



EDUC 8804: Measurement in Survey Research Fall 2007

TUE 9:00-11:30 EDUC 334

Instructor

Derek Briggs

Office: EDUC 211 Tel: (303) 492-6320 E-mail: Derek.Briggs@colorado.edu

Office Hours: Tues/Thur 2:30-3:30

Course Overview

“I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.”

Sir William Thomson, Lord Kelvin. *Electrical Units of Measurement*. Popular Lectures and Addresses, Vol 1 of 3. (London: Macmillan, 1889, p. 73-74)

“Whatever exists, exists in some amount.” E. L. Thorndike.

Measurement in survey research is both art and science. While much of the science can be learned from textbooks and articles, the art can only be learned through experience. The principal objective of this course is to give students an introduction to fundamental concepts of measurement through a semester-long project in which students are expected to develop, pilot test, analyze and evaluate their own survey instruments. Though surveys can be used to measure both manifest variables (i.e., factual variables) and latent variables (i.e., inside a person’s head), in this course the empirical focus will be on developing a survey instrument that can measure at least one latent variable that is cognitive in nature. The aim will be to characterize this variable in terms of a *developmental learning progression* that maps out what survey respondents know and can do in a particular content domain.

This course emphasizes the *process* of developing, analyzing and validating a survey instrument. The concept of validity is central. The most comprehensive definition of the term validity in the context of measurement was established by Samuel Messick in the opening paragraph to his chapter (“Validity”) in the 3rd edition of the book *Educational Measurement* (1989):

Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment...the term test score is used generically here in the broadest sense to mean any observed consistency, not just on tests as ordinarily conceived but on any means of observing or documenting consistent behaviors or attributes. Broadly speaking,

validity is an inductive summary of both the existing evidence for and the potential consequences of score interpretation and use.

The first half of the course focuses on the underappreciated task of developing a survey instrument with items that derive from a clearly delineated theory for the construct to be measured. Construct maps (also known as progress variables) are presented as the instantiation of this theory. By the ninth week of the course, students (working together in small groups of 3-4) should have a survey instrument ready for pilot testing with a minimum of 30-40 respondents. In the second half of the course, the focus shifts to the analysis of item responses and putting together a validity argument to support the proposed uses of the survey. Students will be introduced to classical test theory and item response theory as two measurement models used for the former purpose, with a particular emphasis placed on items response theory models that fall with the Rasch family of measurement models. For the latter purpose, students will be introduced to both historical and recent developments in validity theory.

COURSE READINGS

Books to be Purchased

Wilson, M. (2004). *Constructing measures: an item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates. Available at CU Bookstore.

Masters, G. (1996). *Educational Measurement*. Australian Council for Educational Research. Available to be ordered at <http://shop.acer.edu.au/acer-shop/group/ARK>

Masters, G. & Forster, M. (1996). *Progress Maps*. Australian Council for Educational Research. Available to be ordered at <http://shop.acer.edu.au/acer-shop/group/ARK>

Readings Available Electronically (CULearn or Norlin E-Reserves)

AERA/APA/NCME. (1999). Validity. In *Standards for Educational and Psychological Testing*. American Educational Research Association. (9-24).

Alonzo, A. & Steedle, J. (2007). What do students mean by "Force"?: assessing a force & motion learning progression. Paper presented at the annual meeting of the American Educational Research Association, April 2007, Chicago IL.

Angoff, W. H. (1988). Validity: an evolving concept. In *Test Validity*, H. Wainer & H. Braun. Mahwah, NJ: Lawrence Erlbaum Associates. (19-32).

Biggs, J & Collis, K. (1982). *Evaluating the quality of learning*. New York: NY: Academic Press. (1-31)

Bond, T. & Fox, C. (2001). *Applying the Rasch model*. Mahwah, NJ: Lawrence Erlbaum Associates. (Ch. 12, 173-186)

- Briggs, D., Alonzo, A., Schwab, C. & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33-64.
- Crocker, L & Algina, J. (1986). Item analysis. In *Introduction to Modern and Classical Test Theory*. Harcourt Brace Jovanovic. (311-338).
- Cronbach, L. (1988). Five perspectives on the validity argument. In *Test Validity*, H. Wainer & H. Braun. Mahwah, NJ: Lawrence Erlbaum Associates. (3-15).
- Dillman, D., Eltinge, J., Groves, R., & Little, R. (2002). Survey nonresponse in design, data collection, and analysis. In *Survey Nonresponse*, R. Groves, D. Dillman, J. Eltinge, & R. Little. New York: John Wiley & Sons.
- Fowler, F. (1995). *Improving survey questions: design and evaluation*. Thousand Oaks, CA: SAGE Publications. (Chapter 3, 46-77)
- Gorin, J. (2006). Test design with cognition in mind. *Educational Measurement: Issues & Practice*, 25(4), 21-35.
- Hamilton, L. S., Nussbaum, E. M., & Snow, R. E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, 10, 181-200.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.). (pp. 17-64). Westport, CT: American Council on Education/Praeger.
- Leighton, J. & Gierl, M. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues & Practice*, 26(1), 3-16.
- Marton, F. (1988). Phenomenography: Exploring different conceptions of reality. In *Qualitative approaches to evaluation in education*, D. Fetterman. New York: Praeger. (176-205).
- Mislevy, R. & Haertal, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues & Practice*, 25(4), 6-20.
- Nitko, A. J. (1983) Developing multiple-choice items. In *Educational Tests & Measurement*. Harcourt Brace & Co. (Ch. 8, 189-214).
- Openheim, A. N. (1992). Questionnaire design, interviewing and attitude measurement. London: Pinter. (Chapters 10-11, 174-209).
- Patton, M. (1980) Qualitative Interviewing. *Qualitative Evaluation Methods*. Thousand Oaks, CA: SAGE Publications. (pp. 195-263)

- Sapsford, R. (1999) *Survey Research*. Thousand Oaks, CA: SAGE Publications. (Chapters 2-4, 20-99).
- Thompson, B. (2003) Understanding reliability and coefficient alpha, really. In *Score Reliability*, B. Thompson (ed). Thousand Oaks, CA: SAGE Publications. (Chapter 1, 3-30).
- Toulmin, S. (2003). The layout of arguments. In *The uses of arguments*, S. Toulmin. Cambridge: Cambridge University Press. (87-134).
- Wainer, H. and Thissen, D. (2001) True score theory: the traditional method. In *Test Scoring*, H. Wainer & D. Thissen (eds). Lawrence Erlbaum Associates. (Chapter 2, 23-71).
- Wills, G. D. (2005) *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks, CA: SAGE Publications. (Chapters 1, 3-5)

Course Schedule

WEEK 1: Introduction and Overview

Aug 28

- Course Expectations and Guidelines
- A Framework for Survey Research

Assignments:

1. Handout: What research question(s) will your survey be used to address?
2. Visit HRC website (<http://www.colorado.edu/VCRResearch/HRC/index.html>)

If this is your first time, go to

http://www.colorado.edu/VCRResearch/HRC/First_Timer.html

Complete the CITI course. Otherwise, make sure you have read

http://www.colorado.edu/VCRResearch/HRC/FAQs_for_Investigators.html

WEEK 2: Learning Progressions

Sep 4

- Overview of Wilson's Four Building Blocks of Measurement
- Construct modeling approaches to developing a learning progression
- An argument-based approach to validity

Readings: Wilson, Ch. 1; Kane, 1992; Briggs et al, 2006; Alonzo & Steedle, 2007.

Assignments:

Proposal for Survey Project [Due Sep 25]

WEEK 3: General Issues in Survey Research

Sep 11

- Using a survey to measure manifest variables (i.e., factual information)
- Developing a Sampling Frame
- Survey Nonresponse Bias
- Human Subjects Constraints

Readings: Sapsford Ch. 2-4 (pp. 12-99); Dillman et al, 2002

WEEK 4: Construct Maps, Part 1

Sep 18

- Construct Maps as the Instantiation of a Theory
- Prerequisites to Measurement
- Examples of Construct Maps

Reading: Wilson, Ch. 2, Masters & Forster, (pp. 1-31)

WEEK 5: Construct Maps, Part 2

Sep 25

- Designing a survey to measure an affective latent variable
- Cognitive models of development

Proposal for Learning Progression Project Due

Assignment:

Wilson, Ch. 2: Exercises 1-4 [Creating a first draft of a construct map: Due Oct 2]

Readings: Gorin, 2006; Leighton & Gierl, 2007, Mislevy & Haertel, 2006

WEEK 6: Designing Items

Oct 2

- A Taxonomy of Item Formats
- Working with Likert Items

Readings: Wilson, Ch. 3; Fowler, 1995, Ch. 3, Patton, 1980; Openheim, 1992 (pp. 174-186); Nitko, 1983;

Wilson, Ch. 2: Exercises 1-4 Due

Assignment:

Wilson, Ch. 3: Exercises 1-2 [Designing items for your instrument: Due Oct 9]

WEEK 7: Scoring Items: The Outcome Space

Oct 9

- From Construct Maps to Scoring Rubrics
- Phenomenographic approaches
- The SOLO Taxonomy

Readings: Wilson, Ch. 4; Marton (1988); Biggs & Collis (1982), Openheim, 1992 (pp. 187-209)

Wilson, Ch. 3: Exercises 1-2 Due

Assignment:

Wilson, Ch. 4, Exercises 1-2 [Relating item responses to the construct map: Due Oct 16]

WEEK 8: Critiquing Items: The Item Panel

Oct 16

- Think-alouds
- Cognitive Interviews
- Running an Effective Item Panel

Readings: Wills (2005), Ch 1, 4-5; Hamilton et. al, 1997

Assignments:

Conduct Item Panel [i.e., Wilson, p. 59-61]

Highly Recommended: Conduct Think-Aloud

WEEK 9: Classical Test Theory and Classical Item Analysis

Oct 23

- Observed Scores, True Scores and Error
- The Concept of Reliability
- Item P values and point-biserials

Note: At this point in the course you should have a draft of your survey instrument that is ready to be pilot tested.

Readings:

Wainer & Thissen (2001), [23-34, 52-54, 57-59, 64-70]

Crocker & Algina (1986) [311-338]

WEEK 10: Item Response Theory and The Rasch Model

Oct 30

- Probability, odds and log odds (logits)
- Item Characteristic Curves
- The Wright Map

Readings: Wilson, Ch. 5-6; Masters (1996)

Assignment:

Wilson, Ch. 5, Exercises 2-3 [Calculating response probabilities from the Wright Map]

WEEK 11: Reliability

Nov 6

- Cronbach's Alpha
- The Standard Error of Measurement
- Item and Test Information Curves

Readings: Wilson, Ch. 7; Thompson, (2003), 3-30

WEEK 12: Using ConstructMap to Analyze Item Responses (Part 1)

Nov 13

- Importing item response data into ConstructMap
- Running a classical item analysis
- Running the Simple Rasch Model
- Running the Rating Scale Model & Partial Credit Model
- Producing a Wright Map

Readings: ConstructMap Lite User Manual

Assignment:

Produce classical item statistics for your data using ConstructMap and interpret the output

Note: At this point in the course you should be ready to begin analyzing data from the pilot test of your group's survey instrument.

WEEK 13: Using ConstructMap to Analyze Item Responses (Part 2)

Nov 27

- Plotting Item Characteristic Curves
- Assessing Item and Person Fit
- Examining Item and Test Information Curves

Readings: Bond & Fox, Ch. 12, Wilson, Ch. 8

Assignment:

Fit your data with the appropriate item response model (Simple Rasch, Rating Scale, Partial Credit) using ConstructMap and interpret the output

WEEK 14: Validity Theory

Dec 4

- The evolution of validity since the 1950s.
- The Standards for Validity

Readings: Cronbach (1988); Angoff (1988); Messick (1995), AERA/APA/NCME (1999)

WEEK 15: Building a Validity Argument

Dec 11

- Sources of evidence for a validity argument
- Integrating evidence into a coherent argument

Readings: Toulmin (1958); Kane (2006).

WEEK 16: Group Presentations

Dec 17 (4:30-7:00)

Survey Projects

Preamble

All survey instruments contain items, and each item, or collection of items on the survey is intended to measure something. In some cases the thing to be measured is a manifest variable and a single item may suffice to measure it; for example, we can measure a person's age by asking for his or her date of birth. In other cases the thing to be measured is a latent variable and a set of items will be required to measure it; for example, we might measure a person's ability to reason quantitatively by administering the collection of math items found on the SAT. Of course, in many cases surveys are designed to measure both manifest and latent variables. In educational research, it can be difficult to conceptualize an important research question that does not necessitate the measurement of a latent variable. Latent variables come in two flavors: affective and cognitive. Affective latent variables would include a person's attitude toward some topic, their motivation, or their self-confidence. Cognitive latent variables focus on a person's knowledge, skills and understandings within or across some particular content domain. The term "construct" is often used as a synonym for a latent variable

Conditions

Your task over the 16 week period of this course will be to design, develop, pilot test, analyze and evaluate a survey instrument. Here are some conditions:

1. Your survey instrument must measure at least one cognitive latent variable. You will be attempting to conceptualize this variable in terms of a developmental learning progression. (More on this in week 2 of the course.)
2. You will need to have a MINIMUM of 30 respondents participating in the pilot test of your survey. (The more the better.)
3. My intent is for you to work on this project as part of a small group with no more than 3 students.

STAGE 1: Project Proposal (No more than 10 double-spaced pages)

Your project proposal should address the following issues:

1. Specifying the underlying use for the Survey

All surveys are developed with some use in mind, so the starting point for your project will be to specify the proposed use(s) of a survey designed to measure at least one cognitive construct. One good way to do this is to write down one or more research questions that your survey instrument could help you answer. Here are some examples:

- How well do students understand Newton's 2nd law of motion by the end of a semester-long introductory physics class?
- To what extent does a student's understanding on Newton's 2nd law of motion change over the course of a semester while enrolled in an introductory physics class?
- What is the effect of an innovative lab structure on a student's understanding on Newton's 2nd law of motion?

Sometimes survey "use" is hard to conceptualize in terms of a research question. Here is a different approach:

- The results from this survey will be used by the teacher to diagnose levels of student understanding...
- The results from this survey will be used to help the teacher make changes to the curriculum...
- Each year two scholarships are awarded to our best students. This survey will be used to identify the two best students in class.

2. Why do you need a to design a new survey instrument?

Once you have specified the use or uses for your hypothetical survey instrument, you will need to explain why existing survey instruments will not suffice. This will entail some investigation on your part—it will seldom be the case that no one has ever tried to measure the same construct. How will you be improving upon what has been done in the past? One resource that might be helpful is the *Mental Measurements Yearbook*, available at the Chinook website. Online searches using Google and ERIC can also be helpful.

3. Is there a pre-existing theory (or theories) that informs the construct you wish to measure?

In this project your construct of measurement will be some cognitive latent variable. Imagine the variable in terms of a unidimensional continuum. Some survey respondents will be high on the continuum, some will be low, some will fall in between. Has anyone come up with a theory that explains why? Does anyone have a theory that explains how people get from a low point on the continuum to a high point? If your construct is embedded in a curriculum, is there a theory that links the instructional activities of the curriculum to a student's hypothesized development in terms of the construct of interest? Do pre-existing theories contradict the idea that your construct of measurement can be conceptualized as a continuum?

4. Who are some experts that can/will help you with your survey development?

These should be people with experience and/or knowledge working in the content domain you have chosen for your construct of measurement. For example, if the domain was Chemistry, your experts should have a background in Chemistry.

5. What is your plan for obtaining your sample of pilot test respondents?

STAGE 2: Survey Development

This stage corresponds with three of the four “building blocks of measurement” described by Mark Wilson in Chapter 2-4 of his textbook *Constructing Measures*:

- Developing a Construct Map (Chapter 2)
- Designing Survey Items (Chapter 3)
- Scoring Survey Items (Chapter 4)

The culmination of this stage will be holding an item review panel with your experts, and conducting think-alouds or cognitive interviews with a subsample of your target respondents.

STAGE 3: Pilot Testing of Survey Instrument

This should take place no later than November 12th.

STAGE 3: Analysis of Survey Item Responses

This stage constitutes the fourth “building block of measurement.” In this stage you will be applying a Rasch Measurement Model to your items responses using the (free) software “ConstructMap” (formerly “GradeMap”).

STAGE 4: Evaluation of Survey Instrument

You will be expected to use the evidence gathered in stages 1-3 above to build a preliminary argument to support the validity of the survey for its proposed use. There will, of course, be many holes in the validity argument at this point. The objective is to identify these weaknesses such that revisions to the survey can be made, and a new plan for gathering additional evidence to support a stronger validity argument can be enacted.

Software for Administering and Analyzing Survey Responses

Administering Surveys

In the old days, surveys were administered in hard copy form and item responses had to be transcribed by hand. These days, there are some wonderful internet based programs that can make the act of gathering items responses much more convenient. Here are two sources that students have used in the past:

<http://www.questionpro.com/>
<http://info.zoomerang.com/?CMP=KNC-Gbd1Zoom>

Of course, you may find that an “old-school” paper and pencil survey is still the best way to go...

Analyzing Survey Responses

We will be using the software ConstructMap to analyze item responses using both classical item statistics, and output based on a calibration of responses using a Rasch Measurement Model. You can download the latest version of ConstructMap (along with a user's manual) at <http://bearcenter.berkeley.edu/GradeMap/>

We will be spending two class sessions working with ConstructMap.

Grading Policy

Assessment for EDUC 8804 will consist of

(75%) A research project, which will consist of the development and initial evaluation (pilot study) of a survey instrument. An outline of the proposed project will be developed by the student in conjunction with the instructor. The final project incorporates numerous assignments from throughout the semester:

1. Developing the theory behind your survey instrument (i.e., defining the construct of measurement in terms of a construct map).
2. Designing and critiquing items.
3. Collecting data and analyzing the instrument's measurement properties.
4. Building a preliminary validity argument.

A general outline for the final report can be found on p. 84 of the Wilson text. Components 1-3 above will be written collaboratively by each group of students working together. A separate writeup, based on component 4 above will be written by each student. Student grades on the final project will be a weighted function of the group report (50%) and individual report (50%).

(25%) Participation in classroom discussions and activities. This will include minor presentations by class members throughout the semester and an oral presentation of the research project as a culminating event at the end of the semester. It is expected that all students will complete assigned readings and exercises before each class.

Reasonable Accommodation

If you qualify for accommodations because of a disability, please submit to me a letter from Disability Services in a timely manner so that your needs may be addressed. Disability Services determines accommodations based on documented disabilities. Contact: 303-492-8671, Willard 322, and www.Colorado.EDU/disabilityservices

Disability Services' letters for students with disabilities indicate legally mandated reasonable accommodations. The syllabus statements and answers to Frequently Asked Questions can be found at www.colorado.edu/disabilityservices

Religious Observances

I will make every effort to accommodate all students who, because of religious obligations, have conflicts with scheduled exams, assignments, or other required attendance, provided advance notification of the conflict is given. Whenever possible, students should give at least two weeks advance notice to request special accommodation. For additional information on this policy, see http://www.colorado.edu/policies/fac_relig.html

Classroom Behavior

Students and faculty each have responsibility for maintaining an appropriate learning environment. Students who fail to adhere to such behavioral standards may be subject to discipline. Faculty have the professional responsibility to treat all students with understanding, dignity and respect, to guide classroom discussion and to set reasonable limits on the manner in which they and their students express opinions. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender variance, and nationalities. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. See policies at <http://www.colorado.edu/policies/classbehavior.html> and at http://www.colorado.edu/studentaffairs/judicialaffairs/code.html#student_code

The University of Colorado at Boulder policy on Discrimination and Harassment (<http://www.colorado.edu/policies/discrimination.html>), the University of Colorado policy on Sexual Harassment and the University of Colorado policy on Amorous Relationships applies to all students, staff and faculty. Any student, staff or faculty member who believes s/he has been the subject of discrimination or harassment based upon race, color, national origin, sex, age, disability, religion, sexual orientation, or veteran status should contact the Office of Discrimination and Harassment (ODH) at 303-492-2127 or the Office of Judicial Affairs at 303-492-5550. Information about the ODH and the campus resources available to assist individuals regarding discrimination or harassment can be obtained at <http://www.colorado.edu/odh>

Student Honor Code

All students of the University of Colorado at Boulder are responsible for knowing and adhering to the academic integrity policy of this institution. Violations of this policy may include: cheating, plagiarism, aid of academic dishonesty, fabrication, lying, bribery, and threatening behavior. All incidents of academic misconduct shall be reported to the Honor Code Council (honor@colorado.edu; 303-725-2273). Students who are found to be in violation of the academic integrity policy will be subject to both academic sanctions from the faculty member and non-academic sanctions (including but not limited to university probation, suspension, or expulsion). Other information on the Honor Code can be found at <http://www.colorado.edu/policies/honor.html> and at <http://www.colorado.edu/academics/honorcode/>
