Syllabus
Psychology 557: Psychometric Methods
Spring Semester, 2006
Monday, 4:30-7:10pm
Krug Hall, Room 204

Instructors:	Dr. David Dorsey	Dr. Charles (Chuck) Keil
Daytime Phone:	(703) 812-3066	(703) 269-0159
e-mail:	David.Dorsey@PDRI.com	Chuck.Keil@Monster.com

Office Hours: After class (7:15-8:15) or by appointment

Required Text: Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Stamford, CT: Thomson Learning.

Course Objective

To foster understanding of the fundamentals of developing measuring instruments, interpreting data relevant to those instruments, and knowing which instruments and indices are appropriate for which situations. Examples of course topics and their accompanying subareas follow:

- The construction of measuring instruments
  - Objectives of item analysis
  - Indices of item difficulty and item discrimination
  - The construction and use of scales, including Guttman and Likert scales
- Reliability
  - The various types of reliability (e.g., internal consistency, stability, interrater)
  - The difference between interrater reliability and interrater agreement
  - Issues in calculating the reliability of ratings
  - The relation between the choice of reliability index and the construct being measured
  - The concepts of parallel tests, domain sampling, and "true-and-error score" theory
- Validity
  - Why measure? What is the role of statistics in measurement (i.e., how is this course different from a statistics course?)
  - What are face validity, content validity, criterion-related validity, and construct validity? What is a nomological net?
  - Cross-validation
- Item response theory
  - Its relationship to classical test theory
  - The concepts of information and the conditional standard error of measurement
  - Computerized-adaptive testing
- Additional topics
  - Standard setting (i.e., determining cutoff scores)
  - Establishing test norms
  - Equating tests
  - Computerized testing

Course Requirements

- *Readings*: The course will depend greatly on class participation and discussion of the readings (textbook chapters and articles to be distributed). We would like to have a good bit of interaction during class. Questions and comments are greatly encouraged. To foster the realization of this goal, class participation will provide 20% of your final grade. Participation entails reading the assigned materials *before* class, and answering and asking questions.

- *Midquarter Exam*: A midquarter examination will be given following completion of the first half of the course. The examination will be in a short-answer and essay format and will constitute 40% of your final grade.

- *Final Exam*: The final examination will focus primarily on material presented in the second half of the course, although there will be a few questions regarding topics presented in the first half. Hence, it is "semi-comprehensive." The final will constitute 40% of your final grade and will also be short answer and essay format.

- *Grading Scale*: Grades will follow the following percentage categories:
  92-100 = A; 90-91 = A-; 88-89 = B+; 82-87 = B 80-81 = B-; <80 = C

Honor Code

As a student at George Mason University, each student in this course agrees to uphold the principles of honor set forth by the university, to defend these principles against abuse or misuse, and to abide the university regulations. This class will strictly enforce the stipulations and guidelines delineated in the George Mason University Honor Code. Students have the responsibility of engaging in honorable conduct at all times, including during study and examinations. To see the Honor Code, go to

http://mason.gmu.edu/%7Emontecin/plagiarism.htm

Students with Disabilities

If you are a student with a disability and you need academic accommodations, please see me and contact the Disability Resource Center (DRC) at 703-993-2474. All academic accommodations must be arranged through that office."

Last Add and Drop Dates

Last day to add - Feb. 7th
Last day to drop - Feb. 24th

SCHEDULE OF CLASSES AND ASSIGNMENTS

Date	Topic	Assigned Readings
1/23	Administrative matters; course overview; purpose of measurement; statistical concepts; correlation	C&A 1, 2, 3 (pp. 45-50); Binning & Barrett; Rodgers & Nicewander
1/30	Reliability: What it is; theories of reliability (Classical and Generalizability Theory)	C&A, 6, 8; Dudek; Thompson & Vacha-Haase; Fan & Chen

Date	Topic	Assigned Readings
2/6	Theories (cont.); estimation of reliability	C&A, 7; Cortina; Vacha-Haase; Schmidt & Hunter (1989)
2/13	Interrater reliability and agreement	Tinsley & Weiss; Shrout & Fleiss; McGraw & Wong; James, Demaree, & Wolf (both); Schmidt & Hunter; Kozlowski & Hattrup
2/20	Validity: What it is; theories of validity;	C&A, 10; Binning & Barrett; Murphy & Shiarella; Messick
2/27	Empirical issues in demonstrating validity (bivariate versus multiple correlation; cross-validation); validity generalization	C&A 11 (243-256); Schmidt & Hunter
3/6	**Midquarter exam**	**In class**
3/13	**Spring Break**	**NO CLASS**
3/20	Test plan; strategies for test Construction;	C&A, 4; Russell & Peterson;
3/27	The test item: Item analysis, item difficulty, item discrimination, validity, and reliability	C&A, 5, 14;
4/3	Scaling: Guttman scales, Thurstone scales, Likert scales; multidimensional scaling; Factor Analysis	C&A, 3 (pp. 50-64); Thompson & Daniel
4/10	Item Response Theory: Item characteristic curves; information; adaptive testing	C&A, 15; Drasgow (IRT Lab); Hambleton & Jones
4/17	Norming, standard setting, and equating	C&A, 18, 20
4/24	Technology and Measurement	Drasgow, Luecht, Bennett
5/1	Psychometrics and research methods trends	TBD
5/15	**FINAL EXAM**	**In Class**

Additional Readings

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 74, 478-494.

Cortina, J. M. (1993). What is coefficient alpha: An examination of theory and application. Journal of Applied Psychology, 78, 98-104.

Drasgow, F., Luecht, R., & Bennett, R. (2005). Technology and testing. In R. L. Brennan (Ed.), Educational measurement (4th Ed.). Washington, DC: American Council on Education.

Drasgow (IRT Modeling Lab): IRT Tutorial. http://io.psych.uiuc.edu/irt/.

Dudek, F.J. (1979). The continuing misinterpretation of the standard error of measurement. Psychological Bulletin, 86, 335-337.

Fan X. & Chen, M. (2000). Published Studies of Interrater Reliability Often Overestimate Reliability: Computing the Correct Coefficient. Educational and Psychological Measurement, 60, 532 - 542.

Hambleton, R. K. & Jones, R. W. (1991). Concepts, Models, and Features (pp. 7-32). In R. K. Hambleton & R. W. Jones (Eds.), Fundamentals of Item Response Theory. Newbury Park CA: Sage.

James, L.R., Demaree, R.G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. Journal of Applied Psychology, 69(1), 85-98.

James, L.R., Demaree, R.G., & Wolf, G. (1993). rwg: An assessment of within-group interrater agreement. Journal of Applied Psychology, 78(2), 306-309.

Kozlowski, S.W.J., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. Journal of Applied Psychology, 77(2), 161-167.

McGraw, K.O. & Wong, S.P. (1996). Forming inferences about some intraclass correlations. *Psychological Methods*, 1, 30-46.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50 (9), 741-749.

Murphy, K., & Shiarella, A. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. Personnel Psychology, 50, 823-854.

Rodgers, J.L., & Nicewander, W.A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59-66.

Russell, T. L., & Peterson, N. G. (1997). The test plan (pp. 115-140). In D. L. Whetzel & G. R. Wheaton (Eds.) Applied Measurement Methods in Industrial Psychology. Palo Alto, CA: Consulting Psychologists Press.

Schmidt & Hunter (1996).

Schmidt, F.L., & Hunter, J.E. (1989). Interrater reliability coefficients cannot be computed when only one stimulus is rated. Journal of Applied Psychology, 74(2), 368-370.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420-428.

Thompson, B. & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. Educational and Psychological measurement, 56, 197-208.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. Educational and Psychological Measurement, 60, 174-195.

Tinsley, H.E.A., & Weiss, D.J. (1975). Interrater reliability and agreement of subjective judgments. Journal of Counseling Psychology, 22, 358-376.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. Educational and Psychological Measurement, 58, 6-20.