APSY-GE 2524

PSYCHOLOGICAL MEASUREMENT

Steinhardt School of Culture, Education, and Human Development
New York University

Thursdays, 2-4:30pm
Zoom link: https://nyu.zoom.us/j/98466387576

Instructor:
Alejandro J. Ganimian
Assistant Professor of Applied Psychology and Economics
246 Greene St., Kimball Hall, Room 406W
+1 212 998-4029
alejandro.ganimian@nyu.edu

Office hours:
https://calendly.com/alejandro-ganimian/office-hours

1. Objectives

This course seeks to introduce students to key considerations in the design, administration, and analysis of instruments for psychological research. It focuses on three overarching questions: (a) how can we design instruments to measure our construct(s) of interest?; (b) how can we administer instruments to maximize the amount of useful information we will obtain (and conversely, minimize error)?; and (c) how can we analyze individuals' responses to accurately represent the measurement procedure? It offers an overview of approaches applicable to a wide array of instruments used in psychology, but an important part of the course focuses on educational measurements (e.g., achievement tests and classroom observations). It draws on examples of quantitative research from psychology and economics.

The components of the course aim to achieve different, but complementary, objectives:

- The required readings, to be completed before each lecture, will introduce students to a problem in measurement (e.g., measurement error), the conceptual frameworks that can be used to think about this problem (e.g., classical and generalizability theory), and the analytical strategies employed to address the problem (e.g., Cronbach's alpha and G-studies).
- The lectures will briefly review the problem introduced in the readings, explain the conceptual frameworks, and compare and contrast different approaches to solve the problem, drawing extensively on examples from applied research.
- The problem sets, which can be completed in groups but must be written-up individually, will provide students with opportunities to practice implementing the approaches discussed in lecture on their own using a statistical package.
- The final take-home exam (for master's students) or project (for doctoral students), which must be completed individually, will assess students' ability to apply the material covered in the course independently.

The sequencing of these components (i.e., the fact that students will first complete the readings, then come to lecture, complete problem sets in pairs, and finally apply what they learn independently) aims to provide students with the necessary scaffolding to become critical consumers of research in psychological measurement. By the end of the course, students will be expected to understand the concepts, methods, and analytical strategies on their own.

This course draws on "Statistical and Psychometric Methods for Educational Measurement" at the Harvard Graduate School of Education. The instructor thanks Andrew Ho and Daniel Koretz for sharing their materials and allowing him to use them in this course.

2. Pre-requisites

Students are expected to have taken two semesters of statistics, including analysis of variance and logistic regression. NYU students may fulfill this requirement by taking APSTA-GE 2086 ("Basic Statistics II") *and* APSTA-GE 2003 ("Intermediate Quantitative Methods") *or* APSTA-GE 2002 ("Statistics for Behavioral and Social Sciences II") before taking this course. Students who do not meet these pre-requisites should notify the instructor within the first week of class.

3. Auditing

This course may be taken for a letter-grade only, not on a satisfactory/no credit basis. Auditors are not allowed for two reasons. First, students are unlikely to master the material in the course if they do not complete all requirements (i.e., attend class regularly, participate, and complete the problem sets and exam or project). If a student plans to complete these requirements, he/she should receive credit. Second, the instructor works hard to support registered students throughout the semester. Auditors place additional demands on the instructor, which invariably limit his capacity to provide this support. Therefore, there will be no exceptions.

4. Readings

There will be two recommended texts for this course:

- Koretz, D. M. (2009). *Measuring up: What educational testing really tells us.* Cambridge, MA: Harvard University Press.
  - Consider purchasing the book in print. The e-book version (at least in Google Books) omits all figures due to copyright issues.
- 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.
  - Consider purchasing through APA/AERA/NCME at: http://bit.ly/1iczJWH. Members receive a discount and anyone who buys it also receives an electronic text.

Students may purchase them if they prefer, given that we will read multiple chapters from both of these texts. However, all readings will be posted on the course site through NYU Classes at: www.nyu.edu/its/classes. If students are unable to access the readings before a lecture, they should notify the instructor as soon as possible.

The readings will introduce students to the statistical concepts and methods to be covered during lecture. Students are not expected to understand them before attending lecture, but they must have completed them and made a good-faith effort to understand them.

5. Grading

Each student's grade in the course will be determined as follows:
a) attendance (5%);
b) class participation (15%);
c) four problem sets (50%); and
d) final take-home exam or project (30%).

Attendance: Students are expected to attend all lectures, arriving before the start of each lecture to allow the instructor to start on time. If a student cannot attend a lecture, he/she should e-mail the instructor at least 24 hours in advance stating the reason for the absence. If a student missed a lecture but was unable to notify the instructor (e.g., in the case of a health condition or emergency), he/she should do so afterwards. These communications help the instructor understand students' circumstances and provide additional support whenever necessary.

Both planned and unplanned absences will be considered in the 5% of the unadjusted course grade assigned to attendance and punctuality. There is only one exception. In accordance to NYU's calendar policy on religious holidays, students who let the instructor know of their absences due to religious holidays ahead of time will not incur any penalty on their attendance score. However, they are expected to review the slides and complete the assigned readings to catch up with any missed lectures. If students expect to require extensions due to religious holidays, they should reach out to the instructor as soon as possible so that such extensions may be extended to the rest of the students. No extensions will be granted to individual students.

Each student's attendance score will be calculated as follows. The student will receive a score of 1 for attending each lecture before the official start time and a score of 0 for being absent. The student's total attendance score will be the sum of all the individual scores over the total number of lectures, multiplied by 100. For example, if a student attended 12 of 14 lectures, his/her score will be (12/14)*100 or 86. The maximum attendance score is 100.

For reference, the mean attendance scores for previous iterations of this course were: 94 (fall 2017) and 97 (spring 2020).

Class participation: During each lecture, the instructor may call on students to ask them questions related to the required readings. Students are not expected to understand the concepts, methods, or analytical strategies in the readings before they are discussed in lecture, but they must have made a good-faith effort to understand them. Students' responses, as well as the questions they pose to the class and/or the instructor, will be considered in the 15% of the unadjusted course grade assigned to class participation.

Each student's class participation score will be calculated as follows. The student will receive a score of 1 for making a good-faith effort to answer questions (even if they do so incorrectly), asking clarifying questions, and/or making relevant contributions during each lecture and a score of 0 for attending class but refraining from participating. The student's total participation score will be the sum of all the individual scores over the total number of lectures, multiplied by 100. For example, if a student met expectations in 13 of 14 lectures, his/her score will be (13/14)*100 or 93. The maximum participation score is 100.

For reference, the mean participation scores for previous iterations of this course were: 79 (fall 2017) and 95 (spring 2020).

Problem sets: Students are expected to complete four problem sets throughout the semester. As stated in the course objectives, these problem sets are meant to provide students with opportunities to practice the material covered in lecture. Students can complete problem sets in groups (ideally, pairs), but they must write up their results individually. Instructions on how to format and submit problem sets will be included at the beginning of each assignment. The problem sets from previous iterations of the course are posted on the "Resources" tab of the course site. These are meant to provide students with general guidance on the expected level of detail of their answers and the instructor's approach to grading. Yet, the content and types of questions in problem sets vary from one semester to the next as the course continues to evolve.

Each student's problem-sets score will be calculated as follows. The student will receive a score of 0 to 100 on each problem set, based on the proportion of questions he/she answered correctly. Partial credit will be awarded for partially correct answers, so students are encouraged to show their work. The student's overall problem set score will be the average of the three highest problem-set scores (i.e., the lowest score will not count). This provision is meant to account for the fact that some students may find some of the problem sets more difficult than others, and to prevent one low problem set score from playing a large role in determining students' overall grade. It is also meant to allow students to "drop" (i.e., choose not to complete) one problem set during the semester (e.g., if they cannot complete the problem set on time due to unforeseen circumstances). For example, if a student obtained scores of 90, 80, 100, and 50, his/her score will be (90+80+100)/3 or 90. The maximum problem set score is 100.

For reference, the mean problem-sets scores for previous iterations of this course were: 84 (fall 2017) and 97 (spring 2020).

Final take-home exam or project: Master's students are expected to complete one final take-home exam. As stated in the course objectives, the exam aims to assess students' ability to apply the material covered in lecture independently. Students must complete the exam individually. Doctoral students are expected to complete one final project. Ideally, these projects will help students make progress towards required submissions for their respective doctoral programs (e.g., qualifying papers or dissertations). Note that these are simply the "default" options for master's and doctoral students. If students wish to switch (e.g., a master student wants to complete a final project instead of the take-home exam), they can do so by simply notifying the instructor over e-mail. However, all students who are scheduled to complete a final project must comply with all the milestones outlined in the course calendar below, regardless of whether they were assigned to a project by default or whether they made the switch during the semester.

The final exams and projects from previous iterations of the course are posted on the "Resources" tab of the course site. The exams are meant to provide students with general guidance on the expected level of detail of their answers and the instructor's approach to grading. Yet, the content and types of questions vary from one semester to the next as the course continues to evolve. The projects are meant to illustrate the types of questions students examine, as well as the expected length, structure, and format of the final projects. Yet, these vary based on the intended purpose of the project (e.g., if a student plans to use the project a dissertation appendix, his/her write-up will differ from that of a peer who will use it a second-year paper).

Each student's final-exam or project score will be calculated as follows. The student will receive a score of 0 to 100 on the exam or project, based on criteria to be specified before/after each assignment (in the case of the exam, the instructor will post an answer key after grading all exams; in the case of the projects, the instructor will post instructions for each milestone). In the exam, partial credit will be awarded for partially correct answers, so students are encouraged to show their work. For example, if a student obtained a score of 90, that will be his/her score.

For reference, the mean final-exam scores for previous iterations of this course were: 82 (fall 2017) and 88 (spring 2020). The mean final-project score in the spring of 2020 was 95 (there was no option to complete a final project in the fall of 2017).

<u>Overall course grade:</u> The overall numeric score for each student will be calculated as the weighted average of his/her attendance, class participation, problem sets, and final exam or project. The weights correspond to the percentages allotted to each score above. For example, if a student obtained an 86 for his/her attendance, a 93 for his/her class participation, a 90 for his/her problem sets, and a 90 for his/her final exam or project, his/her overall numeric score will be (86*0.05)+(93*0.15)+(90*0.5)+(90*0.3) or 90.

The overall letter grades will be determined based on the distribution of numeric scores for the corresponding comparison group for each student (i.e., master's or doctoral students). This is meant to account for the fact that some student cohorts may find the material more or less difficult than others. Letter grades will be assigned as follows:

If a student has a numeric score that is…	…he/she will earn a/an…
…0.5 standard deviation (SD) above the mean…	…A
…above the mean by less than 0.5 SD…	…A-
…below the mean by less than 0.5 SD…	…B+
…between 0.5 and 1 SD below the mean …	…B
…between 1 and 1.5 SD below the mean…	…B-
…between 1.5 and 2 SD below the mean…	…C+ or lower

The mean and standard deviation for each group will be posted on the course site when mid-term grades are due (on April 2) and when final grades are due (on May 18 at the latest). Both of these grade releases will be accompanied by a letter to each student explaining how his/her grade was calculated. Mid-term letter grades are not final; they are meant to inform students of their relative standing and to allow students to seek feedback from the instructor on how to improve their grade. For example, if a student obtained a numeric score of 90, his/her group mean is 80, and the group standard deviation is 10, (90-80)/10 or 1 SD above the mean corresponds to an overall letter grade of A. (Please, note this group mean and standard deviation are only examples).

The cutoff scores have varied across semesters as follows:

			Spring 2020	
Criterion	Letter grade	Fall 2017	MA students	PhD students
0.5 standard deviations (SDs) above the mean	…A	88	96	97
above the mean by less than 0.5 SDs	…A-	83	94	90
below the mean by less than 0.5 SDs	…B+	78	92	83
between 0.5 and 1 SD below the mean	…B	72	90	76
between 1 and 1.5 SD below the mean	…B-	67	88	69
between 1.5 and 2 SDs below the mean	…C+ or lower	62	86	62

The instructor may (and often does) adjust a student's final letter grade on the course based on his/her improvement over time and exemplary performance on one or more dimensions, so the actual distribution of letter grades is never determined exclusively by the cutoff scores above.

All grades posted at the end of the semester are final and the instructor will not discuss grades over e-mail. Students interested in better understanding their grades after they are posted are welcome to make an appointment with the instructor at the start of the following semester. There will be no exceptions to ensure no students are given an unfair advantage over others.

Grading criteria for assignments: After each problem set is graded, the instructor will post the answer key, scoring criteria, and student exemplars (i.e., anonymized problem sets with top scores, with students' permission). Students are strongly encouraged to consult these documents to ask the teaching team any questions they might have on the material.

A student may ask for his/her problem sets and/or mid-term exam to be regraded if—after carefully reviewing the answer key, scoring criteria, and exemplars—he/she does not believe that his/her grade is correct. Students who wish to request a regrade should e-mail the instructor no later than one week after scores have been posted. The instructor will conduct all regrades. He will regrade the entire problem set or exam, not just the questions being disputed. Therefore, regrades may result in a lower, equal, or higher scores than the ones originally awarded. The final exam and project scores are final (i.e., not subject to regrades).

6. Classroom policies and expectations

Late assignments: Students should budget enough time to submit their problem sets, project milestones, and exams well ahead of each deadline. Instructions for submitting each assignment will be announced during lecture. Late assignments, regardless of how late they are (even a minute past the deadline), will not be accepted. There will be no exceptions to prevent the instructor from having to determine whether some circumstances are more meritorious of an extension than others. Further, the course already has a mechanism to cope with unanticipated events: dropping the lowest of four problem-set scores (see Grading section above).

Surveys: The instructor will invite students to complete two surveys during the semester: a "student survey" (at the beginning of the semester), which will allow him to get to know students better, and a "feedback survey" (midway through the semester), which will allow students to provide feedback on what is working well and what could be improved in the course. The instructor takes feedback surveys very seriously and it will make a good-faith effort to address the concerns raised by students.

All surveys are optional and there will be no repercussions for students who choose not to answer them. The student survey will ask for identifying information (to avoid asking questions for which the instructor already has information), but the feedback survey will be anonymous. None of the surveys will be considered in students' course grades. All data survey responses will be deleted at the end of the course and it will not be used for other purposes.

7. Statistical programming

All students will need to get access to Stata, a statistical package, to complete the problem sets for this course. All the example code to be provided by the instructor will be written in Stata 15, so students should get access to Stata 15 or above.

Students may get access to Stata on campus, through the computers at Data Services (on the fifth floor of Bobst Library), the Student Technology Centers (LaGuardia Co-op, Kimmel Center Lab, and Third Avenue Lab; see http://bit.ly/2xgqvHg), or the High Performance Computing's Prince cluster (see https://bit.ly/31Rr4Wq).

Students may also get access to Stata off campus through the Virtual Computer Lab at: http://www.nyu.edu//it/vcl.

Finally, students may purchase Stata at a discounted rate through Stata Campus GradPlan at: http://bit.ly/2w1DrCc. An annual license for Stata/IC (the version for mid-sized datasets), which will be sufficient for this course, is $125.

Lectures will not be used to teach students how to code, but the instructor will upload step-by-step guides with all the commands that students will need for the problem sets to the course site. Students are encouraged to attend office hours to ask coding questions.

Additionally, students can seek help with coding from Data Services (on the fifth floor of Bobst Library) either by signing up for their Stata tutorials (see calendar at https://guides.nyu.edu/DS_class_calendar) or by making an appointment for a one-on-one meeting with a consultant (see https://library.nyu.edu/departments/data-services/.)

Students who believe that they would benefit from a book on Stata are encouraged to consult:

- Kohler, U. & Keuter, F. (2009). *Data analysis using Stata (2nd Edition).* College Station, TX: Stata Press.

Students who believe that they would benefit from an introductory book to probability are encouraged to consult:

- Blitzstein, J. K. & Hwang, J. (2019). *Introduction to Probability (2nd Edition).* Free online access: https://bit.ly/38Thki5. Print copies: https://www.crcpress.com.

Students may also consult the introduction to probability course at Harvard University: https://projects.iq.harvard.edu/stat110.

8. Writing

The problem sets, exams, and projects will involve a fair amount of writing (e.g., to define key concepts or explain results from statistical analyses). Students should not take this writing lightly; an important part of becoming a researcher is learning to convey arguments clearly.

Students are expected to review their assignments for typos and grammatical errors before submitting them. They should also take full advantage of the various on-campus resources to help them improve their writing, including the Writing Center (https://bit.ly/2PMe13x) and the University Learning Center (https://bit.ly/2hBrgX0).

9. Plagiarism

Students taking this course are expected to have read in full and agreed to NYU-Steinhardt's statement on academic integrity (http://bit.ly/2vSt2JR).

As the statement specifies, "plagiarism is failure to properly assign authorship to a paper, a document, an oral presentation, a musical score and/or other materials, which are not your original work." Therefore, any student who works together with or receives help from others on the problem sets should recognize their contributions appropriately (instructions for doing so will be provided in each problem set). This will help the instructor understand any similarities in assignments submitted by different students.

Students who have questions about what constitutes appropriate collaboration in this course should contact the instructor at least 24 hours before they submit their problem sets.

If the instructor suspects that a student has committed plagiarism, disciplinary action may be taken following the department procedure or through referral to the Committee on Student Discipline, through the Office of the Associate Dean for Student Affairs. Please, see the statement on academic integrity for details on the steps involved in each procedure.

10. Accommodations

Any student who needs an accommodation due to a chronic, psychological, visual, mobility and/or learning disability, or who is deaf or hard of hearing, should register with the Moses Center for Students with Disabilities (www.nyu.edu/csd) at 212 998-4980, 726 Broadway, 2nd and 3rd Floors.

Students should also notify the instructor within the first week of the semester. Late requests for accommodation will not be honored except in special circumstances (e.g., injury during the semester).

11. Calendar

This course calendar is tentative. The instructor may adjust the topics to be covered in each class based on how students respond to the material during the semester. Students are expected to check the latest version of the calendar on the course site before every lecture.

Date	Topics	Readings	Assignments
Jan 28	Lecture #1: Introduction to the course • What are the course objectives, components, and grading and classroom policies? • What are the different types of instrument that we will cover in the course? • What is the main limitation of any measurement procedure? (the sampling principle of testing) • What is Stata and how does it work?	Recommended: • Thorndike & Thorndike-Christ, Ch. 1 • Raykov & Marcoulides, Chs. 1 and 2, pp. 13-21	Student survey posted
Feb 4	Lecture #2: How can we know if we can use an instrument for a given purpose? (Validity and validation) • What is validity? • What are the different sources of validity evidence? (content, cognition, coherence, correlation, and consequence) • What are common threats to validity? (construct-irrelevant variance and construct underrepresentation)	Required: • AERA/APA/ NCME (2014). Ch. 1 • Koretz, D. M. (2009). Ch. 9 Recommended: • Kane, (2006). Ch. 2 • Raykov & Marcoulides, Ch. 8, pp. 183-192	Student survey due
Feb 11	Lecture #3: How can we know if an instrument yields consistent results? (Reliability) • What is reliability? • What is the most commonly used approach to measure reliability? (classical test theory) • How can we measure inter-item reliability? (split-half reliability, the Spearman-Brown formula, and Cronbach's alpha) • How can we measure inter-rater reliability? (inter-rater agreement and Cohen's kappa)	Required: • AERA/APA/ NCME (2014). Ch. 2 • Koretz, D. M. (2009). Ch. 7 Recommended: • Haertel, E. H. (2006). Ch. 3, pp. 65-79	Problem set 1 posted

Feb 18	*[Legislative Day – classes meet according to a Monday schedule]*		
Feb 25	Lecture #4: In-class journal-article discussions on validity and reliability • How can validity and reliability be used for instrument development?	Required: • Duckworth, A. L., et al. (2007) • Duckworth, A. L. & Quinn, P. D. (2009)	Problem set 1 due
Mar 4	Lecture #5: How can we know if an instrument yields consistent results? (Generalizability, part 1) • What is a more general approach to measure reliability? (generalizability theory) • How can we measure reliability across multiple facets of error? (G-studies with "crossed" designs) • How can we use estimates of reliability to improve measurement procedures? (the D-studies)	Required: • Brenan, R. L. (1992) • Shavelson, R. J. & Webb, N. M. (1991). Chs. 1 and 2	Final-project proposal due
Mar 11	Lecture #6: How do we know if an instrument yields consistent results? (Generalizability, part 2) • How can we measure reliability when facets of error are "nested" within individuals? (G- and D-studies with nested designs) • How can we extend generalizability theory to more complex measurement procedures? (G- and D-studies with mixed designs)	Required: • Shavelson, R. J. & Webb, N. M. (1991). Chs. 3 and 4	Problem set 2 posted
Mar 18	Lecture #7: In-class journal-article discussions on generalizability • How can generalizability theory improve the design of classroom observations?	Required: • Hill, H. C., Charalambous, C. Y. & Kraft, M. A. (2012)	Problem set 2 due Feedback survey posted
Mar 25	Lecture #8: How can we score student achievement tests to account for differences across items? (Item response theory, part 1) • What are the main disadvantages of classical test theory? • What is item response theory (IRT)? • What are the main assumptions of IRT? (local independence and unidimensionality)	Required: • Yen, W. M. & Fitzpatrick, A. R. (2006). Ch. 4, pp. 111-115 • Harris, D. (1989)	Feedback survey due

	• What are the different types of IRT models? (1-, 2-, and 3-PL models)		
Apr 1	Lecture #9: How can we score student achievement tests to account for differences across items? (Item response theory, part 2) • How can we distinguish between items with different characteristics? (item characteristic curves) • How can we identify items that distinguish between similarly performing examinees? (item and test information)		Problem set 3 posted
Apr 8	Lecture #10: How can we map the results of two or more student achievement tests onto a common scale? (Linking and equating) • What is linking? • How is linking different from equating? • What data-collection approaches can we use to establish a link between two tests? (single-group, equivalent-group, counterbalanced, and non-equivalent group with anchor test designs) • What data-analysis approaches can we use to establish a link between two tests? (mean, linear, and equipercentile linking)	Required: • Kolen, M. J. & Brennan, R. L. (2010). Chs. 1 and 10 • Holland, P. W. & Dorans, N. J. (2006). Ch. 6, pp. 197-201	Final-project first draft due
Apr 15	Lecture #11: How can we know if an item works differently across groups of respondents? (Fairness, bias, and differential item functioning) • What are the different dimensions of fairness in measurement procedures? • What is bias? • How can we determine whether an instrument is biased against certain groups? (differential item functioning)	Required: • AERA/APA/ NCME (2014). Ch. 7 • Koretz, D. M. (2009), Ch. 11	Problem set 3 due
Apr 22	Lecture #12: How can we check whether items in an instrument measure the same construct? (Confirmatory factor analysis) • What is factor analysis? • How can we confirm a priori, hypothesized factor structure? (confirmatory factor analysis)	Required: • Raykov, T. & Marcoulides, G. A. Chs. 3 and 4 (skip parts on programming) • Kline, R. B. (2016), Ch. 9	Problem set 4 posted

Apr 29	Lecture #13: How can we check whether items in an instrument measure the same construct? (Explanatory factor analysis) • How can we explore a potential factor structure?		
May 6	Lecture #14: In-class journal-article discussions on factor analysis • How can factor analysis improve the scoring of classroom observations?	Required: • Sandilos, L. E., DiPerna, J. C., & Family Life Project Key Investigators. (2014)	Problem set 4 due Final take-home exam posted
May 9			Final take-home exam due Final project due