

ERSH 4600/6600 Applied Educational Assessment  
Spring, 2003: Tuesday & Thursday 9:30-10:45

**Instructor:** Dr. Deborah Bandalos

**Office/phone/email:** 325V Aderhold; 542-4540; dbandalo@coe.uga.edu

**Office hours:** T & Th, 1-2 and by appointment

**Text:** *Measurement and Assessment in Teaching, 8<sup>th</sup> edition*, Robert L. Linn.

**Supplementary texts:**

*Measurement and Evaluation in Psychology and Education, 6<sup>th</sup> edition*, Robert M. Thorndike.

*Knowing what Students Know: The Science and Design of Educational Assessment*, National Research Council.

Date	Topic	Reading*
1/9	Course Introduction	
1/14	Introduction to the measurement process	L: chapter 1 & 2
1/16	Types of scores	L: chapter 19; Bond, 1996
1/21	Measuring learning outcomes	L: chapter 6; NRC, chapter 2
1/23	Objective test items	L: chapters 7 & 8
1/28	Criticisms of traditional cognitive testing	Wiggins, 1989; Fredericksen, 1984; Nickerson, 1989
1/30	Cognitive psychology and assessment	NRC: chapter 3
2/4	Measuring complex achievement	L: chapter 9
2/6	Measuring complex achievement	NRC: chapters 5 & 6
2/11	Essay questions	L: chapter 10
2/13	Performance-based assessments/Portfolios	L: chapters 11 & 12
2/18	Item analysis: cognitive	L: chapter 14
2/20	Affective measurement	L: chapter 13
2/25	Exam 1	
2/27	Affective measurement	T: chapter 12
3/4	Item analysis for affective measures	
3/6	Item analysis con't	
3 /11	Reliability	L: chapter 5
3/13	Reliability con't	
3/18	SPRING BREAK	
3/20	SPRING BREAK	

Date	Topic	Reading*
3/25	Reliability con't	
3/27	Validity	L: chapter 4
4/1	Validity con't	
4/3	Changing conceptualizations of validity Exam 2 due	Hubley & Zumbo, 1996; Anastasi, 1986
4/8	Changing conceptualizations con't	Shephard, 1993
4/10	Legal and equity issues	T: chapter 13; Childs, 1990; Hambleton & Rodgers, 1995
4/15	Legal and equity issues con't	Bersoff, 1981; 1984
4/17	Ethics and social considerations	T: chapter 14; Haney, 1981
4/22	Interest and personality assessment	T: chapter 11
4/24	Aptitude tests	L: chapter 17
4/29	Standardized achievement tests	L: chapter 16
5/1	Last class: Project 2 poster presentations	

\* I reserve the right to assign other readings at my discretion.

L: Linn; NRC: National Research Council; T = Thorndike

## EVALUATION

The final grades for this course will be based on a weighted average of grades on exams and projects. Exam 1 will be composed of questions submitted by students. I will review all questions submitted and select questions on the basis of technical merit as discussed in class. The second exam will be a take-home exam. The three major projects are discussed below. However, additional smaller projects or homework assignments may be assigned at my discretion.

Weights for final grades will be as follows:

Exams	40% (20% each)
Projects	60% (20% each)

In order to pass the course, all projects must be completed.

## GENERAL COURSE GUIDELINES

Much of class time will be devoted to a discussion of the readings. Therefore it is essential that all students read and come prepared to discuss the readings assigned for that day. Students may be called on randomly to comment on the day's readings.

Because much of the work for this class will be done in a group format, it is important for students to be conscientious in guarding against activities that could be interpreted as cheating behavior. Activities such as working in groups, discussing one's work with other students, and sharing ideas **are encouraged** in this class and are **not** considered to constitute cheating. Cheating is defined as copying work from another student, either past or current, or any misrepresentation of another's work as your own. **This includes copying from class notes, readings, or texts on any assignment or exam.** If you are asked for an evaluation or opinion on an assignment or exam, you should give it **in your own words**. Students found to engage in cheating behavior will receive a grade of zero for the assignment on which the cheating occurred and possibly a failing grade in the course. In addition, the behavior will be reported to the student's advisor and other University authorities.

## PROJECTS

The projects in this class have been designed to give you hands-on experiences with writing and evaluation different types of measurement. I hope that this will provide a basis for class discussion as we observe and discuss issues arising from these measurement projects.

### PROJECT 1

Project 1 is a group project in which you will develop and evaluate an assessment of a complex higher order thinking process in some content area. Examples of such assessments are given in Linn, chapter 9 and in chapters 3, 5, and 6 of the NRC book. To keep this project manageable, your group should carefully delineate the content area and cognitive processes of interest. The assessment should be targeted to the specific content and cognitive areas you have delineated. Because you will be required to actually administer the assessment to several people, you should choose a content area with which people you know or have access to would be familiar. After developing and administering your assessment you will be required to evaluate it in terms of practicality, reliability, and validity.

### PROJECT 2

Project 2 is a group project in which you will actually construct and evaluate a short attitude scale. Each group will decide on an area for measurement and group members will write items for the scale. These items should reflect good item writing principles as discussed in class. In addition to these items, the questionnaire should include four or five items designed to elicit demographic information, such as age or gender. Upon completion of the questionnaire, each group member will be required to administer it. Each group should plan to administer their scale to at least 50 subjects. The subjects need not be chosen randomly.

Each group will be required to complete a set of analyses using the data that is collected on their group's scale. Different sections of the project will be turned in throughout the semester, as the material is covered in class. Groups may redo any portion for which they do not receive all the points possible. Each group is also required to prepare a poster showing their results. These will be shared with the class on the final day of class.

The sections of Project 2 are:

1. A brief description of the affective area that was assessed, as well as any demographic information that was collected.
2. A brief description of the persons who filled out the questionnaire, and a description of the conditions under which the questionnaire was administered. Mention in particular any occurrence during administration that you feel may have affected the responses.
3. A copy of the scale, indicating which, if any, items were reverse coded.
4. A frequency distribution for total scores on the scale.
5. The mean, median, and mode of scores on the scale and a comment on the general shape of the distribution.
6. The standard deviation of the scores on the scale.
7. A brief analysis of the descriptive statistics in 4-6 above, indicating any factors that you feel may affect the interpretation of the scores on the scale.
8. An item analysis for the scale, including
  - a) item variance
  - b) corrected item-total correlation
  - c) alpha if item deleted
  - d) your comments on the adequacy of each item
  - e) suggestions for improvement of flawed items

9. Coefficient alpha and the standard error of measurement for the scale.
10. Your comments on the adequacy of the scale's reliability.
11. Possible reasons for low reliability.
12. Your suggestions for improving the scale's reliability.
13. Evidence of validity for the scale, based on the demographic data. Include a discussion of why the type of validity you present is an appropriate type for this scale and why you chose the evidence you did.
14. A discussion of other types of validity evidence you would like to have and why.

### PROJECT 3

Project 3 has been designed to allow you to apply what you have learned in class about test properties by evaluating a standardized test of your choice. At the beginning of the semester, each student will choose a type of test in which they are interested (e.g., intelligence tests, achievement tests, personality tests, etc). The available instruments in this area should then be researched and each student should choose one instrument to evaluate. Project 3 will consist of an assessment of the instrument in terms of reliability, validity, etc. You must therefore choose an instrument on which such information is available.

In order to assess the instrument you will need to obtain a copy of the test manual, and ideally of the test itself. You may be able to obtain these from professors in your department or schools or agencies with whom you are working. They are also available from testing companies, but this will require that you plan ahead as it will take time to order and receive them. Other sources of information include:

the Buross Mental Measurements Yearbooks (these are now available

online)

specialty area Tests and Reviews

your textbook and other textbooks available from me.

A written report, no longer than ten typewritten, double-spaced pages will be due the last day of class. You will be given a detailed protocol describing what information should be included, but to give you an idea, you will be required to address the following:

- 1) Statement of what the test was designed to measure.
- 2) Information on how the test was constructed.
- 3) Instructions for administering the test.
- 4) Reliability evidence given including type of reliability and standard error of measurement.
- 5) Validity evidence given including type of validity.
- 6) A description of the norm groups and year in which each test was normed.
- 7) Type(s) of score(s) reported and ease of interpretation of the results.
- 8) Practical considerations such as cost, availability, special training or facilities needed, availability of alternate forms, time required, guides for using and interpreting the test.

You should also refer to the article "Questions to ask when evaluating tests" by Rudner, that is available on the online journal *Practical Assessment, Research, and Evaluation*.

## Readings

Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology, 37*, 1-15.

Bersoff, D. (1981). Testing and the law. *American Psychologist, 36(10)*, 1047-1056.

Bersoff, D. (1984). Social and legal influences on test development and usage. In Barbara S. Plake (Ed.), *Social and Technical Issues in Testing: Implications for Test Construction and Use*, pp. 87-110. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Bond, L.A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research, and Evaluation, 5(2)*.

Childs, R. (1990). Gender bias and fairness. *Practical Assessment, Research, and Evaluation, 2(3)*.

Fredericksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39(3)*, 193-202.

Hambleton, R. & Rodgers, J. (1995). Item bias review. *Practical Assessment, Research, and Evaluation, 4(6)*.

Haney, W. (1981). Validity, vaudeville, and values: A short history of social concerns over standardized testing. *American Psychologist, 36(10)*, 1021-1034.

Huble, A.M. & Zumbo, B.D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology, 123*, 207-215.

National Research Council. (2001). The nature of assessment and reasoning from evidence. In *Knowing what Students Know: The Science and Design of Educational Assessment*, pp. 37-54. Washington, D.C.: National Academy Press.

National Research Council. (2001). Advances in the sciences of thinking and

learning. In *Knowing what Students Know: The Science and Design of Educational Assessment*, pp. 59-110. Washington, D.C.: National Academy Press.

National Research Council. (2001). Implications of the new foundations for assessment design. In *Knowing what Students Know: The Science and Design of Educational Assessment*, pp. 177-220. Washington, D.C.: National Academy Press.

National Research Council. (2001). Assessment in practice. In *Knowing what Students Know: The Science and Design of Educational Assessment*, pp. 221-260. Washington, D.C.: National Academy Press.

Nickerson, R.S. (1989). New directions in educational assessment. *Educational Researcher*, 18(9), 3-7.

Rudner, L. (1994). Questions to ask when evaluating tests. *Practical Assessment, Research, and Evaluation*, 4(2).

Shepard, L.A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.

Wiggins, G. (May, 1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 703-713.