Understanding Educational Testing

EDST0213 Syllabus Fall 2022

Steve Hoffman

2022-08-31

Course Description

Achievement testing is now a cornerstone of education policy. It is also complex and routinely misunderstood by educators, policymakers, and the media. In this course students will use statistical methods to explore and address testing issues that arise in both policy and practice. We will examine the uses and abuses of educational assessment. We will examine and interrogate trends and group differences in achievement. And we will broaden our understanding of essential concepts of measurement, such as reliability, validity, and bias, by analyzing both large and small datasets. Prior experience with the statistical package "R" is not required, as learning this package will be part of the course. 3 hrs. lect.

Contact Professor Hoffman

- email: hoffman@middlebury.edu
- *Zoom Hours*
  - Tuesdays: 4:30 - 6:00 PM
  - Thursdays: 11:00 - 12:30 PM
  - And by appointment

Format of the course

Lecture/discussion sessions are scheduled every Tuesday and Thursday from 1:30 - 2:45 in Twilight 301. Should the pandemic require us to pivot, be ready to participate online at the designated class time. **Your active participation and engagement are essential to both your success and your colleagues' success in this course.** Come prepared to participate each Tuesday and Thursday. The general format is a mixture of pre-recorded lectures, in-class lectures, in-class discussion, asynchronous Canvas discussion, small-group work, partner projects and opportunities for students to lead learning.

"Students who have Letters of Accommodation in this class are encouraged to contact me as early in the semester as possible to ensure that such accommodations are implemented in a timely fashion. For those without Letters of Accommodation, assistance is available to eligible students through the Disability Resource Center. Please contact ADA Coordinators Jodi Litchfield and Peter Ploegman in the DRC at ada@middlebury.edu for more information. All discussions will remain confidential."

Mission and Guiding Principals

Mission Statement of the Education Studies Program (Revised June 2015)
In the Education Studies Program, we believe that we must become a more inclusive and just society. We must honestly name and relentlessly address the educational inequities that we have created and that we sustain as individuals and members of multiple communities. In our formal settings, whether a Kindergarten class or senior seminar at Middlebury, and in our informal interactions with each other, we seek to embody the intellectual understandings, the humility, and empathy essential to this work.

The mission of the Education Studies Program is to prepare students to effectively contribute to a more just, compassionate, and equitable society by developing their self-awareness, critical lenses, analytical frameworks, and pedagogical content knowledge manifested in meaningful practices.

This is a learning space. People will make mistakes as we navigate difficult terrain. Please be willing to engage in difficult discussions with humility and respect. Respect means realizing the difference between intent and impact. Respect means helping others consider the impact of their words or the limits of their perspectives. And respect means authentically listening when one's ideas are challenged or questioned. Recognize that people are experts of their own lives, but do not ask them to speak for an entire group.

Course Objectives

In this course, we will:

- critically examine the uses and abuses of educational assessment
- examine and interrogate trends and group differences in achievement
- broaden our understanding of essential concepts of measurement, such as reliability, validity, and bias
- gain facility employing statistical methods to explore and address testing issues that arise in both policy and practice
- use the statistical package "R" to analyze testing data sets
- communicate our understandings and findings clearly to the general public

Required Texts

Koretz, D. (2008). *Measuring up: What educational testing really tells us.* Harvard University Press. (ebook OK)

- Early on (on page 2), Professor Koretz noted that at the time he wrote this book there were many books on educational testing. Some were very pro-testing while others were anti-testing. Nearly 15 years later this is still true. Koretz intended to provide a balanced approach to explaining the field of educational testing to a wide audience, and this book is still required for courses on testing at other colleges and universities — both those for the "informed consumers of test scores" as well as for more technical courses. So while this book is not new, and we will come across explanations that seem dated, it remains an excellent resource for us as we work through today's issues around educational testing.

Wickham, H. & Grolemund, G. (2017). *R for Data Science.* https://r4ds.had.co.nz/index.html (free to use)

- This ebook is about how to "do" data analysis and data science using the "R" package, which you may have used in other courses at Middlebury. We will utilize only a small portion of this book, though you may find it useful to bookmark this for work outside of this course. Here we focus mainly on data visualization (chapters 1 - 3), workflow (chapters 4, 6, 7, and 8), and communication (chapter 27). And for those of you with no experience with R at all, don't fret. I and many others here at Middlebury will help you learn the basic tools in R, allowing you to thrive in this course.

Grading policies

Final course grades will be based on a typical 100-point system. A = 93 – 100; A- = 90 – 92; B+ = 87 – 89; etc. Late work will likely be penalized.

- Attendance and participation – 20%
- Quizzes – 20%
- Discussion posts – 20%
- Data-Analytic Memos – 20%
- Final project – 20%

Attendance and participation

Attendance is required. Come to class on Tuesday and Thursday afternoons prepared and ready to participate. Students who must miss a class should notify Hoffman before the class meeting and arrange for alternate ways to contribute to the class community. (Pro tip: Don't

write "let me know if I missed anything.") Participation grades are calculated based on attendance, punctuality, contributions during small-group activities, and thoughtful engagement in discussions.

Quizzes

Short, low-stakes quizzes to check understanding of the course content will occur often throughout the term. Sometimes quizzes will be administered during face-to-face class sessions. Sometimes quizzes will be administered asynchronously through Canvas. **These are not meant to be high-pressure events.**

Discussion posts

Students will write reflections of 400 words or fewer in response to Canvas discussion prompts about designated assigned readings. Organize your thinking about the ideas and arguments made by the authors. If you reference articles, papers, books, or other media beyond the course texts, cite them appropriately. Make an informed critique of ideas, rather than merely summarizing the readings. Also, list a "burning question" that (perhaps) remains unanswered after reading, placed at the end of the reflection. Finally, **read your colleagues' posts,** and write a brief, thoughtful response to at least three (3) of your classmates' posts. Tentatively due on September 21, October 5, October 19, November 2, November 16, and December 5.

Data-Analytic Briefs

Designed to develop and extend your data-analytic skills and to help you learn to communicate your findings clearly to others, these are collaborative assignments to be completed with a partner. Please engage in a full, fair and mutually-agreeable collaboration with your partner, and do not simply divide up the work. Discuss and plan the analyses together, debate what you have found with each other, and collaborate on the writing of your brief. My objective is to provide a high-quality opportunity for you to learn, review, teach, and communicate the course material to others. Briefs are tentatively due September 28, October 12, October 26, November 9, and December 2

Final project

Due December 19 (coincides with Exam L Schedule).
Assigned during the last week of the term, with an opportunity to draft an outline of your project in early December, this project will serve as our final exam for the course. (There will be no scheduled in-person test during finals week.)

Late Work

If you need an extension, **please ask before the assignment is due**. I generally assume that if you ask, you need an extension for a valid reason. Work with me to come up with a reasonable plan for turning in late work.

Course Outline

September

Sept 13: Introduction to the course

- *Measuring Up*, Prologue & Chapter 1
- *R for Data Science*, Chapters 1 & 2

Sept 15: Fundamental Issues in Measurement

- *Measuring Up*, Chapter 2
- *R for Data Science*, Chapter 3 & 4
- Thorndike, Chapter 2, pp. 23-28, (Canvas file). This supplemental measurement textbook (*Measurement and Evaluation in Psychology and Education*, by Thorndike & Thorndike-Christ) includes a play data set for us to work with.

Sept 20: Measurement concepts

- Watch the introductory video about NAEP

  https://www.youtube.com/watch?v=6WUfX9YZyL8&feature=emb_logo

- *Measuring Up*, Chapter 3 & 4
- Thorndike, Chapter 2, pp. 28-38, (Canvas file). *Just skim the instructions on how to perform various statistical procedures in SPSS and Excel; we will work in R.*
- *R for Data Science*, Chapters 5 & 6
- Optional reading: Lindquist, E.F. (1951) (Canvas file)

Sept 22: Really getting our R on

- *R for Data Science*, Chapters 7, 8, & 27

Sept 27: Trends in achievement tests

- Thorndike, Chapter 2, pages 39-57, (Canvas file). *Skim the instructions on how to perform various statistical procedures in Excel and SPSS.*
- *Measuring Up*, Chapter 5
- NAEP Long-Term Trend Assessment Results https://www.nationsreportcard.gov/ltt/

Sept 29: Achievement patterns and COVID

- *Measuring Up*, Chapter 6
- Kuhfeld, Megan, James Soland, and Karyn Lewis. (2022). Test Score Patterns Across Three COVID-19-impacted School Years. (EdWorkingPaper: 22-521).

  https://www.edworkingpapers.com/sites/default/files/ai22-521.pdf

October

Oct 4: *NAEP: The Nation's Report Card*

- Familiarize yourself with the NAEP website: https://nces.ed.gov/nationsreportcard/
- Read "From The NAEP Primer: A Technical History of NAEP

  https://nces.ed.gov/nationsreportcard/about/newnaephistory.aspx#beginning
- NAEP Report Card: 2019 NAEP Mathematics Assessment

  https://www.nationsreportcard.gov/highlights/mathematics/2019/
- NAEP Report Card: 2019 NAEP Reading Assessment

  https://www.nationsreportcard.gov/highlights/reading/2019/
- For Reading and Math in both 4th and 8th grade, please study national average scores and national score gaps (Asian-White, Black-White, Hispanic-White, female-male, and Eligible-Not Eligible for National School Lunch Program).

Oct 6: Frameworks for NAEP

- Mathematics Framework for the 2026 National Assessment of Educational Progress, Ch. 2, *Mathematics Content*, pp. 14-48. This chapter includes detailed tables of standards, but you don't need to keep track of all of this. The goal is to get a sense of the breadth of the standards as a whole and the nature of the 'chunks' into which mathematics is broken into for the purpose of building the NAEP test.

  https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/mathematics/2026-math-frameowork/NAEP-2026-Mathematics-Framework-Combined.pdf (Note: If you are curious, look at the archived copies of the math frameworks for 2017 or 2019. The majority of the content of these frameworks remain in all versions.)

- Reading Framework for the 2019 National Assessment of Educational Progress, Ch. 2, *Content and Design of NAEP in Reading*, pp. 17-46. The details of the standards are less important than the big picture, but it's important for you to have an understanding of the content breakdown and the assessment issues involved. As of July, 2022, the Reading Framework has not been updated for the 2020s.

  https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/reading/2019-reading-framework.pdf

Oct 11: Performance standards

- *Measuring Up*, Chapter 8 (pages 179 – 200)
- Hansche L., et al. (1999). Handbook for the Development of Performance Standards: Meeting the Requirements of Title I. Washington, D.C.: Council of Chief State School Officers. Chapters 1 (pages 3-6), 3 (pages 11- 16 only), and 10 (pages 87-103). (Canvas file)
- Pages 1 - 2 & Pages 73 - 78 New Meridian Technical Report 2018-2019 (formerly PARCC)

  https://files.eric.ed.gov/fulltext/ED604241.pdf

Oct 13: Scales

- *Measuring Up*, Chapter 8 (pages 200 – 214)
- Thorndike, pp. 72-95 (Canvas). *Skim the instructions on how to perform various statistical procedures in Excel and SPSS.*

Oct 18: International Comparisons

- PISA 2018 Results: What students know and can do, Volume 1, pages 15 - 30 & 41 - 84

  https://www.oecd-ilibrary.org/docserver/5f07c754-en.pdf?expires=1659981276&id=id&accname=guest&
- PISA 2012 Released Mathematics Items (Please read the Forward and then scan through some of the items)

  https://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf
- Highlights TIMSS 2019 International Results in Mathematics and Science, pages 1 - 8 (Canvas)

Oct 20: Error

- *Measuring Up*, Chapter 7

Oct 25: Reliability

- Pages 1 - 37 of the Florida Standards Assessments 2018-2019 Technical Report, Volume 4: Evidence of Reliability and Validity. Read this to understand the general concepts and how they are conveyed. (Don't fret the crazy math!)

  https://www.fldoe.org/core/fileparse.php/5663/urlt/V4-FSATechReportYear1819.pdf

- Pages 4 - 6 & 132 - 135 New Meridian Technical Report 2018-2019 (formerly PARCC)

  https://files.eric.ed.gov/fulltext/ED604241.pdf

Oct 27: Validity

- *Measuring Up*, Chapter 9
- Pages 38 – 42 and pages 65 - 75 of the Florida Standards Assessments 2018-2019 Technical Report, Volume 4: Evidence of Reliability and Validity

  https://www.fldoe.org/core/fileparse.php/5663/urlt/V4-FSATechReportYear1819.pdf

- Pages 136 - 152 New Meridian Technical Report 2018-2019 (formerly PARCC)

  https://files.eric.ed.gov/fulltext/ED604241.pdf

November

Nov 1: Score Inflation

- *Measuring Up*, Chapter 10
- Holcomb, R., Jennings, J. L., & Koretz (2013). The roots of score inflation: An examination of opportunities in two states' tests. In G. Sunderman (Ed.), Charting Reform, Achieving Equity in a Diverse Nation. *Information Age Publishing.* (Canvas file)

Nov 3: Adverse impact and bias

- *Measuring Up*, Chapter 11
- Hutt, E. & Schneider, J. (2018). A history of achievement testing in the United States or: Explaining the persistence of inadequacy. *Teachers College Record, 120*(11) pages 1-34. (Canvas file)
- Papay, J., Mantil, A., & Murnane, R. (2021). On the threshold: Impacts of barely passing high-school exit exams on post-secondary enrollment and completion. (EdWorkingPaper: 22-627). Retrieved from Annenberg Institute at Brown University.: https://doi.org/10.26300/rhx2-yv60.

  https://www.edworkingpapers.com/sites/default/files/ai22-627.pdf

Nov 8: Testing special populations

- *Measuring Up*, Chapters 12 & 13

Nov 10: College Admissions

- Koretz, D., Yu, C., Mbekeani, P., Langi, M., Dhaliwal, T., and Braslow, D. (2016). Predicting freshman grade-point average from college-admissions and state high-school test scores. *AERA Open, 2*(4), 1-13.
- Westrick, P. A., et al. (2019). National SAT Validity Study. https://satsuite.collegeboard.org/media/pdf/n sat-validity-study.pdf *The College Board*.
- Marini, J. P., at al. (2019). Validity of SAT Essay Scores for Predicting First-Year Grades.

  https://research.collegeboard.org/media/pdf/validity-sat-essay-scores-predicting-first-year-grades.pdf *The College Board*.
- Report of the UC Academic Council Standardized Testing Task Force (STTF). (2020). Executive Summary, pp. 3 – 7

  https://senate.universityofcalifornia.edu/_files/underreview/sttf-report.pdf

Nov 15: Cheating

- Aviv, R. (2014). Wrong answer: In an era of high-stakes testing, a struggling school made a shocking choice. *The New Yorker*. July 21, 2014 issue. (Canvas)
- Merrow, J. (2013). Michelle Rhee's reign of error. *The Merrow Report*.

  https://themerrowreport.com/2013/04/11/michelle-rhees-reign-of-error/
- Straus, V. (2011). And now a new standardized testing scandal. *The Washington Post*.

  https://www.washingtonpost.com/blogs/answer-sheet/post/and-now-a-new-standardized-testing-scandal/2011/06/17/AGotFQaH_blog.html
- Herold, B. & Mezzacappa, D. (2011). 2009 report identified dozens of PA schools for possible cheating. *thenotebook*.

  https://thenotebook.org/articles/2011/07/08/2009-report-identified-dozens-of-pa-schools-for-possible-cheating/
- Sanchez, C. (2013). El Paso schools cheating scandal: Who's accountable?

  https://www.npr.org/2013/04/10/176784631/el-paso-schools-cheating-scandal-probes-officials-accountability

Nov 17: More cheating

- Flanagan, C. (April 4, 2019). They had it coming. *The Atlantic*.

  https://www.theatlantic.com/ideas/archive/2019/04/what-college-admissions-scandal-reveals/586468/

- Janowski, E. (May 15, 2021). Cheating investigation embroils Geisel in controversy. *The Dartmouth*.

  https://www.thedartmouth.com/article/2021/05/cheating-investigation-embroils-geisel-in-controversy

- Morey, A. (June 10, 2021). Dartmouth drops cheating charges against med students, apologizes for flawed investigation. *The Fire*.

  https://www.thefire.org/dartmouth-drops-cheating-charges-against-med-students-apologizes-for-flawed-investigation/

- Levenson, M. (2022). Hoping to Identify Cheaters, A Professor Sues His Own Students. *The New York Times*.

  https://www.nytimes.com/2022/03/17/us/chapman-law-cheating-professor.html?referringSource=articleS

- Goldstein, M. (2022). Ernst & Young to Pay $100 Million Fine After Auditors Cheated on Ethics Exams. *The New York Times*.

  https://www.nytimes.com/2022/06/28/business/ernst-young-sec-cheating.html?referringSource=articleSh

Nov 29: Test-based Accountability

- Castellano, K. E., & Ho., A. D. (2013). *A Practitioner's Guide to Growth Models. Washington, D.C.: Council of Chief State School Officers*, February. Up to the heading "Growth Models of Interest" on p. 17. *Note: The rest of the guide provides a clear, detailed, largely nontechnical explanation of the major approaches to measuring "growth."*

- American Statistical Association. (2014). *ASA Statement on Using Value-Added Models for Educational Assessment.*

- Everson, K. C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? *Review of Educational Research, 87*(1), 35-70. DOI: 10.3102/0034654316637199

- Florida Education Association (2019). Value-added measures: Valuing meaningless data over high-quality instruction.

  https://feaweb.org/news/frontline/vam-scam/

- Bitler, et al (2020). Can a teacher really impact student height? A cautionary tale on value-added models. Brookings Institute. January 21, 2020.

  https://www.brookings.edu/blog/brown-center-chalkboard/2020/01/21/can-a-teacher-really-impact-student-height-a-cautionary-tale-on-value-added-models/

- Optional: Bitler, M. Corcoran, S. Domina, T. & Penner, E. (2019). Teacher effects on student achievement and height: a cautionary tale. NBER Working Paper 26480.
- Jacob, B. & Rothstein, J. (2016). The measurement of student ability in modern assessment systems. *Journal of Economic Perspectives, 30*(6), 85-108.

  https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.30.3.85

December

Dec 1: Performance Assessment

- Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan, 80*(9), 662-666. (Canvas)
- Sample performance assessment items from *Smarter Balanced*

  https://sampleitems.smarterbalanced.org/
- SBAC "Penny Argumentative Performance Task" (Grade 8) (Canvas)
- Performance Task Writing Rubric (Grades 6-11)

  https://portal.smarterbalanced.org/library/en/performance-task-writing-rubric-explanatory.pdf

Dec 6: Solutions?

- Klugman, Emma M., and Andrew D. Ho. (2020). How Can Released State Test Items Support Interim Assessment Purposes in an Educational Crisis?. (EdWorkingPaper: 20-292). Retrieved from *Annenberg Institute at Brown University*: https://doi.org/10.26300/yr8t-0v59

Dec 8: NAEP revisited, Course Review and Final Project workshop