#### Psychology 610: PSYCHOMETRICS

Thursdays 11:30 to 1:30 pm, 383 Straub

Instructor: Gerard Saucier, PhD Office: 425 Straub e-mail: gsaucier@uoregon.edu Phone: 346-4927 Office Hours: Mondays 1:15 to 3 pm, or by appointment Readings: See class schedule; all readings are made available (pdf) by the instructor

The purpose of this graduate course is to give students a useful introduction to basic measurement issues in assessment of individual differences. Individual differences of interest might include personality, psychopathology, interests, attitudes, beliefs, values, motives, affects, specific behaviors, and so on. Among the concepts emphasized are reliability, validity, response bias, factor structure and dimensionality of a measure, measurement invariance across populations, content comprehensiveness, cross-cultural generalizability, and key components of item response theory. Measurement issues will be explored broadly, and then in their applications to particular models and measures, including those in which the student has special experience or interest. Indeed, the course will employ measures with which the students are familiar as prime examples, and is designed not only to review the present state of knowledge and practice in psychometrics and assessment, but also to help (present or future) researchers who will be faced with the task of creating a measure of some construct, or evaluating an existing measure with respect to its measurement properties. To that end, the central, unifying theme of the course is "what makes a measure good?"

That can be turned into statistical question, in diverse ways. But it is also a conceptual question, and this is largely a course about concepts and how concepts are best represented in statistical-analysis practice, not purely about statistics. It is assumed that students are familiar with basic statistical techniques of correlation and regression. Those who doubt their facility with these techniques should consult the instructor for suggested readings.

Reading assignments are substantial [though some are optional and not required]. Much of the best work on psychometric principles is classic and highly cited work from decades ago, so not all of the readings are of recent vintage.

Each session will have some instructor presentation but also lots of discussion. We will take a very short break roughly midway through each seminar session.

#### Requirements of the course

1. *Discussion questions based on readings for the current week*. You are responsible for turning in sets of discussion questions based on the readings by two hours before the beginning of <u>five</u> different class meetings (of the nine after the first session). Late discussion questions don't confer credit. Discussion questions are **turned in via e-mail to gsaucier@uoregon.edu**. You can choose the dates of your discussion questions (although any sessions with no assigned reading would be off

limits). Discussion questions, to be worthwhile and to count, should (a) be indicative of having done the reading and (b) be instances of some degree of critical or insightful thinking. Should you ever develop a "block" about coming up with some, you might consider questions of the following form: *Why is this issue important? How are you defining* \_\_\_\_\_\_? *Aren't you assuming* \_\_\_\_\_? *Isn't it debatable whether* \_\_\_\_? *Does the evidence really support the notion that* \_\_\_\_? *Aren't you leaving out* \_\_\_\_? *Isn't there a limitation with regard to* \_\_\_\_\_ (e.g., caused by using that methodology)? What about the rival hypothesis (alternative interpretation) that \_\_\_\_\_? Responses to selected discussion questions (from the week before, or from the current week if they are submitted to the instructor well in advance of the class session) turned in will be a part of the class sessions, starting with week 2 (although sometimes this will be at the beginning of the next class session *after* the date on which they are due). Discussions in the class sessions are a very important part of this course, and student discussion questions are one of the stimulants for such discussions.

2. *Midterm exercise*. Around the middle of the term students will be given instructions for a midterm exercise involving the creation/construction of a short scale (with data that will be provided to you) and an examination of its basic psychometric properties. This exercise is designed to ensure that everyone is on the same page with respect to some basics. The exercise will be **due on February 18**, after our 'scale construction' session There may be a small/brief follow-up exercise two weeks later (due March 4).

3. *A final paper/project*. Students will be asked to identify a model or measure tapping individual differences, and discuss diverse measurement issues with respect to it. A set of generic questions (about 15) that should be addressed in the final paper will be made available by about week 6. The model or measure chosen may be one with which the student has experience, or one in which the student has a particular interest; indeed, think of this work as something that could be incorporated in a dissertation, thesis, or other eventual publication. Selected research-literature references are likely to be useful in the final paper, although none is strictly required. The final paper is **due during final-exam week**. Some kind of **advance outline or plan** for it (no longer than one page double-spaced) should be submitted by e-mail to the instructor **by February 23**.

4. *A brief presentation based on the final paper* (or at least on your early drafts of this paper) during the last two weeks of the seminar. Your presentation should be focused on questions, difficulties, puzzles, or dilemmas you are experiencing with respect to the content of your final paper (after providing just a bit of background). It is not important to create an impressive presentation performance, but rather the brief presentation is primarily an opportunity to get some feedback from the instructor and other class members on the issues involved. These presentations will be allotted a fixed time period (affected by how many course participants there are, but probably 10 minutes); discussion of a presentation may go on longer if issues of interest to many arise.

The **final grade** is based on: 30% for turning in five sets of discussion questions, 10% for the midterm exercise, 5% for a generally acceptable level of in-class contribution, 10% for the brief presentation, 5% for an advance outline of the final paper, and 40% for the final paper itself.

**Bringing in your own data**: The instructor uses real data for numerous examples in the course, and it may be particularly edifying for you to be able to see your own data, involving variables of special interest to you, applied in relation to key psychometric principles. If you have some data (even if incomplete) that you would like to see used for examples in this course, submit in excel or SPSS file format to the instructor, who guarantees it will be used only for educational purposes and only in this class.

# **Course Calendar and Readings**

(readings may be revised: readings added, reduced or changed)

#### January 8

Introduction to the seminar and to psychological measurement in general

## January 15

Validity (and its central place in the evaluation of measures); the received view, strong construct validation and the testing of rival hypotheses as to interpretation of scores **Readings for this session:** Allen & Yen (1979) pp. 95-114; Messick (1988); Cronbach (1990) chapter 5

## January 22

Validity: Alternative views and critiques, related to the survey response process, and the possible potential for better validity via indirect non-survey methods **Readings for this session:** Borsboom, Mellenbergh, & Heerden (2004); Tourangeau, Rips, & Rasinski (2000) chapter 1; Dawes (1972) chapter 9

#### January 29

*Classical test theory; measurement error; reliability assessment Readings for this session:* Allen & Yen (1979) pp. 56-65 and pp. 72-92; Schmitt (1996); Loevinger (1954) [Much more advanced optional reading posted in case you are interested: Sjitsma (2009)

## February 5

Exploratory and confirmatory factor analysis; unidimensionality assessment; criteria (to this point) for evaluating scales

**Readings for this session:** Kline (1998) ch. 3 (pp. 51-69); Slocum-Gori & Zumbo (2011); Saucier and Srivastava (2014) pp. 283-289 [More advanced optional readings posted in case you are interested: Goldberg and Velicer (2006); Zinbarg et al. (2006)]

## February 12

Perspectives on scale construction – and evaluating scales based on how constructed **Readings for this session:** DeVellis (2012) pp. 73-97; Allen & Yen (1979) pp. 118-144; Simms (2009); Tourangeau, Rips, & Rasinski (2000) pp. 23-50

## February 19

Beyond 20<sup>th</sup> century test-construction conventions: Linear composites, measurement invariance and cross-group or cross-cultural generalizability,

**Readings for this session:** DiStefano & Motl (2009); Bontempo & Hofer (); Fischer (2009) [More advanced optional reading posted in case you are interested: Cheung, Leung, & Au (2006)]

# February 26

More postconventional thinking: Equidiscrimination and item response theory (item and test information curves, differential item functioning); perhaps taxometrics (but we probably won't have time for that)

**Readings for this session:** Nunnally & Bernstein (1994) pp. 326-332; Furr & Bacharach (2014) chapter 14; Church et al. (2011) [Non-required but posted reading on taxometrics: Haslam, Holland, & Kuppers (2012)]

# March 5

Halo effects, response biases, response styles, and how to deal with them; also presentations by students

*Readings for this session:* Edwards (1953); Furr & Bacharach (2014) pp. 273-299; Paulhus (1991)

# March 12

Comparative validity studies and what they suggest about scale construction (scale length, content heterogeneity vs. homogeneity, etc.); psychometrics and graduate-admissions decisions; also presentations by students

*Readings for this session:* Thalmayer, Saucier, & Eigenhuis (2011); Credé et al. (2012); Goldberg (1977)

Final paper is due at the conclusion of the final exam date/time for this time slot (10:15 am, Mon. March 16)

Readings listed above are drawn from the following sources:

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.
- Bontempo, D. E., & Hofer, S. M. (2007). Assessing factorial invariance in cross-sectional and longitudinal studies. In A. D. Ong & M. H. M. van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 153-175). New York: Oxford University Press.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.
- Cheung, M., Leung, K., & Au, K. (2006). Evaluating multilevel models in cross-cultural research: An illustration with social axioms. *Journal of Cross-Cultural Psychology*, 37, 522-541.
- Credé, M., Harms, P., Niehorster, S., Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, *10*2, 874-888.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5<sup>th</sup> ed.). New York: Harper/Collins.
- Dawes, R. M. (1972). Fundamentals of attitude measurement. New York: Wiley.
- DeVellis, R. F. 2012). Scale development: Theory and applications (3<sup>rd</sup> ed.). Los Angeles: Sage.
- DiStefano, C. & Motl, R. W. (2009). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling*, *16*, 134-146.
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the

probability that the trait will be endorsed. *Journal of Applied Psychology*, 37, 90-93.

- Fischer, R. (2009). Where is culture in cross-cultural research? An outline of a multilevel research process for measuring culture as a shared meaning system. *International Journal of Cross-Cultural Management, 9*, 25-49.
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2<sup>nd</sup> ed.). Sage. Goldberg, L. R. (1977). Admission to the Ph.D. program in the Department of
- Psychology at the University of Oregon. American Psychologist, 32, 663-668.
- Goldberg, L. R., & Velicer, W. F. (2006). Principles of exploratory factor analysis. In S. Strack (Ed.), *Differentiating normal and abnormal personality*\_ (2<sup>nd</sup> ed.; pp. 239-247). New York: Springer.
- Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine, 42,* 903-920.
- Kline, P. (1998). *The new psychometrics: Science, psychology, and measurement.* London: Routledge.
- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin,* 51, 493-504.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (1988), *Test validity* (pp. 33-45). Hillsdale, NJ: Erlbaum.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3<sup>rd</sup> ed.). New York: McGraw-Hill.
- Paulhus, D. L. (1991). Measurement and control of response bias. In Robinson, J. P., Shaver, P. R., Wrightsman, L. S. (Eds.)\_Measures of personality and social psychological attitudes (pp. 17-59). San Diego, CA: Academic Press.
- Saucier, G. & Srivastava, S. (2014). What makes a good structural model of personality? Evaluating the Big Five and alternatives. In M. Mikulincer, P. R. Shaver, M. L. Cooper & R. J. Larsen (Eds.), Handbook of personality and social psychology. Vol. 3: Personality processes and individual differences (pp. 283-305). Washington, DC: APA Books.
- Schmitt, N. (1996). Uses and abuses of coefficient Alpha. *Psychological Assessment, 8*, 350-353.
- Simms, L. J. (2008). Classical and modern methods of psychological test construction. Social and Personality Psychology Compass, 2/1, 414-443.
- Sjitsma, K. (2009). On the use, misuse, and very limited usefulness of coefficient Alpha. *Psychometrika, 74*, 107-120.
- Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Psychological Medicine*, *4*2, 903-920.
- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief- to medium-length Big Five and Big Six questionnaires. *Psychological Assessment, 23*, 995-1009.
- Tourangeau, R. Rips, L. J. & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.
- Zinbarg, R. E., Yovel, İ., Revelle, W., & MacDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ωh. *Applied Psychological Measurement, 30,* 121-144.