IRT Scoring	8 – Asgn 3 Released	9	10
11	12 – HOLIDAY	13	14 – Class 12 Poly IRT	15	16 – Asgn 3 Due 5PM	17
18	19 – Class 13 Linking	20	21 – Class 14 Bias/DIF	22 – Asgn 4 Released	23	24
25	26 – Class 15 Stand.Setting	27	28 – Class 16 Growth/Acc	29	30 – Asgn 4 Due 5PM	31

November 2020						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1	2 – Class 17 Test Design	3	4 – Class 18 Frontiers	5	6	7
8	9 – Class 19 Review	10	11 - HOLIDAY	12 – Convrs. Celebration!	13 – Convrs. Celebration!	14
15	16 – Class 20 PrctclValidtn	17	18 – Class 21 Inflation	19	20 – Prelim proposal due	21
22	23 – Class 22 Effect Sizes	24	25 – HOLIDAY	26 – Thanksgiving	27	28

*** I may adapt topics and readings for the last 4 classes to relate to student projects.

December 2020						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
29	30 – Class 23 SEDA	1	2 – Class 24 NAEP	3	4	5
6	7 – Presentations	8	9 – Presentations	10	11	12
13	14	15	16 – Projects Due (5PM)	17	18	19

READINGS and MODULES

I will post online modules before synchronous classes. I expect students to watch these modules and skim central readings before each class. To stimulate asynchronous discussion, I may require students to respond to online discussion questions to central readings once a week, by 10PM on Tuesdays between classes. A good response will demonstrate that the student has watched the module or engaged with the central reading. I also encourage students to spend time considering carefully what the discussion question is asking.

Central readings that are very technical need not be mastered in detail, but please 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 library reserves tab here: https://canvas.harvard.edu/courses/80499/external_tools/33439

There are two required textbooks that are available for purchase:

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. I

recommend purchasing this via [APA/AERA/NCME](#), where members can get a discount (\$49.95), and you'll also get an electronic text (very useful). Note that graduate students can also sign up for NCME membership for free! 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 [Harvard Library](#). See also [Amazon](#) or the Coop.

September 2 – Class 1: Validation

Central reading:

- 1) Koretz (2008), Chapter 2, pp. 16-34. (required text)
- 2) AERA/APA/NCME Standards, Chapter 1, pp. 11-31. (required text)
- 3) 2017 MCAS Technical Report Section 2.4 and 3.9: ([link](#))

Supplemental reading:

- 4) Koretz (2008), Chapter 9, pp. 215-234. (required text)
- 5) Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. ([reserves](#))
- 6) Haertel, E. (2013). Getting the help we need. *Journal of Educational Measurement*, 50(1), 84-90. ([reserves](#))
- 7) Ho, A. (2013). The epidemiology of modern test score use: Anticipating aggregation, adjustment, and equating. *Measurement: Interdisciplinary Research and Perspectives*, 11(1-2), 64-67. ([reserves](#))

September 4 – Class 2: Reliability

Central reading:

- 1) AERA/APA/NCME Standards, Chapter 2, pp. 33-47. (required text)
- 2) 2017 MCAS Technical Report Section 3.7: ([link](#))

Supplemental reading:

- 3) Koretz (2008), Chapter 7, pp. 143-178. (required text)

September 9 – Class 3: Classical Test Theory

Central reading:

- 1) Traub, R. E., & Rowley, G. L. (1991). Understanding reliability. *Educational Measurement: Issues and Practice*, 10(1), 37-45. ([reserves](#))

Supplemental reading:

- 2) 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. ([reserves](#))
- 3) Haertel, E. (2006). In R. Brennan (Ed.) *Educational measurement*. Westport, CT: Praeger. Chapter 3, 65-87. ([reserves](#))

September 14 – Class 4: Introduction to Generalizability Theory

Central reading:

- 1) Shavelson, R.J. & Webb, N.W. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage Publications. 1-24. ([reserves](#))

Supplemental reading:

- 1) Brennan, R. L. (2001). Generalizability theory. New York: Springer-Verlag. 1-19. ([reserves](#))

September 16 – Class 5: Generalizability Theory

Central reading:

- 1) Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice*, 11(4), 27-34. ([reserves](#))

Supplemental reading:

- 2) Ho, A.D. & Kane, T.J. (2013). *The reliability of classroom observations by school personnel*. Bill & Melinda Gates Foundation. ([reserves](#))

September 21 & 23 – Class 6 & 7: Generalizability Theory Extensions

Central reading:

- 1) 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. ([reserves](#))

Supplemental reading:

- 2) 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. ([reserves](#))

September 28 & 30 – Class 8 & 9: Introduction to Item Response Theory

Central reading:

- 1) Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, 8(1), 35-41. ([reserves](#))
- 2) 2017 MCAS Technical Report Section 3.6: [link](#)
- 3) Stata Manual (version 16), pp. 1-12. <http://www.stata.com/manuals/irt.pdf>

Supplemental reading:

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

October 5 & 7 – Class 10 & 11: Item Response Theory

Central reading:

- 1) Yen, W. M., & Fitzpatrick, A. R. (2006). In R. Brennan (Ed.), *Educational measurement*. Westport, CT: Praeger. pp. 111-118. Sections 1-2.4 ([reserves](#))
- 2) Stata Manual (version 16), pp. 28-58. <http://www.stata.com/manuals/irt.pdf>

Supplemental reading:

- 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. ([reserves](#))

October 14 – Class 12: Polytomous Extensions

Central reading:

- 1) Stata Manual (version 16), pp. 12-18; 77-92. <http://www.stata.com/manuals/irt.pdf>
- 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. ([reserves](#))

October 19 – Class 13: Linking

Central reading:

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

Supplemental reading:

- 2) Reardon, S. F., Kalogrides, D., & Ho, A. D. (forthcoming). 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*. ([reserves](#))
- 3) 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. ([reserves](#))

October 21 – Class 14: Bias/DIF and Accommodations

Central reading:

- 1) AERA/APA/NCME Standards, Chapter 3, pp. 49-72. (required text)
- 2) Stata Manual (version 16), pp. 234-247. <http://www.stata.com/manuals/irt.pdf>

Supplemental reading:

- 3) Koretz (2008), Chapter 11, pp. 260-280. (required text)
- 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. ([reserves](#))

October 26 – Class 15: Standard Setting and Criterion-Referenced Reporting

Central reading:

- 1) AERA/APA/NCME Standards, Chapter 5. (required text)
- 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. ([reserves](#))

Supplemental reading:

- 3) Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement*, 15(4), 237-261. ([reserves](#))
- 4) 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. ([reserves](#))
- 5) Koretz (2008), Chapter 8. (required text)

October 28 – Class 16: Growth and Vertical Scaling

Central reading:

- 1) Kolen, M. J. (2006). In R. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: Praeger Publishers. pp. 171-180. ([reserves](#))

Supplemental reading:

- 2) Briggs, D. C. (2013). Measuring growth with vertical scales. *Journal of Educational Measurement*, 50(2), 204-226. ([reserves](#))
- 3) Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models*. Council of Chief State School Officers. ([reserves](#))
- 4) Pearl, J. (2014). Lord's paradox revisited - (Oh Lord! Kumbaya!). University of California. Technical Report R-36 (October). ([reserves](#))

November 2 – Class 17: Test Design

Central reading:

- 5) Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking Questions*. San Francisco, CA: John Wiley & Sons. Chapter 12: Asking Questions FAQ, 323-333. ([reserves](#))

November 4 – Class 18: Psychometric Frontiers

Central reading:

- 1) Wainer, H. (2010). 14 conversations about three things. *Journal of Educational and Behavioral Statistics*, 35(1), 5-25. ([reserves](#))

Supplemental reading:

- 2) Thissen, D. (2016). Bad questions: An essay involving Item Response Theory. *Journal of Educational and Behavioral Statistics*, 41, 81-89. ([reserves](#))
- 3) Ho, A. D. (2016). The new (educational) statistics: Properties of scales that matter. *Journal of Educational and Behavioral Statistics*, 41, 94-99. ([reserves](#))

November 9 – Class 19: Review for the Conversational Celebration of Learning

No readings

November 16 – Class 20: Practical Validation

Central Reading:

- 1) Stata codebook from: 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. ([Canvas](#))

November 18 – Class 21: Inflation and Test-Based Accountability

Central reading:

- 1) AERA/APA/NCME Standards, Chapter 12 and 13, pp. 183-213. (required text)
- 2) Koretz (2008), Chapter 10, pp. 235-259. (required text)

Supplemental reading:

- 3) Ho, A. D. (2007). Discrepancies between score trends from NAEP and state tests: A scale-invariant perspective. *Educational Measurement: Issues and Practice*, 26(4), 11-20. ([reserves](#))

November 23 – Class 22: Effect Sizes

Central reading:

- 1) Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49, 241-253. ([reserves](#))

Supplemental reading:

- 2) Lipsey, M. W., Puzio, K., Yun C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., Busick, M. D. (2012). *Translating the statistical representation of the effects of educational interventions into more readily interpretable forms*. (NCSE 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U. S. Department of Education. ([reserves](#))

November 30 – Class 23: Stanford Education Data Archive

Central reading:

- 1) Fahle, E. M., Shear, B. R., Kalogrides, D., Reardon, S. F., Chavez, B., & Ho, A. D. (2019). Stanford Education Data Archive: Technical Documentation (Version 3.0). ([link](#))

December 2 – Class 24: NAEP and Plausible Values

Central reading:

- 1) von Davier, M., Gonzalez, E., & Mislevy R. J. (2009). *What are plausible values and why are they useful?* IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments: Volume 2. 9-36. ([link](#))

Supplemental reading:

- 2) Mislevy, R. J., Johnson, E. G., & Muraki, E. (1992). *Scaling procedures in NAEP. Journal of Educational Statistics*, 17, 131-154. ([reserves](#))

Conversational Celebration of Learning

I and Sophie or Emma will hold a 20-minute conversational “celebration of learning” with each of you on November 12th or 13th during set 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.

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 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. If 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.

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 Sophie, Emma, 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.

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 answered at least one interesting research question in a cogent and thoughtful manner?
- 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?

Additional optional texts

All students should consider #1, below, for their reference library. Application-oriented students should consider #3 and #6 for Generalizability Theory and IRT, respectively. More technically oriented students should consider #4 and #5. Students interested in practical methods for large-scale testing should consider #2.

- 1) Educational Measurement, 4th Edition. ISBN: 978-0275981259

<http://www.amazon.com/Educational-Measurement-American-Council-Education/dp/0275981258>

- 2) Test Equating, Scaling, and Linking, 2nd Edition. ISBN: 978-1441923042

<http://www.amazon.com/Test-Equating-Scaling-Linking-Statistics/dp/1441923047>

- 3) Generalizability Theory: A Primer. ISBN: 978-0803937451

<http://www.amazon.com/Generalizability-Theory-Measurement-Methods-Science/dp/0803937458/>

- 4) Generalizability Theory. ISBN: 978-1441929389

<http://www.amazon.com/Generalizability-Theory-Statistics-Behavioral-Sciences/dp/144192938X/>

- 5) Fundamentals of Item Response Theory. ISBN: 978-0803936478

<http://www.amazon.com/Fundamentals-Response-Measurement-Methods-Science/dp/0803936478/>

6) Applications of Item Response Theory to Practical Testing Problems. ISBN: 978-0898590067

<http://www.amazon.com/Applications-Response-Practical-Testing-Problems/dp/089859006X/>

Statistical and psychometric computing

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

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. Sophie and Emma may also cover coding issues in their sections.

I also allow students to complete problem sets and projects in R. Emma can offer extensive R support. 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.

Collaboration and study groups

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 mandate completion of assignments in pairs starting from the second assignment. Partnerships are optional for the first assignment.

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.

Beyond pairs, study groups can be helpful to you as you prepare to do the assignments, both in terms of how to approach the work (including how to use the computer effectively) and in terms of how to think about important concepts. **However, students must turn in work as pairs or individuals where specified. Papers should be written in the pair or individual’s own words—your text should reflect your own understanding of the material.**

You may also find study groups useful, where multiple pairs work together to discuss concepts. Each study group will undoubtedly develop its own structure; nevertheless, here are a few suggestions:

- Groups with six or more members become less useful and may be harder to organize because finding common meeting times becomes increasingly problematic.
- Plan at least one session of 1 ½ to 2 hours (early enough so that there is sufficient time if an additional session is necessary). After 2 hours of statistics, everyone's eyes will be glazing over.
- Schedule the meetings so that you have sufficient time afterwards to write in pairs or individually. When we read your assignments, we focus on what you say and how you say it. The assignments have been devised to require not only computation and programming skills, but skills in analyzing and reporting the material.
- Use the groups to ask questions, try out interpretations, and so on—you each represent each other's resources. Often one person can explain something that makes you see something in a new way—or the other way around. Different people have different insights and strengths – some are good programmers, some ask good questions, others value contextual analysis—and you can learn from listening to what others in a group have to offer.
- **Be careful about cowriting text outside of partnerships.** You and your partner must write your own paper, on your own, using your own language. **Your papers should be written in your own words, not those of your study group.**
- Be sensitive to the distinction between collaboration to plan for and interpret the assignment and collaboration to write up the assignment. The former is encouraged; the latter is not allowed beyond, when applicable, your partner. If the distinction begins to feel murky, refocus your group's work on lecture content and course materials. And let me know me if you have any questions at all.

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 Maritza Hernandez at maritza_hernandez@gse.harvard.edu, 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 or Sophie or Emma.**

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>

Finally, some tips from recent S-061 cohorts (from the past two years) follow. These are the responses to my favorite question (#9) from our course evaluations. I have not edited or omitted any responses!

What advice would you give to students who are thinking of taking this course (about its level, the amount of work required, any prior training needed, ways to get the most out of the course, etc.)?

As long as you have somewhat of a grasp on basic statistics, you will be able to follow along in class. The class is very focused on psychometrics and the study of tests. It is extremely engaging and interesting, with a small class size and focused support. Andrew is also excellent! The only advice I have is to take this only if you enjoy statistics and are interested in testing (which is of course extremely important for educators to understand), otherwise it might be too focused for you.

Even though this is a doctoral-level course, don't feel daunted if you don't have knowledge of advanced statistics as Andrew does an amazing job of scaffolding material and helping with any coding issues that may crop up during homework or project-related work. That being said, previous knowledge of concepts like principal components analysis and interaction effects in multivariate logistic regression will be useful as it'll allow you to enhance your mental model of statistics by building connections between topics.

However, do feel daunted if you're not prepared to set aside at least 12 - 15 hours per week for lecture revision, homework, readings, and online discussions. A bit like learning a new language, mastery in this course doesn't have much to do with intellect but does have a lot to do with deliberate, frequent, and structured practice, and practicing with peers from class rather than in isolation is much more rewarding and productive. Do not take this class if you plan on taking more than 16 credits in the semester.

Takes a lot of work. Very interesting and useful, even for someone who like me who has no interest in working in psychometrics. It is still useful for me to understand the reliability of tests, since they determine so much of how we evaluate students

TAKE THIS CLASS!!! But make sure you take this class with a lighter course load. If you do take the class, start asking questions and seeking help from Andrew and the TF starting day 1. Don't struggle on your own, form a study group!

This course is a lot of work, especially if these concepts are new. It's like learning a new language; you'll need to keep speaking it to get better at it. Form learning groups at the onset of the semester and commit to meeting as a group. When you meet, take turns re-teaching the concepts to each other (even if you think you already it know it). If everyone commits to support each other's learning, the whole experience will be tremendously rewarding and enriching.

This course is a lot of work but DEFINITELY worth it! Make sure you have the time to dedicate to it.

What advice would you give to students who are thinking of taking this course (about its level, the amount of work required, any prior training needed, ways to get the most out of the course, etc.)?

This course is a real eye-opener about educational and psychological testing and, although the subject matter is complex and the work load is heavy, the course and its delivery make the content and work easy to digest.

Workload pretty consistent over the semester (besides the final project, and that depends on your time management). Use office hours and the google docs and things to ask question and get feedback on your ideas and current understanding. A really solid Stats background and a good level of comfort in Stata should be all the background required.

Would very highly recommend this course.

n/a

Take the course! It was really beneficial BUT HARD. It takes time, folks. No two ways around it. You will have to spend time with the content, time digging into Stata, time putting presentations together, turning that into a proposal, making meaning from the tools. But it really does come together in the end - this from someone who really struggled through this course. I still recommend it for those who want to leave here with solid quantitative chops. Plus Andrew is awesome - I took it, knowing that it would be hard for me, because I knew he could help me get through it.

TAKE THIS COURSE! If you're going to do any kind of quantitative research in education, you're probably dealing with test scores. And if you're dealing with test scores, knowledge of psychometrics is absolutely essential (and in demand. It will help bring in the \$\$\$.)

The workload is totally worth it because the class helps you build extremely useful technical and practical skills in dealing with assessments. Consider balancing out workload by taking some easier classes in order to make time for this. If the technical requirements of the course deter you, don't worry about that. I took this course with no background in Stata and just some knowledge of multiple regression. I worked a bit extra to understand logistic regression before the course started, and to replicate code in R through the semester and I was able to get through the course just fine.

Andrew is likely to be one of the best instructors you have had because of his expertise in the field, his love for teaching and deep enthusiasm for improving education.

This is a great course to get you up and running with measurement theory. It provides hands on experience thinking through your own project. It could be a a steep learning curve if you are new to the field. The time commitment is slightly above average, but by no means unreasonable. The best part is that you get to work on your own project and take it in any direction you'd like. Andrew is amazing at teaching and he really makes an extra effort. The personalized feedback and support was really valuable.

Lots of work and complex content, but if you care about the material and are willing to stay on it week in and week out, you won't get a more rigorous learning experience.

I would recommend them to think about potential data to use in an original project even before starting the course. Also, to be prepared to work A LOT of hours on this course... it will be worthy!

This class is a lot of work, but if you are really into psychometric or the technical aspect of testing that it is worth taking. I like how the class also devoted some time to discuss the less technical issues, such as the implication of test-taking strategies, and incorporated real world examples of common practices in standardized testing. The stats part can sometimes be intimidating, but as long as you have some basic knowledge in logit regression, it will be fine. Dont worry about the Stata coding too, because almost all codes are provided. Not advisable to take if your stats foundation is not really strong.

I remember reading in last year's course reviews someone writing that S-061 should be required for all education researchers, and I completely agree! This was a difficult class-- there is a lot of time that must be spent reading and just digesting all of the concepts presented. But, I am extremely happy that I took it. Looking back, I learned a lot and am proud of the progress I am making on a publishable paper based on the project I did in the class.

Pro tip: Make the most of S-061 by coming in with a dataset you can use.

This course is both a sprint and a marathon. As soon as I felt I mastered the material, we moved onto a completely new, but related, topic. This course, easily, was the most time consuming during my 1+ year at HGSE.

It's a heavy lift and very fast paced, but the material is super valuable. It will change the way you look at any measure or scale, and you'll have the tools to critically assess the data you inevitably use in your research. You also walk away with a lot of solid technical skills for measuring reliability and validity of measures and data. The final project can definitely lead to a published paper, too, but you have to put a lot of time into it, ideally while you're also doing the problem sets (which are time consuming for sure).

It's fast paced but if you can take time to focus on the problem sets, readings, and review the materials regularly you can definitely get a lot out of it.

This course is highly useful if you have the time and energy to fully immerse yourself in the material. Andrew is at his absolute best as an instructor - brilliant, inspiring, humorous - able to continually push you to new levels of understanding. After taking this course, you will never see the world in the same way again.

You must make a full commitment to this course because it consumes 15-20 hours per week for class preparation. You also have a required final project, which will be a nearly publishable manuscript. The instructor and the teaching fellow are quite helpful, and you are sure to ask them a lot of questions. After the semester is over, you will be surprised that you have learned so much about almost all the contemporary psychometric methods and it is totally worth the endeavor to take this amazing course!

This course was a lot more manageable than I expected (I think some changes to content load were made) so I wouldn't feel intimidated by an excessive time commitment. My advice: reference the problem sets while writing the final paper - they end up being very helpful!