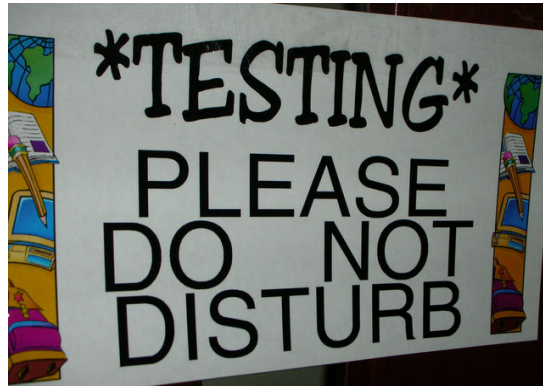

SYLLABUS

EPSY 562 - LARGE-SCALE TESTING (ONLINE)



Fall Semester 2020 (August 24 – December 11)

University of Illinois at Chicago

College of Education

Instructor: Dr. Sarah D. Schnabel

Phone: (847) 373-9675

E-mail: EPSY562@gmail.com

Office hours: by appointment

COURSE OVERVIEW

This course provides an introduction to large-scale testing, presenting an overview of the various tasks that employees perform in testing organizations, city-wide testing bureaus, professional licensing and certification boards, statewide educational testing programs, testing units that are part of state merit systems, etc. The course should be useful for students considering working for such organizations in a variety of capacities (e.g., item writers, statisticians, psychometricians, researchers, testing program managers), employees currently working in these organizations who would like to increase their understanding of the field, and students who want to gain an understanding of the challenges of creating and administering large-scale tests.

The course is organized around the key processes common to all large-scale testing programs: design, administration, scoring, reporting, and validating. The course is not designed to develop the technical skills to carry out specific tasks such as writing items, equating tests, setting cut scores, etc. Rather, the focus is on gaining a conceptual understanding of what is involved in performing these kinds of tasks, and why each task is important. After completing this course, students should have an appreciation for what is involved in producing large-scale tests, as well as an awareness of some of the pressing issues that testing organizations face.

The main text for the course will be the Standards for Educational and Psychological Testing, but we will be supplementing with a variety of other “classic” (and some “not-yet-classic-but-perhaps-destined-to-become-classic”) readings (e.g., book chapters, NCME instructional modules) that address some of the key topics covered by the Standards, such as: defining constructs and content domains, conducting job and practice analyses, writing frameworks and test specifications, reviewing items for content concerns and sensitivity/fairness issues, maintaining test security, scaling, norming, equating, linking tests to establish score comparability, establishing cut scores, testing individuals of diverse linguistic backgrounds, testing individuals with disabilities, fairness in testing and test use, the rights and responsibilities of test takers, and the responsibilities of test users.

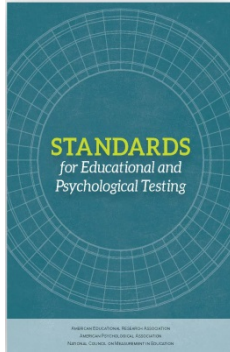
COURSE GOALS

A series of broad, overarching goals undergird this course. In this course, students will work toward acquiring knowledge and a conceptual understanding of:

- the various tasks involved in constructing a large-scale test (e.g., defining a content domain; conducting a job analysis or practice analysis to provide the basis for a test framework; writing test specifications, items, and scoring procedures; assembling item pools and tests; conducting reviews of items for content and fairness considerations; piloting/field testing items; evaluating the quality of items in an item pool and the psychometric properties of a test)
- documents guiding the design and use of large-scale tests (Standards for Educational and Psychological Testing, Code of Fair Testing Practices in Education, Principles for the Validation and Use of Personnel Selection Procedures, Uniform Guidelines on Employee Selection Procedures, Code of Professional Responsibilities in Educational Measurement, Ethical Standards of the American Educational Research Association, Ethical Principles of Psychologists and Code of Conduct, The Rights and Responsibilities of Test Takers)
- key terms used in writing and discussing large-scale testing
- reliability, validity, and errors of measurement and their roles in the construction, selection, interpretation, and use of large-scale tests
- different approaches for scoring, scaling, norming, equating, and linking large-scale tests
- different methods for setting defensible standards/cut scores on large-scale tests
- the steps involved in standardizing a test (e.g., preparing standardized directions, administering tests under standardized conditions, employing standardized scoring procedures, developing materials to interpret test scores)
- what is involved in preparing supporting documentation for a test so that test users will have the information they need to make sound judgments regarding the test's nature and quality, the scores it produces, as well as the interpretations based on the scores
- ways that large-scale tests can be modified to accommodate individuals with various disabilities, and individuals from diverse linguistic backgrounds
- issues of fairness in large-scale testing
- the rights and responsibilities of test takers
- the responsibilities of test users
- key issues that arise in large-scale testing
- key court cases, laws, and legislation that affect large-scale testing programs

REQUIRED TEXTS AND OTHER MATERIALS

TWO TEXTS ARE **REQUIRED** FOR THIS COURSE:



American Educational Research Association, the American Psychological Association, & the National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association. ISBN: 0-935302-35-6

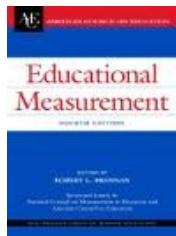
Note: You can order the Standards directly from the American Educational Research Association (<http://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition>). If you are a member of AERA, APA, or NCME, you can purchase the book at a discounted. ***it is very important that you get the 2014 edition of the Standards***



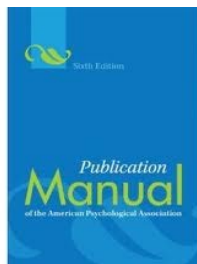
Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2016). Handbook of test development. New York, NY: Taylor & Francis. ISBN: 978-0-415-62602-6

NOTE There is an old (2006) version of this book. Make sure you get the 2016 version.

TWO ADDITIONAL TEXTS ARE **NOT REQUIRED** BUT ARE HIGHLY RECOMMENDED:



Brennan, R. L. (Ed.). (2006). Educational measurement (4th ed). Westport, CT: Praeger. ISBN: 0-275-98125-8



American Psychological Association. (2010). Publication manual of the American Psychological Association. Washington, DC: Author. ISBN-13: 978-1-4338-0561-5

Note: If you prefer, you can order the manual directly from the American Psychological Association (<http://www.apastyle.org/manual/index.aspx>) for \$28.95.

BLACKBOARD COURSE SITE

This course will take place on the UIC Blackboard site. You will need to check the Blackboard course site at least WEEKLY, as this is where I will post course-related documents, direct links to websites related to large-scale testing, assignments, updates, etc. While I intend to follow the syllabus as closely as possible, THINGS ARE SUBJECT TO CHANGE THROUGHOUT THE SEMESTER. Course materials will be provided as Portable Document Format (PDF) files, Microsoft Office documents (such as Word or Excel), or webpages. You will be submitting your memos, taking assessments, and posting to the discussion board through Blackboard, so it will be important that you become familiar with Blackboard and get comfortable using it. **ASSIGNMENTS will typically be due on Sundays by 11:59 PM Central (Chicago) time.** NEW MATERIAL for the upcoming week will always be posted by 9 AM on Monday mornings.

FORMAT AND PROCEDURES

As your instructor, I will assume responsibility for preparing an organized course of study, presenting material in class to expand upon and amplify the topics addressed in the readings, guiding classroom discussions, and facilitating the learning process.

As students in the course, you will be responsible for keeping up with the readings from week to week, actively participating in message board discussions, and handing in assignments on time.

COURSE REQUIREMENTS AND GRADING

This course is worth 4 hours credit. The expectations for receiving credit are:

- participation in a weekly assignment related to the weekly readings (varies by week – usually participation in discussion board posts or an independent formative assessment)
- development of two memos on large-scale testing, and
- completion of a final examination

I have prepared a scoring rubric that I will use to evaluate the memos. You will receive a copy of the rubric in advance so that you will be aware of the criteria I will be using to evaluate your work.

Please note that turning in assignments on time is important to me. See additional information in section entitled “Late/Missing Assignments.”

GRADING SYSTEM

This course will be scored out of 300 possible points. The maximum number of points possible are as follows:

120 points possible for participation (weekly activities)

- 10 points possible for each weekly activity (weeks 2-15, not including week 13 (Thanksgiving break)). There are actually 13 weeks with required activities, so to get the maximum number of points possible, you may skip one week’s activity. You will not get extra credit (only 120 points are possible).

- Typically, I will award you 10 points for participation for sincere attempts (e.g., even if on a formative assessment you answer an item incorrectly, I will award you the full 10 points for your attempt). However, I may award 5 points or 0 points at my discretion if I feel inadequate effort was put forth. You will receive 0 points if the activity is not completed by the due date.

90 points possible for memos

- 45 points possible for each memo
- Points will be taken off for late submissions

90 points possible for the final exam

LETTER GRADES

A

90 – 100% (270-300 points)

Excellent

The student's work demonstrates excellent grasp of all the learning outcomes associated with the course.

B

79 – 89.9% (237-269 points)

Good

The student's work demonstrates mastery of the majority of learning outcomes associated with the course.

C

68 – 78.9% (204 – 268 points)

Average

The student's work demonstrates mastery of approximately two-thirds of the learning outcomes associated with the course.

D

57 – 67.9% (171 – 203 points)

Poor

The student's work demonstrates mastery of fewer than half of the learning outcomes associated with the course.

F

56.9% and Below (170 and below)

Failure

The student's work does not sufficiently demonstrate that he or she has adequately grasped any of the learning outcomes associated with the course.

ASSIGNMENTS

WEEKLY ACTIVITY

Every week, some kind of activity will be required to reinforce your understanding of the readings.

I may post questions about the week's assigned readings in the Discussion Board area of Blackboard. On those weeks, you are required to engage in the discussion by responding to one of my questions or the questions/responses of one of your classmates. Thank you in advance for using the discussion board as an opportunity for meaningful discourse about the readings and posing questions to your classmates. You must participate in the discussion board posts by the end of the week (Sunday night).

Some weeks, I may instead require a different activity, such as a brief formative assessment ("quiz"). On these weeks, please complete the activity by the due date and time listed in the syllabus (usually Sunday night). Feedback on your assessment will usually be provided during the following week.

You are also welcome and encouraged to use the discussion board to post questions and thoughts about the readings even if your weekly activity doesn't require discussion board activity.

MEMOS

You will prepare a series of two, 2-3 page memos (typed, double spaced) in which you will discuss your reactions to various readings on the design, administration, scoring, and/or reporting of results from large-scale tests.

- MEMO #1 is due by 11:59 PM Central (Chicago time) on **Sept. 20**
- MEMO #2 is due by 11:59 PM Central (Chicago time) on **Nov. 15**

The purpose of writing these brief memos is to encourage you to think deeply about the Standards for Educational and Psychological Testing and the implications of those standards for testing organizations and their practices. You can use journal articles, newspaper articles, book chapters, testing monographs, or other resource materials about large-scale testing as the basis for writing your memos. These can be articles that have appeared in the popular press or that you found on the internet, if you like; they need not be peer-reviewed, scholarly academic publications (but they could be). These should be readings other than those that are listed in the syllabus as required or optional readings for class. As the semester progresses, be on the lookout for readings on large-scale testing that raise issues (or concerns) that the Standards for Educational and Psychological Testing address.

Begin each memo with a provocative direct quote from the reading (including the page number on which the quote is found). (The quote need not be more than a couple of sentences.) Following the quote, identify the issue (or concern) raised about large-scale testing in this quote. (It should be an issue (or concern) that the Standards for Educational and Psychological Testing address.) Identify the particular standard(s) that is/are relevant. From the perspective of those standards, discuss why this is an issue (or concern) for large-scale testing. Finally, briefly explain how the issue (or concern) could be handled in a way that would be consistent with the Standards.

Here is the basic outline I would like you to follow for your memos:

- Author, title of article/book/monograph, publisher, date of publication
- Quote (and page number)
- What is the issue (or concern) raised about large-scale testing in this quote?

- What particular standard(s) is/are relevant here?
- From the perspective of these standards, why is this an issue (or concern) for large-scale testing?
- How could the issue (or concern) be handled in a way that would be consistent with these standard(s)?

Additional information about memos and an example memo will be provided to you.

FINAL EXAMINATION

A final examination will be made available during a time period to be specified. The final examination will be open-book and require you to apply your understanding of the readings assigned throughout the semester. More information will be provided as the date nears.

COURSE POLICIES

ACCOMMODATIONS FOR STUDENTS WITH DOCUMENTED DISABILITIES

UIC strives to ensure the accessibility of programs, classes, and services to students with disabilities. I am very willing to provide reasonable accommodations for students with various types of disabilities, such as documented learning disabilities, vision or hearing impairments, and emotional or physical disabilities. If you need accommodations for this class, be sure to register with the Office of Disability Services, 1190 SSB, 413-2183, and let me know your needs. You can find out more about services that this office provides at their website:

<http://drc.uic.edu/>

LATE/MISSING ASSIGNMENTS

The Course Schedule (beginning on the next page) highlights due dates for the assignments for this course. I expect you to keep to the schedule and turn in your assignments on time so that I can provide you with timely feedback on your work. If circumstances arise that make it impossible for you to turn in an assignment on a due date, I expect you to contact me in advance of that date to let me know that the assignment is likely to be late and to determine whether it is possible to negotiate for an extension.

ACADEMIC HONESTY

All work that you submit in this course should be your own original work. Plagiarism (passing off someone else's ideas as your own; copying someone else's class that you have already used in another class, even if it is your own), cheating on exams, or violations of any of UIC's Guidelines Regarding Academic Integrity will not be tolerated. Students who violate the guidelines are subject to disciplinary procedures. Judicial matters include (but are not limited to) the loan or purchase of papers, the use of papers for more than one course, plagiarism, fabrication, and facilitating academic dishonesty. You can obtain a copy of the UIC guidelines from the Office of the Vice Chancellor of Student Affairs. I urge you to read and follow both the letter and spirit of these guidelines.

COURSE SCHEDULE:

WEEK 1 - AUG. 24 – AUG. 30

Topics:

Overview and introduction

To do:

1. Spend some time navigating Blackboard. Click on the links on the left side. The “Weekly Course Content and Readings” section is where you’ll find all the articles and assigned readings that are not from your textbooks.
2. Review the syllabus.
3. Order the two required books for the course.
4. Watch the introduction video (the link is available in the folder: Weekly Course Content and Readings -> Week 1 - <https://www.youtube.com/watch?v=cOyLSUJnTC0>)
5. Post in the discussion board thread "Introduction" to tell the class a little about yourself!
6. Read:

Wise, L. L., & Plake, B. S. (2016). Test design and development following the *standards for educational and psychological testing*. In S. Lane, M.R. Raymond, & T.M. Haladyna (Eds.). (2016). Handbook of test development (pp. 19-39). New York, NY: Taylor & Francis.

WEEK 2 – AUG. 31 – SEPT. 6

Topics:

Test Design and Development

Stating the purpose of a test
Developing test specifications

Readings in Standards for Educational and Psychological Testing:

Part II-4. Test Design and Development, pp. 75-84
and Standards 4.0 – 4.3 pp. 85-86

Readings in Handbook of Test Development:

Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M. (2016). Test development process. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.). (2016). Handbook of test development (pp. 3-18). New York, NY: Taylor & Francis.

[OPTIONAL] Perie, M. & Huff, K. (2016). Determining content and cognitive demand for achievement tests. In S. Lane, M.R. Raymond, & T.M. Haladyna (Eds.). (2016). Handbook of test development (pp. 119-143). New York, NY: Taylor & Francis. (Note: read this if you are interested in getting more information about evidence-centered design and the use of performance-level descriptors (esp. in K-12 assessment))

Other Readings (online):

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 307-324). Westport, CT: Praeger Publishers. (Note: Read just the first two sections of this chapter entitled "Introduction" and "Test and Item Pool Design.")

Watch this video:

Registrar Accreditation Board and The Quality Society of Australasia International. (2011). Exam development process [video]. Milwaukee, WI: Author. Available from http://www.youtube.com/watch?v=LvW13rmRudo&feature=mfu_in_order&list=UL

Assignment Due at End of Week:

Reply to one or more threads in the "Week 2 Forum" discussion board. Feel free to create an additional thread if you have a question of your own.

WEEK 3 - SEPT. 7 – SEPT. 13

Topics:

Test Design and Development (cont.)

Developing test specifications (cont.)
Conducting job and practice analyses
Defining a test framework

Readings in Standards for Educational and Psychological Testing:

Part II-4. Standards 4.4 - 4.6 (pp. 86-87)

Readings in Handbook of Test Development:

Raymond, M. R. (2016). Job analysis, practice analysis, and the content of credentialing examinations. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.). (2016). Handbook of test development (pp. 144-164). New York, NY: Taylor & Francis.

Other Readings (online):

Knapp, J. E., & Knapp, L. G. (1995). Practice analysis: Building the foundation for validity. In J. C. Impara (Ed.), Licensure testing: Purposes, procedures and practices (pp. 93-116). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

Skim **one** of these practice analyses:

The National Council of Architectural Registration Boards. (2013). 2012 NCARB practice analysis of architecture. Washington, DC: Author. Retrieved from:
https://www.ncarb.org/sites/default/files/2013PA_BoxSet_AllReports.pdf

The National Council of State Boards of Nursing, Inc. (2014, January). 2014 RN Practice Analysis: Linking the NCLEX-RN® Examination to Practice U.S. and Canada (NCSBN Research Brief Vol. 62). Chicago, IL: National Council of State Boards of Nursing, Inc. Retrieved from
https://www.ncsbn.org/15_RN_Practice_Analysis_Vol62_web.pdf (no need to read all attachments)

[OPTIONAL] If you have time and would like to read some additional articles by Mark Raymond on how to conduct job and practice analyses, here are three. The NCME instructional module is particularly good.

Raymond, M. R. (2001). Job analysis and the specification of content for licensure and certification examinations. Applied Measurement in Education, *14*(4), 369-415.

Raymond, M. R. (2002). A practical guide to practice analysis for credentialing examinations. Educational Measurement: Issues and Practice, *21*(3), 25-37.

Raymond, M. R. (2005). An NCME instructional module on developing and administering practice analysis questionnaires. Educational Measurement: Issues and Practice, *24*(2), 29-42.

Assignments Due at End of Week:

Feel free to create a thread in the “Week 3 Forum” discussion board if you have any questions about the readings.

Complete the Week 3 Formative Assessment by 11:59 PM Chicago (Central) time on Sept. 13.

WEEK 4 - SEPT. 14 – SEPT. 20

Topics:

Test Design and Development (cont.)

Writing items and scoring procedures
Item selection procedures; assembling item pools
Expert reviews of items for content concerns
Sensitivity/fairness review procedures
Field testing/pilot testing procedures
Determining the psychometric properties of items and of the test
Test assembly
Writing instructions to test takers
Writing standardized procedures for test administration
Conducting studies of differential item functioning

Readings in Standards for Educational and Psychological Testing:

Part II-4. Test Design and Development, Standards 4.7 – 4.25, pp. 87-93

Readings in Handbook of Test Development:

N/A

Other Readings (online):

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 324-353). Westport, CT: Praeger Publishers. **(Note: Read the sections of this chapter entitled “Item Development,” “Item Review,” “Item Evaluation and Test Assembly,” “Test Review,” “Item Banking,” and “Quality Control.”)**

Plake, B. S. (1995). Differential item functioning in licensure tests. In J. C. Impara (Ed.), Licensure testing: Purposes, procedures and practices (pp. 205-218). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

[OPTIONAL] If you have additional interest in item development, watch one or more of these videos:

Registrar Accreditation Board and The Quality Society of Australasia International. (2011). Item writing, part 1 [video]. Milwaukee, WI: Author. Available from <http://www.youtube.com/watch?v=ePORGqoYpkQ&feature=related>

Registrar Accreditation Board and The Quality Society of Australasia International. (2011). Item writing, part 2 [video]. Milwaukee, WI: Author. Available from <http://www.youtube.com/watch?v=U1KBjxrrSmU&feature=related>

Registrar Accreditation Board and The Quality Society of Australasia International. (2011). Item review, part 2 [video]. Milwaukee, WI: Author. Available from http://www.youtube.com/watch?v=WCrct1COQBE&feature=mfu_in_order&list=UL

Assignments Due at End of Week:

1. Reply to one or more threads in the “Week 4 Forum” discussion board. Feel free to create a thread if you have a question of your own.
2. ***MEMO #1 is due by 11:59 PM Central (Chicago time) on Sept. 20***

WEEK 5 – SEPT. 21 – 27

Topics:

Validity

Definitions of validity

Establishing a conceptual framework for investigating validity

Making decisions about types of validity evidence to gather

Articulating propositions to test and conducting logical analyses to evaluate those propositions

Sources of validity evidence

Integrating validity evidence

Developing a “validity argument”

Construct underrepresentation, construct-irrelevant variance

Readings in Standards for Educational and Psychological Testing:

Part I-1. Validity, pp. 11-31 (Standards 1.0 – 1.25)

[OPTIONAL] Readings in Handbook of Test Development:

Kane, M. T. (2016). Validation strategies. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.). (2016). Handbook of test development (pp. 64-79). New York, NY: Taylor & Francis.

Other Readings (online):

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 17-64). Westport, CT: Praeger Publishers.

[OPTIONAL] If you want to read more about content-related validity and its role in testing, see:

Kane, M. T. (2006). Content-related validity evidence in test development. In S. M. Downing & T. M. Haladyna (Eds.), Handbook of test development (pp. 131-153). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

[OPTIONAL] If you want to read more about the argument-based approach to validity, here's an excellent resource:

Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? Educational Measurement: Issues and Practice, *29*(1), 3-13.

[OPTIONAL] If you want to read more about the consequences of test score use as validity evidence, here's a useful resource:

Nichols, P. D., & Williams, N. (2009). Consequences of test score use as validity evidence: Roles and responsibilities. Educational Measurement: Issues and Practice, *28*(1), 3-9.

[OPTIONAL] If you want to learn more about validation in a licensure/certification context, this is a GREAT resource. I re-read this article frequently:

Kane, M. T. (1994). Validating interpretive arguments for licensure and certification examinations. *Evaluation & the Health Professions*, 17(2), 133-159. doi: 10.1177/016327879401700202

Assignments Due at End of Week:

Reply to one "Application Question" in the "Week 5 Forum" discussion board. Note: Your response for this question should be a little longer, as each application question has a few components. Feel free to create a thread if you have a question of your own.

WEEK 6 – SEPT. 28 – OCT. 4

Topics:

Reliability/Precision and Errors of Measurement

Definitions of reliability and measurement error
Flexibility in large-scale testing and increasing measurement error
Characteristics of measurement error
Summarizing reliability data
Interpreting reliability data

Readings in Standards for Educational and Psychological Testing:

Part I-2. Reliability/Precision and Errors of Measurement, pp. 33-47 (Standards 2.0 – 2.20)

Readings in Handbook of Test Development:

NONE

Other Readings (online):

Parkes, J. (2007) Reliability as argument. Educational Measurement: Issues and Practice, 26(4), 2-10.

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

2002 GED Technical Manual (pp. 45-62 only)

[OPTIONAL] If you'd like more information on reliability, read this too:

Miller, M.D., Linn, R.L., Gronlund, N.E. (2008). Reliability and other desired characteristics. Measurement and Assessment in Teaching. (10th ed., p. 107-137). Upper Saddle River, New Jersey: Merrill.

Assignments Due at End of Week:

Feel free to create a thread in the "Week 6 Forum" discussion board if you have any questions about the readings.

Complete the Week 6 Formative Assessment by 11:59 PM Chicago (Central) time on Oct. 4.

Topics:

Test Administration, Scoring, Reporting, and Interpretation

Adhering to standardized test administration procedures
Monitoring test takers to eliminate opportunities for cheating
Managing testing environments
Protecting the security of test materials
Training scorers
Assuring accuracy of scoring
Providing appropriate interpretations of scores
Protecting confidentiality when transmitting scores

Supporting Documentation for Tests

Providing appropriate materials for interpreting test scores
Preparing technical documentation to support tests
Preparing descriptive materials to inform examinees about the nature and content of a test

Readings in Standards for Educational and Psychological Testing:

Part II-6. Test Administration, Scoring, Reporting and Interpretation, pp. 111-121 (Standards 6.0 – 6.16)

Part II-7. Supporting Documentation for Tests, pp. 123-129 (Standards 7.0 – 7.14)

Readings in Handbook of Test Development:

McCallin, R. C. (2016). Test administration. In S. Lane, M. R. Raymond & T. M. Haladyna (Eds.) Handbook of test development 2nd edition (pp. 567-584). New York, NY: Routledge.

Zenisky, A. L., & Hambleton, R. K. (2016). A model for good practice in score reporting. In S. Lane, M. R. Raymond & T. M. Haladyna (Eds.) Handbook of test development 2nd edition (pp. 585-602). New York, NY: Routledge.

Other Readings (online):

[OPTIONAL] If you want to read about quality control monitoring procedures, here's a useful instructional module:

Allalouf, A. (2007). Quality control procedures in the scoring, equating, and reporting of test scores. Educational Measurement: Issues and Practice, 26(1), 36-43.

[OPTIONAL] More information on test administration, security, scoring, and reports:

Cohen, A. S., & Wollack, J. A. (2006). Test administration, security, scoring, and reporting. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 355-386). Westport, CT: Praeger Publishers.

Assignments Due at End of Week:

Reply to one or more threads in the “Week 7 Forum” discussion board. Feel free to create a thread if you have a question of your own.

WEEK 8 – OCT. 12 – OCT. 18

Topics:

Scores, Scales, Norms, Score Linking, and Cut Scores – Part 1

Scaling

Norming

Equating

Linking tests that measure different constructs to establish score comparability

Readings in Standards for Educational and Psychological Testing:

Part II-5. Scores, Scales, Norms, Score Linking, and Cut Scores, pp. 95-100 (stop reading where “cut scores” section starts)

Standards 5.0 – 5.20, pp. 102-107

Readings in Handbook of Test Development:

NONE

Other Readings (online):

Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 155-186). Westport, CT: Praeger Publishers.

Choose one of the following two chapters on equating to read. (The Holland and Dorans article is more technical; the Kolen and Brennan article is more conceptual):

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 187-220). Westport, CT: Praeger Publishers.

Kolen, M. J., & Brennan, R. L. (1995). Introduction and concepts. In Test equating: Methods and practices (pp. 1-27). New York: Springer-Verlag, Inc.

[OPTIONAL] If you want to read more about how to carry out various methods of scaling, here’s an excellent reference:

Tong, Y., & Kolen, M. J. (2010). Scaling: An ITEMS module. Educational Measurement: Issues and Practice, 29(4), 39-48.

[OPTIONAL] If you want to read a basic “how to” on equating methods, here is a good resource (though somewhat dated now):

Cook, L. L., & Eignor, D. R. (1991). IRT equating methods (An NCME instructional module). Educational Measurement: Issues and Practice, 10, 191-199.

[OPTIONAL] If you want to read more about the challenges of linking exams, here are three useful resources:

National Research Council. (1999). Technical aspects of links. In Uncommon measures: Equivalence and linkage among educational tests (pp. 20-47). Washington, DC: National Academy Press.

Holland, P. W. (2005). Assessing the validity of test linking: What has happened since Uncommon measures? In C. A. Dwyer (Ed.), Measurement and research issues in a new accountability era (pp. 185-195). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kolen, M. J. (2001). Linking assessments effectively: Purpose and design. Educational Measurement: Issues and Practice, 20(1), 5-19.

Assignments Due at End of Week:

Reply to one threads in the “Week 8 Forum” discussion board. This week’s discussion board activity requires you to review a technical report and answer some questions about it. See the folder *Weekly Course Content and Readings > Week 8 > Materials for Discussion Board Post*.

As always, feel free to create a thread if you have a question of your own.

WEEK 9 – OCT. 19 – OCT. 25

Topics:

Scores, Scales, Norms, Score Linking, and Cut Scores

Establishing defensible cut scores

Readings in Standards for Educational and Psychological Testing:

Part II-5. Scores, Scales, Norms, Score Linking, and Cut Scores, pp. 100-101
Standards 5.21 – 5.23, pp. 107-109

Readings in Handbook of Test Development:

Cizek, G. J. & Earnest, D. S. (2016). Setting performance standards on tests. In S. Lane, M. R. Raymond & T. M. Haladyna (Eds.) *Handbook of test development* 2nd edition (pp. 212-237). New York, NY: Routledge.

Other Readings (online):

Plake, B. S. (2008). Standard setters: Stand up and take a stand! *Educational Measurement: Issues and Practice*, 27(1), 3-9.

Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425-461. doi: 10.2307/1170678 (this one is long butt has great info.! Intro. video for this week let's you know which areas you can skim over)

[OPTIONAL] More overview of standard setting:

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433-470). Westport, CT: Praeger Publishers.

[OPTIONAL] Watch this video:

Registrar Accreditation Board and The Quality Society of Australasia International. (2011). *Cut score study, Part 2* [video]. Milwaukee, WI: Author. Available from http://www.youtube.com/watch?v=-P-ZuMgQ_XA&feature=mfu_in_order&list=UL

[OPTIONAL] This NCME Instructional Module might also be useful:

Cizek, G., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods (An NCME instructional module). *Educational Measurement: Issues and Practice*, 23(4), 31-50.

Assignments Due at End of Week:

Feel free to create a thread in the “Week 9 Forum” discussion board if you have any questions about the readings.

Complete the Week 9 Formative Assessment by 11:59 PM Chicago (Central) time on Oct. 25.

WEEK 10 – OCT. 26 – NOV. 1

Topics:

Fairness in Testing

General views of fairness

Threats to fair and valid interpretations of test scores

Minimizing construct-irrelevant components through test design and testing adaptations

Readings in Standards for Educational and Psychological Testing:

Part I-3. Fairness in Testing, pp. 49-72 (Standards 3.0 – 3.20)

Readings in Handbook of Test Development:

TBD

Other Readings (online):

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 221-256). Westport, CT: Praeger Publishers.

SKIM these next two documents:

Educational Testing Service. (2009). ETS guidelines for fairness review of assessments. Princeton, NJ: Author. Retrieved from http://www.ets.org/Media/About_ETS/pdf/overview.pdf

Joint Committee on Testing Practices. (2004). Code of fair testing practices in education. Washington, DC: Joint Committee on Testing Practices, Science Directorate, American Psychological Association. Retrieved from <http://aac.ncat.edu/Resources/documents/Code%20Final%20Edit%209-02revFINAL12Wall.pdf>

[OPTIONAL] Lots more great testing fairness resources from ETS are available here:

http://www.ets.org/understanding_testing/test_development/quality_fairness/

Assignments Due at End of Week:

Reply to one or more threads in the “Week 10 Forum” discussion board. Feel free to create a thread if you have a question of your own.

Topics:

The Rights and Responsibilities of Test Users

Test security
Access to test results and to understandable information about a test
Testing irregularities
Informed consent requirements
Protection from improper disclosure
Forms of cheating

Psychological Testing and Assessment

Test selection and administration
Test score interpretation
Collateral information used in psychological testing and assessment
Types of psychological testing and assessment
Purposes of psychological testing and assessment

Uses of Tests for Program Evaluation, Policy Studies, and Accountability

Evaluation of programs and policy initiatives
Test-based accountability systems
Issues in program and policy evaluation and accountability
Additional considerations

Readings in Standards for Educational and Psychological Testing:

Part II-9. The Rights and Responsibilities of Test Users, pp. 139-148 (Standards 9.0 – 9.23)

Part III-10. Psychological Assessment and Testing, pp. 151-168 (Standards 10.1 – 10.18)

Part III-13. Uses of Tests for Program Evaluation, Policy Studies, and Accountability, pp. 203-213 (Standards 13.1 - 13.9)

Readings in Handbook of Test Development:

NONE

Other Readings (online):

Cizek, G. J. (1999). Detecting cheating on tests. In Cheating on tests: How to do it, detect it, and prevent it (pp. 127-150). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Cizek, G. J. (1999). Responding to cheating. In Cheating on tests: How to do it, detect it, and prevent it (pp. 151-162). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. Applied Measurement in Education, 17(2), 145-220.

Hambleton, R. K. (2007, June 18). A new challenge: Making test score reports more understandable and useful [PowerPoint presentation].

SKIM this document:

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. Washington, DC: Author. Retrieved from <http://www.apa.org/ethics/code/index.aspx>

[OPTIONAL] I have posted on Blackboard a number of direct links to YouTube videos that students have produced showing various methods of cheating, if you are interested in viewing them.

[OPTIONAL] If you are interested in reading about qualifications that the APA considers important for appropriate use of psychological tests, here is a resource:

Practice and Science Directorates. (2000, August). Report of the task force on test user qualifications. Washington, DC: American Psychological Association. Retrieved from <http://www.apa.org/science/programs/testing/qualifications.pdf>

Assignments Due at End of Week:

Reply to one or more threads in the “Week 11 Forum” discussion board. Feel free to create a thread if you have a question of your own.

Topics:

Technology and Testing

Technology and test design
Automatic item generation
Item pretesting
Test assembly and packaging
Examinee registration, scheduling, and accommodations
Test delivery models
Systems for scoring responses
Post-administration analysis systems
Reporting scores

The Rights and Responsibilities of Test Takers

Testing Individuals with Disabilities

Accommodation issues when testing individuals with disabilities
Strategies of test modification
Using modifications in different testing contexts
Reporting scores on modified tests

Testing Individuals of Diverse Linguistic Backgrounds

Test translation, adaptation, and modification
Issues of equivalence
Language proficiency testing
Testing bilingual individuals
Administration and examiner variables
Use of interpreters in testing
Cultural differences and individual testing

Readings in Standards for Educational and Psychological Testing:

Part II-8. The Rights and Responsibilities of Test Takers, pp. 131-137 (Standards 8.0 – 8.12)

Readings in Handbook of Test Development:

N/A

Other Readings (online):

Dragow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 471-515). Westport, CT: Praeger Publishers.

Coltrane, B. (2002, November). English language learners and high-stakes tests: An overview of the issues. Educational Resources Information Center (ERIC) Digest, EDO-FL-02-07, pp. 1-4.

Menken, K. (2000, September). What are the critical issues in wide-scale assessment of English language learners? National Clearinghouse for Bilingual Education Issue & Brief, No. 6, pp. 1-7.

Test Taker Rights and Responsibilities Working Group of the Joint Committee on Testing Practices. (1998). Rights and responsibilities of test takers: Guidelines and expectations. Washington, DC: American Psychological Association. Retrieved from <http://www.apa.org/science/programs/testing/rights.aspx>

[OPTIONAL] See blackboard content folder entitled “Additional Optional Readings (not listed in syllabus)” for MANY additional resources on these topics. Too many to list them all here!

Assignments Due at End of Week:

MEMO #2 is due by 11:59 PM Central (Chicago time) on Nov. 15

Feel free to create a thread in the “Week 12 Forum” if you have a question or comment about the readings. This week, discussion board engagement is not required (turning in your memo on time will get you all the participation points for Week 12!)

Topics:

Workplace Testing and Credentialing

Employment testing

Testing in professional and occupational credentialing

Key Court Cases, Laws, and Legislation Affecting Licensure and Certification Programs

Readings in Standards for Educational and Psychological Testing:

Part III-11. Workplace Testing and Credentialing, pp. 169-182 (Standards 11.1 – 11.16)

Readings in Handbook of Test Development:

NONE

Other Readings (online):

Clauser, B. E., Margolis, M. J., & Case, S. M. (2006). Testing for licensure and certification in the professions. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 701-731). Westport, CT: Praeger Publishers.

Phillips, S. E., & Camara, W. J. (2006). Legal and ethical issues. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 733-755). Westport, CT: Praeger Publishers.

Liptak, A. (2009, June 29). Supreme Court Finds Bias Against White Firefighters. *The New York Times*. Retrieved from http://www.nytimes.com/2009/06/30/us/30scotus.html?_r=0

SKIM this web page:

U.S. Department of Labor, Office of Federal Contract Compliance Programs, Equal Employment Opportunity. (1978). Uniform guidelines on employee selection procedures. Washington, DC: Author. Retrieved from <http://www.uniformguidelines.com/uniguideprint.html>

[OPTIONAL] If you want to read more about legal issues in licensure and certification testing, here are some excellent resources:

Carson, J. D. (2001). Legal issues for standard setting for licensure and certification. In G. J. Cizek (Ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 427-444). Mahwah, NJ: Lawrence Erlbaum Associates.

Mehrens, W. A. (1995). Legal and professional bases for licensure testing. In J. C. Impara (Ed.), Licensure testing: Purposes, procedures, and practices (pp. 33-58). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

Assignments Due at End of Week:

Reply to one or more threads in the “Week 11 Forum” discussion board. Feel free to create a thread if you have a question of your own.

WEEK 14 – NOV. 23 – NOV. 29



No readings or assignments – be thankful!

WEEK 15 – NOV. 30 – DEC. 6

Topics:

Educational Testing and Assessment

Design and development of educational assessments
Use and interpretation of educational assessments
Administration, scoring, and reporting of educational assessments

Key Court Cases, Laws, and Legislation Affecting Large-Scale Testing Programs

Ethics and Large-Scale Testing

Readings in Standards for Educational and Psychological Testing:

Part III-12. Educational Testing and Assessment, pp. 183-201 (Standards 12.1 – 12.19)

Readings in Handbook of Test Development:

NONE

Other Readings (online):

Choose one of these two chapters to read:

Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 531-578). Westport, CT: Praeger Publishers.

Zwick, R. (2006). Higher education admissions testing. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 647-679). Westport, CT: Praeger Publishers.

Choose one of these two resources to read:

Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. Educational Measurement: Issues and Practice, 25(3), 27-34.

Parkes, J., & Stevens, J. J. (2003). Legal issues in school accountability systems. Applied Measurement in Education, 16(2), 141-158.

SKIM these next two documents:

American Educational Research Association. (2000). Position statement on high-stakes testing in pre K-12 education. Washington, DC: Author. Retrieved from <http://www.aera.net/About-AERA/AERA-Rules-Policies/Association-Policies/Position-Statement-on-High-Stakes-Testing>

American Educational Research Association. (1992). Ethical standards of the American Educational Research Association. Washington, DC: Author. Retrieved from http://www.aera.net/uploadedFiles/About_AERA/Ethical_Standards/EthicalStandards.pdf

[OPTIONAL] If you want to read about legal issues in assessing students with disabilities, here are some excellent resources:

Phillips, S. E. (2002). Legal issues affecting special populations in large-scale testing programs. In G. Tindal & T. M. Haladyna (Eds.), Large-scale assessment programs for all students: Validity, technical adequacy, and implementation (pp. 109-148). Mahwah, NJ: Lawrence Erlbaum Associates.

Shriner, J. G. (2000). Legal perspectives on school outcomes assessment for students with disabilities. Journal of Special Education, 33(4), 232-239.

Green, P.C., & Sireci, S. G. (1999). Legal and psychometric issues in testing students with disabilities. Journal of Special Education Leadership, 12(2), 21-29.

Assignments Due at End of Week:

Feel free to create a thread in the "Week 15 Forum" if you have a question or comment about the readings. This week, you will receive participation points automatically. 😊

WEEK 16 (FINAL EXAM WEEK) – DEC. 7 - 11

Topics:

No new readings.

Assignments Due at End of Week:

FINAL EXAM (exact dates are TBD and will be communicated to you. You should expect that over a period of 7 days, you'll need to set aside about 5 hours to take the exam in one sitting.)