Syllabus For
Measurement Theory and Methods In Behavioral Research
6J:273

Updated Summer 2006
Prof. Frank Schmidt
Room W252 Pappajohn Business Building
College of Business Administration
University of Iowa
Phone: 335-0927/0949
e-mail: frank-schmidt@uiowa.edu

“No other contribution of psychology has had the social impact equal to that created by the psychological test. No other technique and no other body of theory in psychology has been so fully rationalized from the mathematical point of view.”

“When you can measure what you are speaking about, and express it in numbers, you know something about it; when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.”
---Lord Kelvin, 1891, pp. 80 – 81.

“If something exists, it exists in some amount. If it exists in some amount, then it is capable of being measured.”
---Descartes, 1644.

“We must measure what is measurable and make measurable what cannot be measured.”
---Galileo

This graduate level course covers measurement and statistical methods needed for the conduct of methodologically sound, publishable research. Topics include: kinds and levels of measurement; role of measurement in theory development and cumulative knowledge; the theory of measurement error; types of reliability and their estimation; reliability of difference scores and composite scores; corrections for bias due to measurement error; basic scaling methods; criterion-related, content, and construct validity; cross-validation and shrinkage formulas; the role of base rates in validity; test validity and minority groups; practical utility of tests; statistical power in validity studies; introduction of meta-analysis; item analysis and scale construction; and other topics (e.g., suppressor variables).

Course includes seven (7) exercises based on class presentations. Additional exercises are added from time to time. There is no paper or term project. There is a final examination.

Texts:

Other Readings:

Other readings are assigned that are on Reserve in the Business Library. Reserve materials are indicated at the end of this syllabus.

<table>
<thead>
<tr>
<th>Week (approximate)</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &amp; 2</td>
<td>1. Principles of Psychological Measurement</td>
</tr>
<tr>
<td>3 &amp; 4</td>
<td>2. Estimation of Reliability</td>
</tr>
<tr>
<td>5 &amp; 6</td>
<td>3. Criterion Construction and Scaling Methods</td>
</tr>
<tr>
<td>7 &amp; 8</td>
<td>4. Validity</td>
</tr>
<tr>
<td>9</td>
<td>Validity (Cont.)</td>
</tr>
<tr>
<td>10 &amp; 11</td>
<td>5. Combining Tests in Batteries; Cross Validation</td>
</tr>
<tr>
<td>12 &amp; 13</td>
<td>6. Scale Construction and Item Analysis</td>
</tr>
</tbody>
</table>

I would like to hear from anyone who has a disability that may require some modification of seating, testing, or other class requirements so that appropriate arrangements can be made. Please see me after class or during my office hours.

In connection with the exercises that are part of this course, I expect that the answers turned in by each student will reflect only that student’s work. It is permissible for students to discuss general aspects of the exercises with other students prior to working the exercises, but the actual calculations and other work should be done by the student alone. I will contact the student if I find evidence that this is not the case and that the Tippie College honor code has been violated.
**Topic 1**

First and Second Weeks

Nature and Role of Psychological Measurement: Basic Principles

A. Kinds and Levels of Measurement

Lemke and Wiersma, 2 and 3
Helmstadter, 20-24 and 40-58
Guilford, 1
Lindquist, 14
Nunnally and Bernstein, 1, 2, 4, 5
Lord and Novick, 1
(Associate in Mehrens and Ebel, No. 1.)
Anastasi, 2
Eysenck, 11-17 (Handout)

B. Introduction to Basic Test Theory

Magnusson, 1 (on reserve)
Lord and Novick, 3
Guilford, 349-354; 358-361
Ghiselli, 3

C. Relevant Background Reading

Cronbach, L.J. (1957). The two disciplines of scientific psychology. *Amer. Psychologist*. (Also in Jackson and Messick, No. 2.) (on reserve)
Guion, 1
Horst, 1
Chase and Ludlow, 1
D. Other Readings


E. The Attenuation Paradox


Topic 2

Third and Fourth Weeks

The Estimation of Reliability

A. General

Lemke & Wiersma, 4, 5
Helmstader, 58-68 and 74-86
Anastasi, 5
Magnusson, 5, 6, 8, 9
Guilford, pp. 373-398
Lord & Novick, 5, 6
Thorndike (1949), 4  
Nunnally & Bernstein, 6, 7; also pp. 338 – 347 (effects of guessing)
Gulliksen, 8, 10, 15, 16  
Lindquist, 15
APA Standards, 48-55  
Guion, 2
Ghiselli, 8, 9
Cronbach, L.J. (1951). Coefficient Alpha and the internal structure of tests. Psychometrika, 16, 297-334. (Also in Mehrens and Ebel, No. 18.)

B. Reliability of Ratings of Job Performance


C. Reliability of Difference Scores

Traub (1994), 127-138
Magnusson, 7

D. Speed and Power Tests and Related Reliability Problems

Gulliksen, 17
Guilford, 365-370
Nunnally, 628-641

E. Special Topics in Reliability (Selected References)

Hoyt, C.J. (1951). Test reliability estimated by analysis of variance. Psychometrika, 6, 153-160. (Also in Mehrens and Ebel, No. 16.)
Lord, F.M. (1959). Tests of the same length do have the same standard error of
10, 43-56.
Rajaratnam, N., Cronbach, L.J., & Gleser, G.C. (1964). Generalizability of stratified-
paralleled tests. Psychometrika, 29, 39-56.
Psychometrika, 33, 1-18.
Brogden, H.E. (1946). The effect of bias due to difficulty factors in product-moment item
6, 517-520.
Rulon, P.J. (1939). A simplified procedure for determining the reliability of a test by split
halves. Harvard Educ. Review, 9, 99-103. (Also in Mehrens and Ebel, No. 15.)
computed when only one stimulus is rated. Journal of Applied Psychology, 74,
368-370. (In Topic 2 readings packet)
asymptote level with increasing opportunity to observe. Journal of Applied
Psychology, 75, 322-327.
Psychologist, 44, 922-932. (In readings packet)
Mangione, T.W., & Quinn, R.P. (1975). Job satisfaction, counterproductive behavior, and
drug use at work. Journal of Applied Psychology, 60, 114-116. (Example of study
in which all relations among variables are greatly attenuated due to low reliability
--and authors do not know this.)
Lessons from 26 research scenarios. Psychological Methods, 1, 199 – 223. (In
readings packet)

F. Some Applications

Msmt., 26, 13-14.

Topic 3

Fifth and Sixth Weeks

Criterion Construction and Scaling Methods
A. Determination of Incumbent Trait Requirements (Personnel Specifications)

Thorndike (1949), 2
Tiffin & McCormick, 3
Dunnette, 4, 5
Ghiselli & Brown, 3
Shantle, 6, 11
McCormick, E.J. Job analysis: Methods and Applications. New York, AMACOM.

B. Criterion Construction (General)

Thorndike (1949), 5
Tiffin & McCormick, 227-258
Guion, 4
Lindquist, 626-640
Lawsh & Balma, 3
(Students buy this in copy center.)

C. Criterion Construction: Scaling Methods

1. Pair Comparisons

Review Ch. 2 of Nunnally & Bernstein (assigned earlier in Topic 1)
Edwards, 19-52
Guilford, 154-177

2. Forced Choice Rating

Guilford, 274-278
Ghiselli & Brown, 114-121

Applications


3. **Rating Scales**

Guilford, 11
Ghiselli & Brown, 103-110
Tiffin & McCormick, 227-232
Guion, 97-103
Rothstein (1990) (See Topic 2, Section D.)

4. **Ranking**

Guilford, 8
Ghiselli & Brown, 96-103
Guion, 100-101

5. **Additional References on Scaling**

Edwards, 172-199
Guilford, 456-462

D. Multiple vs. Composite Criteria


**Topic 4**

Seventh, Eighth and Ninth Weeks

Validity and Utility

A. General

Nunnally & Bernstein, 3
Anastasi, 6
Magnusson, 11, 12, 13
Thorndike, (1949), 6
Guilford, 398-409; 356-357
APA Standards, 9-18
Guion, 6
Gulliksen, 9
Ghiselli, 11
Lord & Novick, 12 (parts)
Lindquist, 674-693; 640-675
B. Statistical Power


C. Range Restriction


D. Interpretations of Validity Coefficients

Anastasi, 157-173
Cronbach & Gleser, 4
Guion, 150-158
Tiffin & McCormick, 127-145


E. Validity Generalization and Related Topics


**F. Role of the Base Rate in Validity**


Rorer, et al. (See two articles under "Interpretation of Validity Coefficients.")


**G. Criterion-Related Validity: General**


**H. Test Validity and Minority Groups**


I. Suppressor Variables


J. Special Topics in Noncriterion Validity (Content and Construct Validity)


Campbell, D.T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. (In Jackson & Messick.)


**Topic 5**

**Tenth and Eleventh Weeks**

**Combining Tests into Batteries; Cross-Validation**

**A. Selecting and Weighting Tests**

Blum & Naylor, 3
Thorndike (1949), 185-204
Lord & Novick, 284-288
Lindquist, 778-794
Ghiselli, 10
Gulliksen, 20
Guilford, 403-406


**B. Clinical vs. Actuarial Combining of Test Scores**


Meehl, P.E. What can the clinician do well? In Jackson and Messick, No. 48.


Cronbach, P. 441 ff.


**C. Cross-Validation and Double Cross-Validation**

Guilford, 405-406; 440-441.

D. Use of Shrinkage Formulae in Lieu of Cross-Validation

Wherry, R.J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. The annals of mathematical statistics, 2, 440-457. (This widely but incorrectly used formula estimates \( \rho^2(\mathbf{B}) \) instead of \( \rho^2(\hat{\mathbf{B}}) \). Also, is a biased estimate.)


E. Some Applications


**Topic 6**
Twelfth and Thirteenth Weeks

Scale Construction and Item Analysis

A. Scale Construction and Invention

Thorndike (1949), 3
Adkins, 4, 5, 6, 7, 10
Furst, 7, 8, 9, 10, 11, 12, 13
Guilford, 414-417
Guion, 187-198; 205-209
Lord & Novick, 284-293
Nunnally & Bernstein, 8, 9
Lindquist, 5, 6, 7, 8
Traub, Ch. 7

B. Item Analysis: General

Adkins-Wood, 9
Anastasi, 8
Magnusson, 2, 4, 14
Nunnally, 8
Thorndike (1949), 8
Thorndike (in Jackson & Mossick)
Guilford, 417-443
Lord & Novick, 15
Lindquist, 9
Gulliksen, 21
Guion, 198-205
C. Cross Validation of Item Analysis


D. Different Approaches to Item Analysis

Guilford, p. 442. (Negative item analysis.)

E. Comparisons of Different Item Selection Techniques


F. Other Contributions of Interest


Additional Topics

A. Effects of Guessing


Magnusson, 15

Nunnally, 641-655
B. Scoring Problems

Thorndike, 7, 9
Lindquist, 17
Gulliksen, 18
Lord and Novick, 14

C. Administration of a Testing Program

Thorndike, 9, 10, 11
Lindquist, 10
Michigan State University Guidance Department. Designing and Implementing a Testing Program. (In Payne and McMorris, No. 47.)
Reference Texts


Eysenck, H.J. The structure and measurement of intelligence. New York: Springer-Verlag, 1979 (pp. 11-17). (Class handout)


Lemke, E. & Wiersma, W. Principles of psychological measurement. (On Reserve)


Magnusson, D. Test theory. Addison-Wesley, 1966. (On Reserve)


