

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.”

---J.P. Guilford, 1954, p. 341.

“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:

Nunnally, J.C., & Bernstein, I.H. (1994). Psychometric Theory (3rd Ed.). New York: McGraw-Hill.

Other Readings:

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

Brief Outline of Topics

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

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

Stevens, S.S. (1946). On the theory of scales of measurement. Science, 103, 677-680.

(Also in Mehrens and Ebel, No. 1.)

Aftanas, M.S. (1988). Theories, models, and standard systems of measurement. Applied Psychological Measurement, 12, 325-338.

Stine, W.W. (1989). Meaningful information: The role of measurement in statistics. Psychological Bulletin, 105, 147-155.

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. Psychological Bulletin, 100, 398-407.

Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. Psychological Bulletin, 74, 167-184.

Johnson, C.E., Wood, R., & Blinkhorn, S.F. (1988). Spuriousness and Spuriousness: The use of ipsative personality tests. Journal of Occupational Psychology, 61, 153-162.

Anastasi, 2

Eysenck, 11-17 (Handout)

Schwager, K.W. (1991), The representational theory of measurement: An assessment. Psychological Bulletin, 110, 618-626.

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

Ghiselli, 1, 2

Humphreys, L.G., & Fleishman, A. (1974). Pseudo-orthogonal and other ANOV designs involving individual differences variables. J. Educ. Psychol., 66, 464-472.

Gardner, P.L. (1975). Scales and statistics. Review of Educational Research, 43-57.

D. Other Readings

Lord, F.M. (1962). Estimating norms by item sampling. Educ. Psych. Msmt., 22, 259-267. (Also in Mehrens and Ebel, No. 10.)

Lord, F.M. (1959). Test norms and sampling theory. J. Exper. Educ., 27, 247-263.

Schmidt, F.L. (1973). Implications of a measurement problem for expectancy theory research. Organizational Behavior and Human Performance, 10, 243-251. (in Topic 1 Packet)

E. The Attenuation Paradox

Brogden, H.E. (1946). Variations in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. Psychometrika, 11, 197-214.

Humphreys, L.G. (1956). The normal curve and the attenuation paradox in test theory. Psych. Bull., 53, 472-476.

Loevinger, Jane (1954). The attenuation paradox in test theory. Psych. Bull., 51, 493-504.

Tucker, L.R. (1946). Maximum validity of a test with equivalent items. Psychometrika, 11, 1-13.

Cronbach, L.J., & Warrington, W.G. (1952). Efficiency of multiple-choice tests as a function of item difficulties. Psychometrika, 17, 127-147.

Richardson, M.W. (1936). The relation of difficulty to the differential validity of a test. Psychometrika, 1, 33-49.

Engel, J. (1977). The attenuation paradox and latent trait theory. U.S. Civil Service Commission: Personnel Research and Development Center.

Topic 2

Third and Fourth Weeks

The Estimation of Reliability

A. General

Lemke & Wiersma, 4, 5

Helmstadter, 58-68 and 74-86

Anastasi, 5

Guilford, pp. 373-398

Magnusson, 5, 6, 8, 9

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
- Feldt, L.S., & Brennan, R.L. (1989). Reliability. In R.L. Linn (Ed.), Educational Measurement (3rd Ed., 105-146). NY: Macmillan.
- Kuder, G.F., & Richardson, M.W. (1937). The theory of the estimation of test reliability. Psychometrika, 2, 151-160. (Also in Mehrens and Ebel, No. 14.)
- Cronbach, L.J. (1951). Coefficient Alpha and the internal structure of tests. Psychometrika, 16, 297-334. (Also in Mehrens and Ebel, No. 18.)
- Cureton, E.E. (1958). The definition and estimation of test reliability. Educ. Psych. Msmt., 18, 715-738. (Also in Mehrens and Ebel, No. 19.)
- Cronbach, L.J. (1947). Test "reliability": Its meaning and determination. Psychometrika, 12, 1-16.
- Tyron, R.C. (1957). Reliability and behavior domain validity: Reformulation and historical critique. Psych. Bull., 229-249.
- Le, H., Schmidt, F.L., & Lauver, K. How reliable are measures of job satisfaction? New answers from Generalizability Theory. Unpublished paper. (In Topic 2 readings packet)
- Schmidt, F.L., & Hunter, J.E. (1999). Theory testing and measurement error. Intelligence, 27, 183 – 198. (In Topic 2 readings packet)

B. Reliability of Ratings of Job Performance

- King, L.M., Hunter, J.E., & Schmidt, F.L. (1980). Halo in a multidimensional forced choice performance evaluation scale. Journal of Applied Psychology, 65, 507–516.
- Rothstein, H.R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. Journal of Applied Psychology, 75, 322–327.
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlation: Uses in assessing rater reliability. Psych. Bull., 86, 420 – 428.
- Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (1996). Comparative analysis of the reliability of job performance ratings. Journal of Applied Psychology, 81, 557–574.

C. Reliability of Difference Scores

- Traub (1994), 127-138
- Magnusson, 7
- Cronbach, L.J., & Furby, L. (1970). How should we measure change--or should we? Psych. Bull., 74, 47-67.
- Lord, F.M. (1956). The measurement of growth. Educ. Psych. Msmt., 16, 421-437.

- Lord, F.M. (1958). Further problems in the measurement of growth. Educ. Psych. Msmt., 18, 437-454. (Also in Jackson and Messick, No. 17.)
- McNemar, Q. (1958). On growth measurement. Educ. Psych. Msmt., 18, 47-55.
- Stanley, J.C. (1967). General and special formulas for reliability of differences. J. Educ. Msmt., 4, 249-252.
- Traub, R.E. (1967). A note on the reliability of residual change scores. J. Educ. Msmt., 4, 253-256.
- Trimble, H.C., & Cronbach, L.J. (1943). A practical procedure for the rigorous interpretation of test-retest scores in terms of pupil growth. J. Educ. Msmt., 35, 481-488.
- Lord, F.M. (1958). The utilization of unreliable difference scores. J. Educ. Psych., 49, 150-152.

D. Speed and Power Tests and Related Reliability Problems

- Gulliksen, 17
- Guilford, 365-370
- Nunnally, 628-641
- Cronbach, L.J., & Warrington, W.G. (1951). Time limit tests: Estimating their reliability and degrees of spending. Psychometrika, 16, 157-168.

E. Special Topics in Reliability (Selected References)

- Cronbach, L.J., Rajaratnam, N., & Gleser, Goldine C. (1963). Theory of generalizability: A liberalization of reliability theory. British Journal of Statistical Psychology, 16, 137-163.
- Cronbach, L.J., Rajaratnam, N., & Gleser, Goldine C. (1959). Interpretation of reliability and validity coefficients: Remarks on a paper by Lord. J. Educ. Psych., 50, 230-237.
- Cronbach, L.J., & Hartman, W. (1954). A note on negative reliabilities. Educ. Psych. Msmt., 14, 342-346.
- Cureton, E.F., et al. (1973). Length of test and standard error of measurement. Educ. Psych. Msmt., 33, 63-68.
- Horst, P. (1954). The estimation of immediate retest reliability. Educ. Psych. Msmt., 14, 705-708.
- Horst, P. (1953). Correcting the K-R reliability coefficient for dispersion of item difficulties. Psych. Bull., 50, 371-374.
- 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. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. Psychometrika, 17, 181-194.
- Lord, F.M. (1956). Sampling error due to choice of split in split-half reliability coefficients. J. Experimental Educ., 24, 245-249.

- Lord, F.M. (1959). Tests of the same length do have the same standard error of measurement. Educ. Psych. Msmt., 233-239.
- Coombs, C.H. (1950). The concepts of reliability and homogeneity. Educ. Psych. Msmt., 10, 43-56.
- Rajaratnam, N., Cronbach, L.J., & Gleser, G.C. (1964). Generalizability of stratified-paralleled tests. Psychometrika, 29, 39-56.
- Hunter, J.E. (1968). Probabilistic foundations for coefficients of generalizability. Psychometrika, 33, 1-18.
- Brogden, H.E. (1946). The effect of bias due to difficulty factors in product-moment item intercorrelations of the accuracy of estimation of reliability. Educ. Psych. Msmt., 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.)
- Schmidt, F.L., & Hunter, J.E. (1989). Interrater reliability coefficients cannot be computed when only one stimulus is rated. Journal of Applied Psychology, 74, 368-370. (In Topic 2 readings packet)
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.
- Rothstein, H.R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. Journal of Applied Psychology, 75, 322-327.
- Shavelson, R.J., Webb, N.M., & Rowley, G.L. (1989). Generalizability theory. American 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.)
- Schmidt, F.L., & Hunter, J.E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. Psychological Methods, 1, 199 – 223. (In readings packet)

F. Some Applications

- Anastasi, A., & Drake, J. (1954). An empirical comparison of certain techniques for estimating the reliability of speeded tests. Educ. Psych. Msmt., 15, 529-540.
- Cureton, E.E. (1966). Kuder-Richardson reliability of classroom tests. Educ. Psych. Msmt., 26, 13-14.
- Strong, E. K. (1954). Validity vs. reliability. J. Applied Psych., 38, 103-104.

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
 Otis, J.L. (1952). Job Analysis. Pers. Psych., 25-29.
 Hill, J.M. (1956). The time span of discretion in job analysis. Human Relations, 9, 295-324.
 Pearlman, K. (1980). Job families: A review and discussion of their implications for personnel selection. Psych. Bull., 87, 1-28.
 McCormick, E.J. Job analysis: Methods and Applications. New York, AMACOM.
 Prien, E.P., & Ronan, W.W. (1971). Job Analysis: A review of research findings. Pers. Psych., 24, 371-396.
 Bemis, S., Schmidt, F.L., & Caplan, J.R. Manual for the behavioral consistency examination procedure. (BRE Exam Preparation Manual) USCSC, June 1977.

B. Criterion Construction (General)

- Thorndike (1949), 5
 Tiffin & McCormick, 227-258
 Guion, 4
 Astin, A.W. (1964). Criterion centered research. Educ. Psych. Msmt., 24, 807-821.
 Brogden, H.E., & Taylor, E.K. (1950). A theory and classification of criterion bias. Educ. Psych. Msmt., 10, 159-186.
 Lindquist, 626-640
 Lawshe & Balma, 3
 Schmidt, F.L. (1979). The measurement of job performance. Unpublished paper. (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
 Guilford, J.P. (1928). The method of paired comparisons as a psychometrics technique. Psych. Review, 35, 494-506.
 Bartlett, C.J., Heerman, E., & Retting, S. (1960). A comparison of six different scaling techniques. J. Soc. Psych., 51, 343-348.
 Kephart, N.C., & Oliber, J. (1952). A punched card procedure for use with the method of paired comparisons. J. Applied Psych., 36, 47-48.

- Lawshe, C.H., & Kephard, N.C. (1950). Manual for use with the Personnel Comparison System, Lafayette, Ind., Southworth Book Store.
- Lawshe, C.H., Kephard, N.C., & McCormick, E.J. (1949). An investigation of the method of paired comparison technique for rating performance of industrial employees. Journal of Applied Psychology, *33*, 69-77.
- McCormick, E.J., & Bachus, J.A. (1952). Paired comparisons. I. The effect on ratings of reductions in the number of pairs. Journal of Applied Psychology, *36*, 123-127.
- McCormick, E.J., & Robers, W.K. (1952). Paired comparison ratings. 2. The reliability of ratings based on partial pairings. Journal of Applied Psychology, *36*, 188-192.
- Oliver, J.E. (1953). A punched card procedure for use with partial pairings. Journal of Applied Psychology, *37*, 129-130.
- Rambo, W.W. (1959). The effects of partial pairings on scale values derived from the method of paired comparisons. Journal of Applied Psychology, *43*, 379-381.
- Rambo, W.W. (1959). Paired comparison scale value variability as a function of partial pairings. Psych. Reports, *5*, 341-344.
- Schucker, R.E. (1959). A note on the use of triads for paired comparison. Psychometrika, *24*, 273-276.

2. Forced Choice Rating

- Guilford, 274-278
- Ghiselli & Brown, 114-121
- Zavala, A. (1965). Development of the forced choice rating scale technique. Psych. Bull., *63*, 117-124.
- Highland, R.W., & Berkshire, J.R. (1951). A methodological study of forced-choice performance rating. Res. Bull., 51-9. San Antonio, TX: Human Resources Research Center. Also Educ. Psych. Msmt., 1957, 1958.
- Berkshire, J.R. (1958). Comparison of five forced-choice indices. Educ. Psych. Msmt., *18*, 553-561.
- Harris, F.J., Howell, M.A., & Newman, S.H. (1956). Forced-choice tetrads-effect of scoring procedures and key length on validity and reliability. Educ. Psych. Msmt., *16*, 454-464.
- Waters, L.K., & Wherry, R.J. (1962). The effect of intent to bias on forced-choice indices. Pers. Psych., *15*, 207-214.
- Travers, R.M.W. (1951). A critical review of the validity and rationale of the forced-choice technique. Psych. Bull., *48*, 62-70.
- Hicks, L.E. (1970). Some properties of ipsitive, normative and forced-choice measures. Psych. Bull., *74*, 167-184.
- King, L., Hunter, J.E., & Schmidt, F.L. (1980). Halo in a multi-dimensional forced choice rating scale. Journal of Applied Psychology, *65*, 507-516.

Applications

- Norman, W.T. (1963). Personality measurement, faking and detection: An assessment method for use in personnel selection. Journal of Applied Psychology, 47, 225-236.
- Schwartz, S.L., & Gekoski, N. (1960). The Supervisory Inventory: A forced-choice measure of human relations attitude and technique. Journal of Applied Psychology, 44, 233-236.
- Maher, H. (1959). Follow-up on the validity of a forced-choice study activity questionnaire in another setting. Journal of Applied Psychology, 43, 293-295.

3. Rating Scales

- Guilford, 11
 Ghiselli & Brown, 103-110
 Tiffin & McCormick, 227-232
 Guion, 97-103
 King, L.M., Hunter, J.E. & Schmidt, F.L. (1989). (See Section 2, above.)
 Rothstein (1990) (See Topic 2, Section D.)

4. Ranking

- Guilford, 8
 Ghiselli & Brown, 96-103
 Guion, 100-101
 Bartlett, C., Heermann, E., & Rettig, S. (1960). A comparison of six different scaling techniques. Journal of Social Psychology, 51, 343-348. (Shows ranking has higher reliability than ratings; nearly as high as pair comparisons.)

5. Additional References on Scaling

- Edwards, 172-199
 Guilford, 456-462
 Lickert, R. (1932). A technique for the measurement of attitudes. Arch. Psych., No. 140, 55.
 Barclay, J.E., & Weaver, M.B. (1962). Comparative reliabilities and ease of construction of Thurstone and Lickert attitude scales. J. Soc. Psych., 58, 109-120.
 Bartlett, C.J., Quay, L.C., & Wrightsmon, L.W., Jr. (1960). A comparison of two methods of attitude measurement: Lickert-type and forced-choice. Educ. Psych. Msmt., 20, 699-704.

Edwards, A.L., & Kirkpatrick, F.P. (1948). A technique for the construction of attitude scales. Journal of Applied Psychology, 32, 374-384.

D. Multiple vs. Composite Criteria

Schmidt, F.L., & Kaplan, L.B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. Pers. Psych., 24, 419-434.

Dunnette, M.D. (1963). A note on the criterion. Journal of Applied Psychology, 47, 251-254.

Dunnette, M.D. (1963). A modified model for test validation and selection research. Journal Applied Psychology, 47, 317-323.

Nagle, B.F. (1953). Criterion development. Pers. Psych., 6, 271-289.

Wallace, S.R. (1965). Criteria for what? Amer. Psych., 20, 411-417.

Gaylord, R.H., & Brogden, H.E. (1964). Optimal weighting of unreliable criterion elements. Educ. Psych. Msmt., 24, 529-533.

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

Schmidt, F.L., & Hunter, J.E. (1980). The future of criterion-related validity. Pers. Psych., 33, 41-60. (In readings packet)

Schmidt, F.L., & Hunter, J.E. (1981). Employment testing: Old theories and new research findings. Amer. Psych., 36, 1128-1137. (Special Issue on Testing) (Also in Readings in professional personnel assessment. Washington, D.C.: The International Personnel Management Association, in press; in Rynes & Milkovich (Eds.), Readings in industrial relations, in press; and in C.E. Schneider, R.W. Beatty, & G.M. McEvoy (Eds.), Personnel/human resource management today, 2nd Ed. Addison-Wesley Co., in press). (In readings packet)

B. Statistical Power

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). Hillsdale, N.J.: Erlbaum.
- Raju, N.S., Edwards, J.E., & LoVerde, M.A. (1985). Corrected formulas for computing sample sizes under indirect range restriction. Journal of Applied Psychology, *70*, 565-566.
- Schmidt, F.L., Hunter, J.E., & Urry, V.W. (1976). Statistical power in criterion-related validation studies. Journal of Applied Psychology, *61*, 473-485.(In readings packet)
- Sackett, P.R., & Wade, B.E. (1983). On the feasibility of criterion related validity: The effects of range restriction assumptions on needed sample size. Journal of Applied Psychology, *68*, 374-381.
- Sedlmeier, P., & Gigerenger, G. (1989). Do studies of statistical power have an effect on the power of studies? Psychological Bulletin, *105*, 309-316.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. American Psychologist, *47*, 1173-1181.

C. Range Restriction

- Hoffman, C.C. (1995). Applying range restriction corrections using public norms: Three case studies. Personnel Psychology, *48*, 913-924.
- Held, J.D., Foley, P.P. (1994). Explanations for accuracy of the general multivariate formulas for correcting for range restriction. Applied Psychological Measurement, *18*, 335 – 367.
- Sackett, P.R., & Ostgaard, D.J. (1994). Job-specific applicant pool and national norms for cognitive ability tests: Implications for range restriction corrections in validation research. Journal of Applied Psychology, *79*, 680-684.
- Ree, M.J., Carretta, T.R., Earles, J.A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. Journal of Applied Psychology, *79*, 298-301.
- Linn, R.L. (1968). Range restriction problems in the use of self-selected groups for test validation. Psychological Bulletin, *69*, 69-73.
- Linn, R. L., Harnish, D.L., & Dunbar, S.B. (1981). Correction for range restriction: An empirical investigation of conditions resulting in conservative correction. Journal of Applied Psychology, *66*, 655-663.
- Linn, R.L. (1983). The Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. Journal of Educational Measurement, *20*, 1 – 15.
- Sackett, P.R., & Yang, H. (2000). Correction for range restriction: An expanded typology. J. Applied Psychology, *85*, 112 – 118.
- Hunter, J.E., Schmidt, F.L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. Journal of Applied Psychology, *91*, 594 – 612.

Hunter, J.E., & Schmidt. (2004). Methods of meta-analysis (2nd Edition). Sage. [Chapter 5: Discussion of indirect range restriction.]

D. Interpretations of Validity Coefficients

Anastasi, 157-173

Cronbach & Gleser, 4

Guion, 150-158

Tiffin & McCormick, 127-145

Boldt, R.F. (1978). Robustness of range restriction in court. Paper presented at 1978 APA Convention, Toronto, Canada, August 28-Sept. 1.

Richardson, M.W. (1944). The interpretation of a test validity coefficient in terms of increased efficiency of a selected group of personnel. Psychometrika, 9, 245-248.

Rorer, L.G., Hoffman, P.J., LaForce, G.E., & Hsieh, K.C. (1966). Optimum cutting scores to discriminate groups of unequal size and variance. Journal of Applied Psychology, 50, 153-164.

Linn, R.L. (1985). The Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. J. Educ. Msmt.

Rorer, L.G., Hoffman, P.J., & Hsieh, K.E. (1966). Utilities as base-rate multipliers in the determination of optimum cutting scores for the discrimination of groups of unequal size and variance. Journal of Applied Psychology, 50, 364-368.

Curtis, E.W. and Alf, E.F. (1969). Validity, predictive efficiency, and practical significance of selection tests. Journal of Applied Psychology, 53, 327-337.

Brogden, H.E. (1949). A new coefficient: Application to biserial correlation and to estimation of selective efficiency. Psychometrika, 14, 169-182.

Jarrett, R.F. (1948). Percent increase in output of selected personnel as an index of test efficiency. Journal of Applied Psychology, 32, 135-145.

Curtis, E.W. (1966). The application of decision theory and scaling methods to selection test validation. Dissert. Abstracts, 26, 4794.

Brewer, J.K., & Hills, J.R. (1969). Univariate selection: The effects of size of correlation, degree of skew, and degree of range restriction. Psychometrika, 34, 347-361.

Brogden, H.E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. J. Educ. Psych., 37, 65-76.

Wickert, F.R. Some implications of decision theory for occupational selection. In Payne and McMorris, No. 40.

Brogden, H.E. (1949). When testing pays off. Pers. Psych., 2, 171-184.

Schmidt, F.L., Hunter, J.E., & Urray, (1976). (See under "Statistical Power"; discussion of range restriction.)

Schmidt, F.L., & Hoffman, B. (1973). An empirical comparison of three methods of assessing the utility of a selection device. Journal of Industrial and Organizational Psychology, 1, 14-23. (Also in W.C. Hamner, & F.L. Schmidt (Eds.), Contemporary problems in personnel, St. Clair Press, 1974.)

Schmidt, F.L., & Hunter, J.E. (1979). Poor selection procedures lower productivity. Civil Service Journal, 19, 9.

- Hunter, J.E., & Schmidt, F.L. (1982). Fitting people to jobs: Implications of personnel selection for national productivity. In E.A. Fleishman, & M.D. Dunnette (Eds.) Human performance and productivity. Volume 1: Human capability assessment. Hillsdale, NJ: Earlbaum, 233-284.
- Schmidt, F.L., Hunter, J.E., McKenzie, R.C., & Muldrow, T.W. (1979). The impact of a valid selection procedure on work-force productivity. Journal of Applied Psychology, 64, 609-626. (In readings packet)
- Schmidt, F.L., & Hunter, J.E. (1982). The money test. Across the Board, The Conference Board Magazine, 19, 7, 35-38. (Also in L.N. Jewell (Ed.), Industrial organizational psychology for the eighties, West Publishing Co., in press.)
- Schmidt, F.L., & Hunter, J.E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. Journal of Applied Psychology, 68, 407-415.
- Schmidt, F.L., Hunter, J.E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on workforce productivity. Pers. Psych., 35, 333-347. (Also in R.S. Schuler, & S.A. Youngblood (Eds.), Personnel and human resource management. NY: West Publishing Co., 1984.)
- Schmidt, F.L., Mack, M.J., & Hunter, J.E. (1984). Selection utility in the occupation of U.S. Park Ranger for three modes of test use. Journal of Applied Psychology, 69, 490-497.
- Schmidt, F.L., Hunter, J.E., Outerbridge, A.M., & Trattner, M.H. (1986). The economic impact of job selection methods on the size, productivity, and payroll costs of the Federal work-force: An empirical demonstration. Personnel Psychology, 39, 1-29. (In readings packet)
- Hunter, J.E., Schmidt, F.L., & Judiesch, M.K. (1990). Individual differences in output variability as a function of job complexity. Journal of Applied Psychology, 75, 28-42.
- Hunter, J.E., Schmidt, F.L., & Coggin, T.D. (1988). Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing the utility of personnel programs. Journal of Applied Psychology, 73, 522-528.

E. Validity Generalization and Related Topics

- Lawshe, C.H., & Sternberg, M.D. (1955). Studies in synthetic validity. I. An exploratory investigation of clerical jobs. Pers. Psych., 8, 291-301.
- Dawson, R.I. (1952). A new approach to test validation for clerical jobs. Doctoral dissertation, Purdue Univ.
- Balma, M.J. (1959). The concept of synthetic validity. Pers. Psych., 12, 395-396.
- McCormick, E.J. (1959). The development of a procedure for indirect or synthetic validity, 3. Application of job analysis to indirect validity. Pers. Psych., 12, 402-413.
- Guion, R.N. (1965). Synthetic validity in a small company: A demonstration. Pers. Psych., 18, 49-63.

- Coward, W.M. & Sackett, P.R. (1990). Linearity of ability-performance relationships: A re-confirmation. Journal of Applied Psychology, 75, 297-300. (In readings packet)
- Hunter, J.E. & Schmidt, F.L. (1994). The estimation of sampling error variance in meta-analysis of correlations: The homogenous case. Journal of Applied Psychology, 79, 171-177.
- Primoff, E.S. (1957). The J-coefficient approach to jobs and tests. Pers. Psych., 20, 3, 34-40.
- Primoff, E.S. (1959). Empirical validations of the J-coefficient. Pers. Psych., 413-418.
- Schmidt, F.L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. J. Applied Psych., 62, 529-540.
- Schmidt, F.L., Hunter, J.E., Pearlman, K., & Shane, G.S. (1979). Further tests of the Schmidt-Hunter Bayesian Validity Generalization Model. Pers. Psych., 32, 257-281.
- Ones, D.S., Viswesvaran, C., & Schmidt, F.L. (1993). Meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. Journal of Applied Psychology Monograph, 78, 679-703.
- Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalization results in tests used to predict job proficiency and training criteria in clerical occupations. Journal of Applied Psychology, 65, 373-407.
- Rothstein, H.R., Schmidt, F.L. et al. (1990). Biographical data in employment selection: Can validities be made generalizable? Journal of Applied Psychology, 75, 175-184.
- Schmidt, F.L., Hunter, J.E., & Pearlman, K. (1980). Task difference and validity of aptitude tests in selection: A red herring. Journal of Applied Psychology, 66, 166-185.
- Schmidt, F.L., Gast-Rosenberg, I.F., & Hunter, J.E. (1980). Validity generalization results for computer programmers. Journal of Applied Psychology, 65, 643-661.
- Schmidt, F.L., Hunter, J.E., & Caplan, J.R. (1981). Validity generalization results for two job groups in the petroleum industry. Journal of Applied Psychology, 66, 261-273.
- Schmidt, F.L., & Hunter, J.E. (1984). A within-setting test of the situational specificity hypothesis in personnel selection. Pers. Psych., 37, 317-326.
- Schmidt, F.L., Hunter, J.E., Pearlman, K., & Hirsh, H.R. (1985). Forty questions about validity generalization and meta-analysis. Pers. Psych., 38, 697-798.
- Schmidt, F.L., Ocasio, B.P., Hillery, J.M., & Hunter, J.E. (1985). Further within setting empirical tests of the situational specificity hypothesis in personnel selection. Pers. Psych., 39, 509-524.
- Schmidt, F.L., Law, K., Hunter, J.E., Rothstein, H.R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. Journal of Applied Psychology, 78, 3-13.
- Hirsh, H.R., Northrup, L., & Schmidt, F.L. (1986). Validity generalization results for law enforcement occupations. Pers. Psych., 39, 337-344.

- Raju, N.S., Fralicx, R., & Steinhaus, S.D. (1986). Covariance and regression slope models for studying validity generalization. Applied Psychological Measurement, 10, 195-211.
- Hunter, J.E., & Schmidt, F.L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage.

F. Role of the Base Rate in Validity

- Meehl, P.E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. Psych. Bull., 52, 94-216. (Also in Jackson and Messick, No. 33, and Mehrens and Ebel, No. 29.)
- Rorer, et al. (See two articles under "Interpretation of Validity Coefficients.")
- Dawes, R.M. (1962). A note on base rates and psychometric efficiency. J. Consulting Psych., 26, 422-424.
- Schmidt, F.L. (1974). Probability and utility assumptions underlying use of the Strong Vocational Interest Blank. Journal of Applied Psychology, 59, 456-464.

G. Criterion-Related Validity: General

- Schmidt, F.L., Hunter, J.E., Croll, P.R., & McKenzie, R.C. (1983). Estimation of employment test validities by expert judgement. Journal of Applied Psychology, 68, 590-601.
- Schmidt, F.L., & Hunter, J.E. (1978). Moderator research and the law of small numbers. Pers. Psychology, 31, 215-232. (In readings packet)
- Hunter, J.E., & Schmidt, F.L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. Journal of Applied Psychology, 75, 334-349.

H. Test Validity and Minority Groups

- Schmidt, F.L., Berner, J.G., & Hunter, J.E. (1973). Racial differences in validity of employment tests: Reality or illusion? Journal of Applied Psychology, 58, 5-9. (Also in Ford, D.L. (Ed.), Readings in minority group relations, University Associates, 1974.)
- Schmidt, F.L., & Hunter, J.E. (1974). Ethnic and racial bias in psychological tests: Divergent implications of two definitions of test bias. Amer. Psych., 29, 1-8.
- Hunter, J.E., Schmidt, F.L., & Raushenberger, J. (1977). Fairness of psychological tests: Implications of three definitions for selection utility and minority hiring. Journal of Applied Psychology, 62, 245-260.
- Hunter, J.E., & Schmidt, F.L. (1977). A critical analysis of the statistical and ethical implications of various definitions of test fairness. Psych. Bull., 83, 1053-1071.
- Hunter, J.E., & Schmidt, F.L. (1978). Differential and single group validity of employment tests by race: A critical analysis of three recent studies. Journal of Applied Psychology, 63, 1-11.
- Hunter, J.E., Schmidt, F.L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. Psych. Bull., 86, 721-735.

- Schmidt, F.L., Pearlman, K., & Hunter, J.E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. Pers. Psych., 33, 705-724.
- Hunter, J.E. & Schmidt, F.L. (1982). Ability tests: Economic benefits versus the issue of fairness. Industrial Relations, 21, 3, 293-308.
- Hunter, J.E., Schmidt, F.L., & Rauschenberger, J. (1984). Methodological and statistical issues in the study of bias in mental testing. In C.R. Reynolds & R.T. Brown (Eds.), Perspectives on bias in mental testing. New York: Plenum Press.

I. Suppressor Variables

- Collins, J.M., & Schmidt, F.L. (1997). Can suppressor variables enhance criterion-related validity in the personality domain? Educational and Psychological Measurement, 57, 924-936.
- Mark, M.R., Christal, R.E., & Bottenberg, R.A. (1961). A simple formula for understanding the joint action of two predictors. Journal of Applied Psychology, 45, 285-288.

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

- Ebel, R.L. (1956). Obtaining and reporting evidence on content validity. Educ. Psych. Msmt., 16, 269-282. (Also in Chase and Ludlow, No. 9.)
- Mosier, C.I. (1947). A critical examination of the concepts of face validity. Educ. Psych. Msmt., 7, 191-205. (Also in Mehrens and Ebel, No. 22.)
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. Psych. Bull., 52, 281-302. (Also in Mehrens and Ebel, No. 25.) (On Reserve)
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psych. Bull., 56, 81-105. (Also in Mehrens and Ebel, No. 27.) (On Reserve)
- Darlington, R.B. (1970). Some techniques for maximizing a test's validity when the criterion variable is unobserved. J. Educ. Msmt., 7, 1-14.
- Campbell, D.T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. (In Jackson & Messick.)
- Schmitt, N., & Stults, D.M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10, 1-22.
- Schmitt, N., Coyle, B.W., & Saari, B. (1977). A review and critique of analyses of multitrait-multimethod matrices. Multivariate Behavioral Research, 12, 447-448.
- Vance, R.J., MacCallum, R.C., Covert, M.D., & Hedge, J.W. (1988). Construct validity of job performance measures using confirmatory factor analysis. Journal of Applied Psychology, 73, 74-80.
- Marsh, H.W., & Hocerar, D. (1988). A new, more powerful approach to multitrait-multimethod analysis: Application on second order confirmatory factor analysis. Journal of Applied Psychology, 73, 107-117.
- Levin, J. (1988). Multiple group factor analysis of multitrait-multimethod matrices. Multivariate Behavioral Research, 23, 469-479.

- Widaman, K.F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. Applied Psychological Measurement, 9, 1-26.
- Dye, D.A., & Reck, M. (1990). Moderators of the validity of written job knowledge measures. Human Resources Management Review.

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
- Schmidt, F.L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educ. Psych. Msmt., 31, 699-714.
- Schmidt, F.L. (1972). The reliability of differences between linear regression weights in applied differential psychology. Educ. Psych. Msmt., 32, 879-886.
- Lord, F.M. (1962). Cutting scores and errors of measurement. Psychometrika, 27, 19-30.
- Dvorak, Beatrice J. (1956). Advantages of the multiple cut-off method. Pers. Psych., 9, 45-47.
- Darlington, R.B., & Stauffer, G.F. (1966). A method for choosing a cutting point on a test. Journal of Applied Psychology, 229-231.
- Gordon, Mary A. (1954). Empirical comparisons of three multiple correlation techniques. Educ. Psych. Msmt., 14, 133-137.
- Grimely, G. (1949). A comparative study of the Wherry-Doolittle and a multiple cutting score method. Psych. Monog., 63, No. 2. (Also No. 297, pp. 1-24.)
- Jenkins, W.L. (1952). An improved short-cut method for the multiple R. Educ. Psych. Msmt., 12, 316-322.
- Lawshe, C.H., & Patinka, P.J. (1958). An empirical comparison of two methods of test selection and weighting. Journal of Applied Psychology, 42, 210-212.
- Lawshe, C.H., & Schucker, R.E. (1959). The relative efficiency of four test weighting methods in multiple prediction. Educ. Psych. Msmt., 19, 103-114.
- Lawshe, C.H. (1969). Statistical theory and practice in applied psychology. Pers. Psych., 22, 117-124.
- Sevier, F.A. (1957). Testing the assumptions underlying multiple regression. J. Exper. Educ., 25, 323-330.

- Trattner, M.N. (1963). Comparison of three months for assembling aptitude test batteries. Pers. Psych., 16, 221-232.
- Wesman, A.G., & Bennett, G.K. (1959). Multiple regression vs. simple addition of scores in prediction of college students. Educ. Psych. Msmt., 19, 243-246.
- Madden, J.M., & Bottenberg, R.A. (1963). Use of an all possible combinations solution of certain multiple regression problems. Journal of Applied Psychology, 47, 365-366.
- Ghiselli, E.E., & Kahneman, D. (1962). Validity and nonlinear heteroscedostic models. Pers. Psych., 15, 1-11.
- Ghiselli, E.E. (1964). Dr. Ghiselli comments on Dr. Tupes' note. Pers. Psych., 17, 61-63.
- Rock, D.A., Linn, R.L., Evans, F.R., & Patrick, C. (1970). A comparison of predictor selection techniques using Monte Carlo methods. Educ. Psych. Msmt., 30, 873-874.
- Tupes, E.C. (1964). A note on "Validity and nonlinear heteroscedostic models." Pers. Msmt., 17, 59-61.
- Hawk, J.A. (1970). Linearity of criterion-GATB relationships. Measurement and Evaluation in Guidance, 2, 249-251.
- Wilkinson, L. (1979). Tests of significance in stepwise regression. Psychological Bulletin, 86, 168-174. (Shows how ex post facto selection of predictors inflates R-mult and invalidates F-test.)

B. Clinical vs. Actuarial Combining of Test Scores

- Meehl, P.E. (1954). Clinical versus statistical prediction. Minneapolis: University Minnesota Press.
- Meehl, P.E. What can the clinician do well? In Jackson and Messick, No. 48.
- Meehl, P.E. (1956). Wanted--a good cookbook. Amer. Psych., 11, 263-272. (Also in Jackson and Messick, No. 44.)
- Cronbach, L.J. Report on a psychometric mission to Clinicia. In Jackson and Messick, No. 13.
- Sarbin, T.R. (1942). A contribution to the study of actuarial and individual methods of prediction. Amer. J. Sociology, 48, 593-602.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. Psych. Bull., 66, 178-200.
- Cronbach, P. 441 ff.
- Pankoff, L.D., & Roberts, H.V. (1968). Bayesian synthesis of clinical and statistical prediction. Psych. Bull., 70, 762-773.
- Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. American Psychologist, 59, 595 – 613. (A follow up on Sawyer article)

C. Cross-Validation and Double Cross-Validation

- Guilford, 405-406; 440-441.

- Symposium: The need and means of cross-validation. Educ. Psych. Msmt., 1951, 11, 5-28.
- Cureton, E.E. (1950). Reliability, validity, and baloney. Educ. Psych. Msmt., 10, 94-96.
- Kurtz, A.K. (1948). A research test of the Rorschach Test. Pers. Psych., 1, 41-51.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. Psych. Bull., 76, 105-110.

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 $\rho^2(\underline{B})$ instead of $\rho^2(\hat{\underline{B}})$. Also, is a biased estimate.)
- Lord, F.M. (1950). Efficiency of prediction when a regression equation from one sample is used in a new sample. Res. Bull., 40-50. Princeton, NJ: Education Testing Service.
- Schmitt, N., Coyle, B., & Rauschenberger, J. (1977). A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. Psych. Bull., 84, 751-758.
- Nicholson, G.E. (1960). Prediction in future samples. In I. Olkin, et al., Contributions to probability and statistics; essays in honor of Harold Hotelling. Stanford Univ. Press.
- Burket, G. R. (1964). A study of reduced rank models for multiple prediction. Psychometric Monograph., No. 12.
- Cattin, P. (1980). Estimation of the predictive power of a regression model. J. Applied Psych., 65, 401-414. [Provides best formulas for estimating $\rho^2(\hat{\underline{B}})$.] (In readings packet)
- Lautenschlager, G.J. (1990). Sources of imprecision in formula cross-validated multiple correlations. Journal of Applied Psychology, 75, 460-462.

E. Some Applications

- Harker, J.B. (1960). Cross-validation of an IBM proof machine test battery. Journal of Applied Psychology, 44, 237-246.
- Kirkpatrick, J.J. (1951). Cross-validation of a forced-choice personality inventory. Journal of Applied Psychology, 35, 413-416.
- Merenda, P.F., Clarke, W.V., & Hall, C.E. (1961). Cross-validity of procedures for selecting life insurance salesmen. Journal of Applied Psychology, 45, 376-380.
- Schmidt, F.L., Johnson, R.H., & Gugel, J.F. (1978). Estimation of the utility of policy capturing as an approach to graduate admissions decision making. Applied Psych. Msmt., 2, 347-359.

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
 DeVellis, R.E. (1991). Scale development: Theory and applications. Vol. 26 in Applied Social Research Methods Series. Thousand Oakes, CA: Sage (113 pp.)
 Mosier, C.I., Myers, M.C., & Price, Helen G. (1945). Suggestions for the construction of multiple-choice test items. Educ. Psych. Msmt., 5, 261-271. (Also in Chase and Ludlow, No. 31.)
 Engelhart, M.D. (1947). Suggestions for writing achievement test exercises. Educ. Psych. Msmt., 7, 357-374. (Also in Payne and McMorris, No. 20.)
 Travers, R.M.W. (1951). Rational hypotheses in the construction of tests. Educ. Psych. Msmt., 11, 128-137. (Also in Mehrens and Ebel, No. 5.)

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
 Davis, F.B. (1952). Item analysis in relation to educational and psychological testing. Psych. Bull., 40, 97-119.
 Richardson, M.W. (1936). The relation of item difficulty to the differential validity of a test. Psychometrika, 1, 33-49.
 Brozek, J., & Tiede, K. (1952). Reliable and questionable significance in a series of statistical tests. Psych. Bull., 49, 339-341.

C. Cross Validation of Item Analysis

Baker, P.C. (1952). Combining tests of significance in cross-validation. Educ. Psych. Msmt., 12, 300-306.

Katzell, R.A. (1951). Cross-validation of item analysis. Educ. Psych. Msmt., 11, 16-22.

D. Different Approaches to Item Analysis

Walker, Helen M. (1949). Item selection by sequential sampling. Teachers College Record, 50, 404-409.

Lawshe, C.H. (1942). A nomograph for estimating the validity of test items. Journal of Applied Psychology, 26, 846-849.

Lawshe, C.H., & Baker, P.C. (1950). Three aids in the evaluation of the significance of the differences between percentages. Educ. Psych. Msmt., 10, 263-270.

Flanagan, J.C. (1939). General considerations in the selection of test items and a short method of estimating product-moment coefficient from data at the tails of the distribution. J. Educ. Psych., 30, 674-680.

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

Richardson, M.W., & Adkins, D.C. (1938). A rapid method of selecting test items. J. Educ. Psych., 29, 547-552,

Johnson, A.P. (1951). Notes on a suggested index of item validity: The U-L index. J. Educ. Psych., 42, 499-504. (Also in Mehrens and Ebel, No. 32.)

Findley, W.G. (1956). A rationale for evaluation of item discrimination statistics. Educ. Psych. Msmt., 16, 175-180. (Also in Mehrens and Ebel, No. 33.)

Kelly, T.L. (1939). The selection of upper and lower groups for the validation of test items. J. Educ. Psych., 30, 17-24.

Ferguson, G.A. (1949). On the theory of test discrimination. Psychometrika, 14, 61-68.

Feldman, M.J. (1953). The effects of the size of criterion groups and the level of significance in selecting test items on the validity of tests. Educ. Psych. Msmt., 13, 273-279.

Kirkpatrick, J.J., & Cureton, E.E. (1954). Simplified tables for item analysis. Educ. Psych. Msmt., 14, 709-714.

E. Comparisons of Different Item Selection Techniques

Tiffin, J., & Hudson, T.W. (1956). Comparison of sequential and conventional item analysis when used with primary groups varying in size and composition. Educ. Psych. Msmt., 16, 333-344.

Anastasi, A. (1953). An empirical study of the applicability of sequential analysis to item selection. Educ. Psych. Msmt., 13, 3-13.

Guilford, J.P., & Lacey (Eds.) (1947). Printed classification tests. Army Air Forces Aviation Psych. Res. Prgm. Reports, No. 5, Washington D.C.: U.S. Government Printing Office, (Reliability of item indices).

Ely, J.H. (1951). Studies in item analysis. 2. Effects of various methods upon test reliability. Journal of Applied Psychology, 35, 194-203.

- Jurgensen, C.E. (1951). Note on Ely's "Effects of various methods on test reliability." Journal of Applied Psychology, 35, 204.
- Lawshe, C.H., & Mayer, J.S. (1947). Studies in item analysis: The effect of two methods of item validation on test reliability. Journal of Applied Psychology, 31, 271-277.
- Adams, J.F. (1960). Test item difficulty and the reliability of item analysis methods. Journal of Psych., 49, 255-261.
- Adams, J.F. (1960). The effect of nonnormally distributed criterion scores on items analysis. Educ. Psych. Msmt., 20, 317-319.
- Engelhart, M.D. (1965). A comparison of several item discrimination indices. J. Educ. Msmt., 2, 69-76. (Also in Mehrens and Ebel, No. 34.)
- Lawshe, C.H. (1969). Statistical theory and practice in applied psychology. Pers. Psych., 22, 117-124.

F. Other Contributions of Interest

- Lord, F.M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. Psychometrika., 17, 181-194.
- Myers, C.T. (1962). The relationship between item difficulty and test validity and reliability. Educ. Psych. Msmt., 22, 565-571.
- Stanley, J.C., & Wang, M.D. (1970). Weighting test items and test item options: An overview of the analytical and empirical literature. Educ. Psych. Msmt., 30, 21-35.
- Swineford, F. (1959). Some relations between test scores and item statistics. Journal of Educ. Psych., 50, 26-30.
- Travers, R.M. (1942). A note on the value of customary measures of item validity. Journal of Applied Psychology, 26, 625-632.
- Guilford, J.P. (1953). The correlation of an item with a composite of the remaining items in a test. Educ. Psych. Msmt., 13, 87-93.
- Howard, K.I., & Forehand, G.A. (1962). A method for correcting item-total correlations for the effect of relevant item inclusion. Educ. Psych. Msmt., 22, 731-735.
- Fan, C.T. (1952). Item analysis table. Princeton, N.J.: ETS, (For r-tet).

Additional Topics

A. Effects of Guessing

- Carroll, J.B. (1945). The effect of difficulty and chance success on correlations between items or between tests. Psychometrika, 101-119.
- Bryan, M.M., Burke, P.J., & Stewart, N. (1952). Correction for guessing in the scoring of pretests: Effects upon item difficulty and item validity indices. Educ. Psych. Msmt., 12, 45-56.
- 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

- Adkins-Wood, Dorothy. Test construction. Merrill, 1961.
- Anastasi, A. Psychological testing (8th Edition). New York: Macmillan, 1988. (On Reserve)
- APA, AERA, NCME. Standards for Educational and Psychological Testing. American Psychological Association, 1985.
- APA, Division of Industrial-Organizational Psychology. Principles for the Validation and Use of Personnel Selection Procedures (3rd Edition). East Lansing, MI: Author, 1986.
- Barnette, W.L., Jr. Readings in psychological tests and measurements. Dorsey Press, 1968.
- Chase, C.I., & Ludlow, H.G. Readings in educational and psychological measurement. Houghton Mifflin, 1966.
- Cronbach, L.J., & Gleser, Goldine C. Psychological tests and personnel decisions. University of Illinois Press, 1965.
- Dunnette, M.D. Personnel selection and placement. Wadsworth, 1966.
- Edwards, A.L. Techniques of attitude scale construction. Appleton-Century-Crofts, 1957.
- Eysenck, H.J. The structure and measurement of intelligence. New York: Springer-Verlog, 1979 (pp. 11-17). (Class handout)
- Furst, E.J. Constructing evaluation instruments. Longmans, Green, & Co., 1958.
- Ghiselli, E.E. Theory of psychological measurement. McGraw-Hill, 1964.
- Ghiselli, E.E., & Brown, C.W. Personnel and industrial psychology. McGraw-Hill, 1955.
- Guilford, J.P. Psychometric methods. McGraw-Hill, 1954. (On Reserve)
- Guion, R.M. Personnel testing., McGraw-Hill, 1965.
- Gulliksen, H. Theory of mental tests. Wiley, 1950.
- Helmstadter, G.C. Principles of psychological measurement. McGraw-Hill. (On Reserve)
- Horst, P. Psychological measurement and prediction. Brooks-Cole, 1968.

- Hunter, J.E., Schmidt, F.L. & Jackson, G.B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.
- Hunter, J.E. & Schmidt, F.L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Sage.
- Jackson, D.N., & Messick, S. (Eds.) Problems in human assessment. McGraw-Hill, 1967. (On Reserve)
- Lawshe, C.H., & Balma, M.J. Principles of personnel testing. McGraw-Hill, 1966.
- Lemke, E. & Wiersma, W. Principles of psychological measurement. (On Reserve)
- Lindquist, E.F. (Eds.) Educational measurement. American Council on Education, 1951. (On Reserve)
- Linn, R.L. (Ed.) (1989). Educational measurement (3rd Ed.). New York: Macmillan.
- Lord, F.M., & Novick, M.R. Statistical theories of mental test scores. Addison-Wesley, 1968.
- Magnusson, D. Test theory. Addison-Wesley, 1966. (On Reserve)
- Mehrens, W.A., & Ebel, R.L. (Eds.) Principles of educational and psychological measurement. Rand McNally, 1967. (On Reserve)
- Nunnally, J.C. Psychometric theory. McGraw-Hill, 1978. (On Reserve)
- Payne, D.A., & McMorris, R.F. (Eds.) Educational and psychological measurement. Blaisdell, 1967.
- Shartle, C.L. Occupational information (3rd ed.). New York: Prentice-Hall, 1959.
- Tiffin, J., & McCormick, E.J. Industrial psychology. Prentice-Hall, 1965.
- Thorndike, R.L. Personnel selection. Wiley, 1949. (On Reserve)
- Thorndike, R.L. Applied psychometric theory. Wiley, 1982.
- Thorndike, R.L. (Ed.) Educational measurement (2nd Edition). American Council on Education, 1971. (On Reserve)
- Traub, R.E. (1994). Reliability for the social sciences: Theory and applications. Vol. 3 in the series Measurement Methods For the Social Sciences. Thousand Oaks, CA: Sage.