

EDPSY 592: Advanced Educational Measurement Spring Quarter 2004

Instructor: Min Li
Office: 312D Miller

Phone: 616-6305
e-mail: minli@u.washington.edu

TA: Brad Wu
Office Hours: Thurs 2-4pm, by appointment.

e-mail: ccwu@u.washington.edu

Text: M. J. Allen & W. M. Yen (2002). Introduction to measurement theory. Prospect Heights, IL: Waveland Press.

COURSE DESCRIPTION

Tests are used extensively in education, counseling, health, employment, and mental health settings. Professional organizations have been writing "guidelines" and "standards" for test development and use for more than 40 years. Still, one of the major problems related to testing is the *misuse* and *overuse* of tests to make decisions about human beings. The purpose of this course is to help students of education, psychology, or other social science fields understand the classical test theory and its applications to the real world of testing. Specifically, you will learn (a) the concepts of reliability and validity, (b) the mathematical models underlying these concepts, and (c) the application of these concepts to problems in measurement. Though the examples I use are paper-and-pencil measures of cognitive abilities and academic achievement, most of the concepts and methods developed apply equally to performance testing, as well as the assessment of attitudes and personality constructs, ratings based on systemic observations, and other kinds of measurements of individuals or groups.

For this course, I hope to create a hands-on environment that balances the quantitative foundations of classical test theory with the real world applications and issues that arise from testing.¹ As such, we may spend less time drilling quantitative methods. Instead, more time is spent on class discussions and small group explorations. My task will be to make certain that the in-class work and external assignments are useful to you as professionals. Your job is to do the readings (including supplemental articles) and work with each other to explore and engage with the concepts of the course. The assignments for the course are designed for yourselves to enhance your understanding of the concepts rather than for me to test you. There will be three basic types of assignments:

- ♦ Readings and homework assignments
- ♦ Small group explorations/discussions of issues and problems
- ♦ Class projects

The following pages include the course learning objectives, grading policy, and the course schedule.

¹ This course is not designed for students with exceptionally strong mathematical and statistical preparation, but such students may nonetheless desire a systematic introduction to measurement theory. If you fall into this category, please see me individually to arrange some special projects in connection with the course, such as a critique of measurement methods in published research, a critique of the measurement of some psychological construct, or a review of one or more published tests.

COURSE GOALS AND OBJECTIVES

1) **Students will understand classical test theory and the methods of data gathering and analysis that follow from it.**

- a. Students will learn the fundamental principles of classical test theory.
- b. Students will learn how to obtain evidence for the validity of assessments.
- c. Students will learn how to obtain evidence for the reliability of assessments.
- d. Students will learn how to use item level data to evaluate items and to create reliable and valid tests.
- e. Students will learn the meanings of different types of derived test scores.
- f. Students will learn how to interpret standardized test score reports.

2) **Students will become empowered to make thoughtful, ethical, and reasonable decisions about assessments and assessment practices.**

- a. Students will learn to discriminate between appropriate and inappropriate uses of assessment information.
- b. Students will learn how assessments can be used to support or adversely affect examinees.
- c. Students will learn appropriate roles for educators and psychologists in administering and interpreting tests.

3) **Students will understand origins and criticisms of classical test theory and prevailing theories of individual differences.**

- a. Students will learn the origins of intelligence testing.
- b. Students will learn the major criticisms of the use of test theory to measure psychological variables.

GRADING POLICY

Grading in this course will be based on your learning of the targeted objectives (above). A grade of 3.3 or better will mean that you can have the capacity of learning about new tests; that you can participate (in an informed way) in debates and decision-making situations related to tests; and that you understand the foundational assumptions behind most contemporary testing practices. Grades will be based on the class projects and your participation in class work related to the readings. Periodically, there may be small exercises assigned in class or as homework for me to check your understanding, but these will be non-graded.

You are expected to complete two individual class projects (or some identifiable parts of a group project), one with an emphasis on reliability issues by conducting item and test analysis and another with an emphasis on validity issue - test review and construction. Both projects are intended to relate materials we discuss in the course to real problems. You are strongly encouraged to bring your own research projects or datasets. All project ideas must be approved by the instructor. Below are some brief descriptions of the possible projects you can select.

Project 1: Item Analysis and Test Analysis (40% of grade)

The purpose of this project is for you to get a glimpse of what life is like for test developers. You will look at data for a set of test items and decide which items you would like to put together for a test. For example, you can estimate the reliability of the test, difficulty index for each item,

describe the decision-making process you went through to select items for your test, and discuss the quality of your test given the item and reliability data as well as the content represented by the items. Or you will conduct G-studies. For example, you can estimate the Generalizability coefficients for a math achievement test, the conduct D-studies how many raters and tasks are needed in order to obtain a reliable estimate for student math achievement.

Project 2: Test Review and Construction Project (50% of grade)

The purpose of the project is to give you a chance to integrate all the ideas we talk about and apply them in a meaningful way in your field. You can review two tests, or construct one test of your choosing, or conduct a validity study related to test items. For example when reviewing the test, you need to provide the items, describe the technical information regarding evidence for the reliability and validity of the test scores as well as studies that have been done to assess the bias of items or test scores, and critique the test from the literature and your own thinking. For test construction, you need to provide the theoretical framework for your constructs, the items you develop, pilot your items with a reasonable sample size, and examine and discuss the reliability and validity of the test scores. For validating test items, you need to provide the theoretical framework for the constructs and conduct a study using empirical data, quantitative or qualitative, to validate the constructs. Sometimes the tests you select may not include enough information for you to conduct this project or the test you develop may need several testing occasions, so be sure to start early and allow yourself to have enough time to improve your work!

Another part of this project is to prepare and offer a brief class presentation at the 9th-10th week. The write-up will account for 40% of grade whereas the presentation for 10%.

The reports for both projects should be research papers in APA style (double space and no more than 30 pages). The first project will be due on 4/29 and the second will be due on 6/7.

Class Participation (10% of grade)

As one who was a wall-flower in college and graduate school, I know that this part of the grade is the most discomfoting for some people. Nevertheless, one of the ways that you will learn best is if you attend class and get actively involved. Class participation will be evaluated on a rating scale each week. The criteria for class participation are:

- ♦ Asks questions when confused by the presentations or demonstrations
- ♦ Comes to class daily unless prior arrangements are made or a family/personal emergency occurs
- ♦ Contributes ideas and personal thoughts about the topics discussed as a whole class
- ♦ Contributes substantive ideas to small group work
- ♦ Shows some understanding of the concepts under investigation or the content of the readings through group work or class discussion

Scores will be assigned as follows:

4.0 Meets all relevant criteria;

3.7 Meets only four of the criteria;

3.4 Meets only three criteria, e.g., *typically* waits until after class to bring up topic related questions and issues;

3.0 Meets only two criteria, e.g., does not participate in class discussions, does not bring up

questions and issues;

2.0 Meets one criterion, e.g., attends class but does not contribute to small or large group discussion.

Grade Summary Example

If a student receives the scores in the table below for assignments and participation, the grading summary process and final grade would be as follows:

SOURCE	SCORE	WEIGHT	WEIGHTED VALUE
Item/Test Analysis	3.5	40%	1.40
Test Review/construction	4	50%	2.00
Participation	3.7	10%	0.37
Grade			3.77 ~ 3.8

COURSE SCHEDULE

DATE	TOPICS	READINGS & ASSIGNMENTS
4/1	Introduction/overview; Covariance, correlation; linear combinations of variables	Ch1; Ch2
4/8	Test construction and item analysis (1); Classical true score model	Ch 6.1-6.3, 6.7, 6.8; Ch3
4/15	AERA week; no class	
4/22	Reliability	Ch 4
4/29	Introduction to Generalizability theory; G theory	Ch 10.7; Shavelson & Webb (1991)
5/6	Validity (1)	Ch 5; Messick (1989); Linn (1997)
5/13	Validity (2)	Ch 9; Ch 10.4-10.5
5/20	Test scores; Sharing ideas for 2 nd project	Ch 7;
5/27	Test construction and item analysis (2) Bias in testing;	Ch 6.4-6.6, 6.8 (review); 6.9; Shepard (1982);
6/3	Item Response theory, basic principles Class presentation of final projects	Ch 10.6; Ch 11