

Psychology 7304/8304

Measurement Theory and Psychometrics

Fall 2008

Ronald S. Landis, Ph.D.

310 Psychology Building
678-4690
rlandis@memphis.edu

Office Hours: W 1-3 or by appointment

Course Basics

Class Time and Room: 2:30pm – 5:30pm, T
244 Psychology Building

Textbooks & Readings:

I do not assign a particular textbook for this course. I will make a copy of my course notes available on the first day of class. These notes represent the bulk of the material we will cover in the course and will serve as the foundation for all tests and project assignments. If you are interested in completing some background readings you may find the following texts useful.

1. DeVellis, R.F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
2. McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Additional readings are also referenced in the syllabus. Readings associated with each of the main topics are broken down into two categories: suggested and additional. The suggested readings are those that I think will be helpful for following the primary course concepts. You are responsible for securing copies of the readings listed as 'Suggested.' Alternatively, the additional readings are provided for those interested in alternative perspectives, current topics, etc.

COURSE OVERVIEW

Measurement is fundamental to psychological research. In fact, some have argued that advances in psychological theory follow new breakthroughs in measurement technology. This course is designed to introduce you to central concepts of psychological measurement. I will take a "nuts and bolts" approach that will cover the critical aspects of psychometric theory including scaling, test construction, item analysis, reliability, validity, and bias. Emphasis will be put on learning the logic and assumptions of various concepts and methods along with providing ample opportunity to apply them. The format of the course will include a combination of lecture, class discussion, and projects.

PROJECTS

There will be six projects for this course. Project topics will include:

1. Scale Construction & Refinement
2. Item Analysis & Item Response Theory
3. Reliability
4. Generalizability Theory, Interrater Reliability, & Agreement
5. Construct Validity & Confirmatory Factor Analysis
6. Multiple Regression

More information about the projects is provided in a separate document.

COURSE EVALUATIONS

Your grade will be based on your performance on the projects (60%) as well as two examinations (40%). Each examination will include an in-class component as well as a take-home component. I expect all projects and examinations to be completed individually.

GRADING SCALE

A	=	90 - 100%
B	=	80 - 89%
C	=	70 - 79%
D	=	60 - 69%
F	=	below 60%

ACADEMIC HONESTY

Plagiarism or cheating may result in your case being referred to the Academic Discipline Committee that addresses academic misconduct. This decision is up to the discretion of the instructor. Students are expected at all times to behave in accordance with the American Psychological Association Code of Ethics. Students should be familiar with the academic regulations outlined in the Graduate Issues of the Bulletin of the University of Memphis and to observe policies regarding student conduct published in the Student Handbook.

DIVERSITY STATEMENT

Diversity means the fair representation of all groups of individuals, the inclusion of minority perspectives and voices, and the appreciation of different cultural and socioeconomic group practices. We aspire to foster and maintain an atmosphere that is free from discrimination, harassment, exploitation, or intimidation. Academic courses will aim at providing opportunities for students to discuss issues of diversity including, but not limited to, ethnicity, gender, disability, and sexual orientation as they can be related to course content. The University of Memphis has adopted policies prohibiting discrimination based upon race, sex, disability, or sexual orientation. In addition, the American Psychological Association has explicit policies regarding the issues of and writing about race, gender, class, sexual orientation, disability, ethnicity, and religion. You may find information on these standards in the APA Publication Manual or on the APA webpage: <http://apa.org/pi/oema/>.

STUDENTS WITH SPECIAL NEEDS OR DISABILITIES

If you have a disability that interferes with completion of this course, please let the instructor know privately at the beginning of the course, and s/he will seek consultation on how best to adapt course materials or instruction. Students with disabilities are encouraged to contact Student Disability Services for the university at 678-2880.

ELECTRONIC COMMUNICATION

Course announcements as well as consultation with the instructor may occur via e-mail messages. Each student must maintain an e-mail account and is responsible for notifying the instructor if their e-mail address changes during the term. The University of Memphis offers free e-mail address changes during the term. Students without an e-mail account can bring their university ID cards, swipe them, and set up an account at the Smith or McWhirter computer labs or students can bring their cards to the information technology helpdesk, room 124 in the Administration Building for assistance.

COURSE READINGS

PART 1

SUGGESTED READING

DeVellis, R.F. (2003). [**Chapters 1 & 2**]

McDonald, R.P. (1999). [**Chapters 1, 4, & 18**]

Judd, CM & McClelland, GH (1998). Measurement. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 180-232). Boston: McGraw-Hill.

Additional References

American Psychologist (Whole Issue) October 1981.

Campbell, J.P. (1976). Psychometric theory. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 185-222). Chicago: Rand McNally.

Cowles, M. (1989). *Statistics in psychology: An historical perspective*. Hillsdale, NJ: Erlbaum.

Cronbach, L.J. (1968). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.

Cronbach, L.J. (1975). Five decades of public controversy over mental testing. *American Psychologist*, 30, 1-14.

Embretson, S.E. & Hershberger, S.L. (1999). Summary and future of psychometric methods in testing. In S.E. Embretson and S.L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know*. Mahwah, NJ: Erlbaum.

Gould, S.J. (1981). *The mismeasure of man*. New York: Norton.

Green, B.F. (1978). In defense of measurement. *American Psychologist*, 33, 664-670.

Jones, L.V. & Appelbaum, M.I. (1989). Psychometric methods. *Annual Review of Psychology*, 40, 23-43.

- Meier, S.T. (1994). *The chronic crisis in psychological measurement and assessment: An historical survey*. San Diego, CA: Academic Press.
- Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, 21, 215-237.
- Owens, W.A. (1968). Toward one discipline of scientific psychology. *American Psychologist*, 23, 782-785.
- Owens, W.A. & Schoenfeldt, L.F. (1979). Toward a classification of persons. *Journal of Applied Psychology*, 65, 569-607.
- Ross, L. & Nisbett, R.E. (1991). *The person and the situation*. New York: McGraw-Hill.
- Stigler, S.M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Skinner, B.F. (1948). *Walden two*. New York: MacMillan.

PARTS 2 & 3

SUGGESTED READING

- DeVellis, R.F. (2003). **[Chapter 5]**
- McDonald, R.P. (1999). **[Chapters 2 & 3]**
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93-105.

Additional References

- Anderson, J.C. & Gerbing, D.W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, 76, 732-740.
- Chen, W. & Bentler, P.M. (1993). The covariance structure analysis of ipsative data. *Sociological Methods of Research*, 22, 214-247.
- Dudycha, A.L. & Carpenter, J. (1973). Effects of item format on item

- discrimination and difficulty. *Journal of Applied Psychology*, 58, 116-121.
- Ebel, R.L. (1982). Proposed solutions to two problems of test construction. *Journal of Educational Measurement*, 19, 267-268.
- Edwards, A.L. (1957). *Techniques for attitude scale construction*. Appleton-Century-Crofts.
- Gardner, P.L. (1975). Scales and statistics. *Review of Educational Research*, 45, 43-57.
- Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Keller, T. & Dansereau, F. (2001). The effect of adding items to scales: An illustrative case of LMX. *Organizational Research Methods*, 4, 131-143.
- Levine, E. (1983). *Everything you wanted to know about job analysis*. Tampa, FL: Mariner.
- Reagan, R.T., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, 74, 433-442.
- Roid, G.H. & Haladyna, T.M. (1982). *A technology for test-item writing*. New York: Academic Press.
- Schmitt, N. & Klimoski, R.J. (1991). *Research methods in human resource management*. Cincinnati, OH: South-Western.
- Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Stine, W.W. (1989). Meaningful inference: The role of measurement in science. *Psychological Bulletin*, 105, 147-155.
- Thorndike, R.L. (1971). *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education.
- Warwick, D.P. & Liniger, C.A. (1975). *The sample survey: Theory and practice*. New York: McGraw-Hill.

PART 4

SUGGESTED READING

DeVillis, R.F. (2003). **[Chapter 7]**

McDonald, R.P. (1999). **[Chapters 11-14]**

Additional References

Applied Psychological Measurement, 1999. [Whole Issue] Volume 23, pp. 187-261.

Adams, J.F. (1960). Test item difficulty and the reliability of item analysis. *Journal of Psychology*, 49, 255-261.

Adams, J.F. (1960). The effect of non-normally distributed criterion scores on item analysis. *Educational and Psychological Measurement*, 20, 317-320.

Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Beuchert, A.K. & Mendoza, J.L. (1979). A Monte-Carlo comparison of ten item discrimination indices. *Journal of Educational Measurement*, 16, 109-118.

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

Ely, J.H. (1951). Studies in item analysis. *Journal of Applied Psychology*, 35, 194-203.

Englehart, M.D. (1965). A comparison of several item discrimination indices. *Journal of Educational Measurement*, 2, 69-76.

Findley, W.G. (1956). A rationale for evaluation of item discrimination statistics. *Educational and Psychological Measurement*, 16, 175-180.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). *Item response theory: Applications to psychological measurement*. Homewood, IL: Dow-Jones-Irwin.

- James, L.R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 1*, 131-163.
- Mislevy, R.J. & Stocking, M.L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement, 13*, 57-76.
- Reckase, M. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 23*, 25-36.
- Reise, S.P. & Widaman, K.F. (1999). Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. *Psychological Methods, 4*, 3-21.
- Steinberg, L. & Thissen, D. (1996). Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychological Methods, 1*, 81-97.
- Thissen, D. & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin, 104*, 385-395.
- Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Erlbaum.
- Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: MESA.
- Wright, B.D. & Stone, M.H. (1979). *Best test design*. Chicago: MESA.

PART 5

SUGGESTED READING

- DeVillis, R.F. (2003). **[Chapter 3]**
- McDonald, R.P. (1999). **[Chapters 5-7]**
- Mislevy, R.J. (1996). Test theory reconceived. *Journal of Educational Measurement, 33*, 379-416.

Additional References

- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

- Brennan, R.L. & Kane, M.T. (1977). An index of dependability for mastery tests. *Journal of Educational Measurement, 14*, 277-290.
- Burke, M.J., Finkelstein, L.M., & Dusig, M.S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods, 2*, 49-68.
- Cardinet, J., Tourneur, Y., & Allal, L. (1981). Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement, 18*, 183-204.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249-254.
- Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.
- Cronbach, L.J. (1947). Test reliability: Its meaning and determination. *Psychometrika, 12*, 1-16.
- Cronbach, L.J. & Furby, L. (1970). How should we measure change - Or should we? *Psychological Bulletin, 74*, 68-80.
- Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Dudek, F.J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin, 86*, 335-337.
- Edwards, J.R. (2001). Ten difference score myths. *Organizational Research Methods, 4*, 265-287.
- Ghiselli, E.E., Campbell, J.P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. New York: Freeman.
- Gibson, W.M. & Weiner, J.A. (1998). Generating random parallel test forms using CTT in a computer-based environment. *Journal of Educational Measurement, 35*, 297-320.
- Glass, G.V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*, 237-262.

- Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ: Lawrence Erlbaum. (Originally published in 1950 by Wiley)
- Hambleton, R.K. & Eigner, D.R. (Winter, 1978). A practitioner's guide to criterion-referenced test development, validation, and test score usage. *Journal of Educational Measurement*.
- Hambleton, R.K., Swaminathan, H., Algina, J., & Coulson, D.B. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research*, 48, 1-47.
- Huynh, H. (1976). On consistency of decisions in criterion referenced testing. *Journal of Educational Measurement*, 13, 253-264.
- Li, H., Rosenthal, R., & Rubin, D.B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, 1, 98-107.
- Livingston, S.A. (1972). Criterion-referenced applications of classical test theory. *Journal of Educational Measurement*, 9, 13-21.
- Lindell, M.K. & Brandt, C.J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of the CVI, T, $r_{WG}(J)$, and $r_{WG}^*(J)$ indexes. *Journal of Applied Psychology*, 84, 640-647.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- McGraw, K.O. & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46.
- Miller, M.B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, 2, 255-273.
- Novick, M.R. & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1-13.
- Osburn, H.G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5, 343-355.
- Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 92, 726-748.

- Shavelson, R.J. & Webb, N.M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- 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.
- Schmidt, F.L. & Hunter, J.E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1, 199-223.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 81-84.
- Schmitt, N., Pulakos, E.D., & Lieblein, A. (1984). Comparison of three techniques to assess group-level beta and gamma change. *Applied Psychological Measurement*, 8, 249-260.
- Stanley, J.C. (1967). General and special formulas for reliability of differences. *Journal of Educational Measurement*, 4, 249-252.
- Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- Subkoviak, M.J. (1976). Estimating reliability from a single administration of a mastery test. *Journal of Educational Measurement*, 13, 265-276.
- Suen, H.K. (1990). *Principles of test theories*. Hillsdale, NJ: Erlbaum.
- Tinsley, H.A. & Weiss, D.J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22, 358-376.
- van den Bergh, H. & Eiting, M.H. (1989). A method of estimating rater reliability. *Journal of Educational Measurement*, 26, 29-40.

PARTS 6 & 7

SUGGESTED READING

- DeVillis, R.F. (2003). **[Chapters 4 & 6]**

McDonald, R.P. (1999). **[Chapters 8-10]**

Angoff, W.H. (1988). Validity: An evolving concept. In H. Wainer & H.I. Braun (Eds.), *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum.

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.

Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer & H.I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.

Hubley, A.M. & Zumbo, B.D. (1996). A dialectic on validity: Where have we been and where we are going. *Journal of General Psychology*, 123, 207-215.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H.I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.

Additional References

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1985). *Standards for Educational and Psychological Tests*. Washington, DC: APA.

Aiken, L.S. & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.

Astin, A.W. (1964). Criterion centered research. *Educational and Psychological Measurement*, 24, 807-821.

Bangert-Downs, R.L. (1986). Review of developments in meta-analytic methods. *Psychological Bulletin*, 99, 388-399.

Barrett, G.V., Caldwell, M.S., & Alexander, R.A. (1985). The concept of dynamic criteria: A critical reanalysis. *Personnel Psychology*, 38, 41-56.

Barrett, G.V., Phillips, J.S., & Alexander, R.A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied*

Psychology, 66, 1-6.

- Berry, W.D. (1984). *Nonrecursive causal models*. Newbury Park, CA: Sage.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bollen, K. & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305-314.
- Brogden, H.E. & Taylor, E.K. (1950). A theory and classification of criterion bias. *Educational and Psychological Measurement*, 10, 159-186.
- Brogden, H.E. & Taylor, E.K. (1950). The dollar criterion: Applying cost accounting concepts to criterion construction. *Personnel Psychology*, 3, 133-154.
- Bullock, H.E., Harlow, L.L. & Mulaik, S.A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling*, 1, 253-267.
- Byrne, B.M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Campion, J.E. (1972). Work sampling for personnel selection. *Journal of Applied Psychology*, 56, 40-44.
- Cascio, W.F., Alexander, R.A., & Barrett, G.V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1-24.
- Cattin, P. (1980). Estimation of the predictive power of a regression model. *Journal of Applied Psychology*, 65, 407-414.
- Cohen, J. (1978). Partialled products are interactions: Partialled powers are curve components. *Psychological Bulletin*, 85, 858-866.
- Covert, M.D., Craiger, P.J., Teachout, M.S. (1997). Effectiveness of the direct product versus confirmatory factor model for reflecting the structure of multimethod-multirater job performance data. *Journal of Applied Psychology*, 82, 271-280.

- Cortina, J.M., Chen, G., & Dunlap, W.P. (2001). Testing interaction effects in LISREL: Examination and illustration of available procedures. *Organizational Research Methods, 4*, 324-360.
- Crocker, L., Llabre, M., & Miller, M.D. (1988). The generalizability of content validity ratings. *Journal of Educational Measurement, 25*, 287-299.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Cudek, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*, 317-327.
- Cureton, E.E. (1950). Reliability, validity, and baloney. *Educational and Psychological Measurement, 10*, 94-96.
- Doty, D.H. & Glick, W.H. (1998). Common methods bias: Does common methods variance really bias results? *Organizational Research Methods, 1*, 374-406.
- Duncan, O.D. (1975). *Introduction to structural equations models*. New York: Academic Press.
- Dunnette, M.D. (1963). A note on criterion. *Journal of Applied Psychology, 47*, 251-254.
- Dvorak, B.J. (1956). Advantages of the multiple cutoff method. *Personnel Psychology, 9*, 45-47.
- Edwards, J.R. (2001). Multidimensional constructs in organizational research: An integrative analytical framework. *Organizational Research Methods, 4*, 144-192.
- Edwards, J.R. & Bagozzi, R.P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*, 155-174.
- Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.
- Fine, S.A. & Wiley, W.W. (1974). An introduction to functional job analysis. In E.A. Fleishman & A.R. Bass (Eds.), *Studies in personnel and industrial*

- psychology*. Homewood, IL: Dorsey.
- Fiske, D.W. (1973). Can a personality construct be empirically validated? *Psychological Bulletin*, 80, 89-92.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-355.
- Frederiksen, N. (1986). Construct validity and construct similarity: Methods for use in test development and test validation. *Multivariate Behavioral Research*, 21, 3-28.
- Gael, S. (1983). *Job analysis: A guide to assessing work activities*. San Francisco, CA: Jossey-Bass.
- Gael, S. (1988). *The job analysis handbook for business, industry, and government*. New York: Wiley.
- Ghiselli, E.E. (1956). Dimensional problems of criteria. *Journal of Applied Psychology*, 40, 1-4.
- Ghiselli, E.E., Campbell, J.P., & Zedeck, S. (1981). Chapter 10. *Measurement theory for the behavioral sciences*. New York: Freeman.
- Ghiselli, E.E. & Haire, M. (1961). The validation of selection tests in the light of the dynamic character of criteria. *Personnel Psychology*, 14, 141-149.
- Guion, R.M. (1977). Content validity - the source of my discontent. *Applied Psychological Measurement*, 1, 1-10.
- Guion, R.M. (1978). Content validity in moderation. *Personnel Psychology*, 31, 205-214.
- Guion, R.M. & Cranny, C.J. (1982). A note on concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, 67, 239-244.
- Gulliksen, H.O. (1950). Intrinsic validity. *American Psychologist*, 5, 511-517.
- Hall, R.J., Snell, A.F., & Foust, M.S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, 2, 233-256.
- Hattrup, K., Schmitt, N., & Landis, R.S. (1992). Equivalence of constructs

- measured by job-specific and commercially available aptitude tests. *Journal of Applied Psychology*, 77, 298-308.
- Hayduk, L.A. (1987). *Structural equation modeling with LISREL*. Baltimore, MD: Johns Hopkins University Press.
- Hedges, L.V. (1988). The meta-analysis of test validity studies: Some new approaches. In H. Wainer & H.I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Erlbaum.
- Hinkin, T.R. & Tracey, J.B. (1999). An analysis of variance approach to content validation. *Organizational Research Methods*, 2, 175-186.
- Hu, L. & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Hu, L. & Bentler, P.M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
- Jackson, D.L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analyses: A Monte Carlo investigation. *Structural Equation Modeling*, 8, 224-256.
- James, L.R. & Brett, J.M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, 69, 307-321.
- James, L.R., Demaree, R.G., Mulaik, S.A., & Mumford, M.D. (1988). Validity generalization: Rejoinder to Schmidt, Hunter, and Raju. *Journal of Applied Psychology*, 73, 673-678.
- James, L.R., Mulaik, S.A., & Brett, J.M. (1982). *Causal analysis: Assumptions, models, & data*. Beverly Hills, CA: Sage.
- Jöreskog, K.G. & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood and least squares methods*. Mooresville, IN: Scientific Software, Inc.

- Kenny, D.A. (1979). *Correlation and causality*. New York: Wiley.
- Landis, R.S., Beal, D.J., & Tesluk, P.E. (2000). A comparison of approaches to forming composite measures in structural equation models. *Organizational Research Methods, 3*, 186-207.
- Landis, R.S., Dempsey, M.T., & Overstreet, S. (2003). Conceptualizing exposure to community violence: An application of confirmatory factor analysis. *School Psychology Quarterly, 18*, 303-324.
- Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.
- Lawshe, C.H. & Schucker, R.E. (1959). The relative efficiency of four test weighting methods in multiple prediction. *Educational and Psychological Measurement, 9*, 45-47.
- Linn, R.L. (1983). Person selection formulas: Implications for studies on predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement, 20*, 1-16.
- Linn, R.L., Harnisch, D.L., & Dunbar, S.B. (1981). Corrections for range restriction: An empirical investigation of conditions resulting in conservative corrections. *Journal of Applied Psychology, 66*, 665-663.
- Livingston, S.A. & Zieky, M.J. *Passing scores*. Princeton, NJ: Educational Testing Service.
- Lord, F.M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement, 21*, 239-244.
- MacCallum, R.C., Roznowski, M., & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490-504.
- MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-99.
- Marsh, H.W., Balla, J.R., & MacDonald, R.P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Journal of Applied Psychology, 103*, 391-410.
- McCormick, E.J., Jeanneret, P.R., & Mechem, R.C. (1972). A study of job

- characteristics and job dimensions as based on the PAQ. *Journal of Applied Psychology*, 56, 347-368.
- McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
- Meehl, P.E. (1954). *Clinical versus actuarial prediction*. Minneapolis, MN: University of Minnesota Press.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Mulaik, S.A., James, L.R., Van Alitine, J., Bennett, N., Lind, S., & Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.
- Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalization results for tests used to predict training success and job proficiency in clerical occupations. *Journal of Applied Psychology*, 65, 373-406.
- Primoff, E.S. (1973). *How to prepare and conduct job element examinations*. Washington, DC: Personnel Research and Development Center. U.S. Civil Service Commission.
- Ree, M.J., Carretta, T.R., & Earles, J.A. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulae. *Journal of Applied Psychology*, 79, 298-301.
- Ree, M.J., Carretta, T.R., & Earles, J.A. (1998). In top-down decisions, weighting variables does not matter: A consequence of Wilks' theorem. *Organizational Research Methods*, 1, 407-420.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 66, 178-200.
- Schmidt, F.L. (1971). The relative efficiency of regression and simple unit weights in applied differential psychology. *Educational and Psychological Measurement*, 31, 699-714.
- Schmidt, F.L. & Hunter, J.E. (1978). Moderator research and the law of small

- numbers. *Personnel Psychology*, 31, 215-231.
- Schmidt, F.L. & Hunter, J.E. (1980). The future of criterion-related validity. *Personnel Psychology*, 33, 41-60.
- Schmidt, F.L. & Hunter, J.E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128-1137.
- Schmidt, F.L., Hunter, J.E., & Raju, N.S. (1988). Validity generalization and situational specificity: A second look at the 75% rule and Fisher's Z transformation. *Journal of Applied Psychology*, 73, 665-672.
- Schmidt, F.L., Hunter, J.E., & Urry, V.W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, 66, 473-485.
- Schmidt, F.L. & Kaplan, L.B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. *Personnel Psychology*, 24, 419-434.
- Schmitt, N., Cortina, J.M., & Whitney, D. (1993). Appropriateness fit indices and criterion-related validity. *Applied Psychological Measurement*, 17, 143-150.
- Schmitt, N., Coyle, B.W., & Rauschenberger, J.M. (1977). A monte-carlo evaluation of three formula estimates of cross-validated multiple correlation. *Psychological Bulletin*, 84, 751-758.
- Schmitt, N., Gilliland, S.W., Landis, R.S., & Devine, D. (1993). Computer-based testing applied to selection of secretarial applicants. *Personnel Psychology*, 46, 149-166.
- Schmitt, N., Gooding, R.Z., Noe, R.A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Schmitt, N. & Landy, F.J. (1992). The concept of validity. In N. Schmitt & W.C. Borman (Eds.), *Personnel selection*. San Francisco, CA: Jossey-Bass.
- Schmitt, N. & Levine, R. (1977). Statistical versus subjective weights: Some problems and proposals. *Organization Behavior and Human Performance*, 18, 95-106.
- Schmitt, N. & Ostroff, C. (1986). Operationalizing the behavioral consistency approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, 39, 91-108.

- Schwab, D.P. (1980). Construct validity in organizational behavior. *Research in Organizational Behavior*, 2, 3-43.
- Scullen, S.E. (1999). Using confirmatory factor analysis of correlated uniquenesses to estimate method variance in multitrait-multimethod matrices. *Organizational Research Methods*, 2, 275-292.
- Society for Industrial and Organizational Psychology (1987). *Principles for the validation and use of personnel selection procedures*. Lake Mills, IA: Graphics Publishing.
- Stone, E.F. (1988). Moderator variables in research: A review and analysis of conceptual and methodological issues. In K.R. Rowland & G. Ferris (Eds.), *Research in Human Resources Management*. Greenwich, CT: JAI.
- Sussman, M. & Robertson, D.U. (1986). The validity of validity: An analysis of validation study designs. *Journal of Applied Psychology*, 71, 461-468.
- Travers, R.M.W. (1951). Rational hypotheses in the construction of tests. *Educational and Psychological Measurement*, 11, 128-137.
- Turban, D.B., Sanders, P.A., Francis, D.J., & Osburn, H.G. (1989). Construct equivalence as an approach to replacing validated cognitive ability tests. *Journal of Applied Psychology*, 74, 62-71.
- Vandenberg, R. & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations. *Organizational Research Methods*, 3, 4-69.
- Wainer, H. & Braun, H.I. (1988). *Test validity*. Hillsdale, NJ: Erlbaum.
- Wallace, S.R. (1965). Criteria for what. *American Psychologist*, 20, 411-418.
- Widaman, K.F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9, 1-26.
- Zedeck, S. (1971). Problems with the use of moderator variables. *Psychological Bulletin*, 76, 295-310.

PART 8

SUGGESTED READINGS

- Bartlett, C.J., Bobko, P., Mosier, S.B., & Hannan, R. (1978). Testing for fairness with a moderated regression strategy: An alternative to differential analysis. *Personnel Psychology, 31*, 233-242.
- Cascio, W.F., Outtz, J., Zedeck, S., & Goldstein, I.L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance, 4*, 233-264.
- Chan, K.Y., Drasgow, F. & Sawin, L.L. (1999). What is the shelf life of a test? The effect of time on the psychometrics of a cognitive ability test battery. *Journal of Applied Psychology, 84*, 610-619.
- Humphreys, L.G. (1986). An analysis and evaluation of test and item bias in the prediction context. *Journal of Applied Psychology, 71*, 327-333.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H.I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.

Additional References

- American Psychologist* (1981). Entire Issue, October.
- Journal of Vocational Behavior* (1988). Entire Issue, Vol. 33, No. 3.
- Journal of Educational Measurement* (1976). Entire Issue, Spring.
- Anastasi, A. Some emerging trends in psychological measurement: A fifty year perspective. *Applied Psychological Measurement, 9*, 121-138.
- Boehm, V.R. (1977). Differential prediction: A methodological artifact? *Journal of Applied Psychology, 62*, 146-154.
- Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement, 5*, 115-124.
- Cleary, T.A., Humphreys, L.C., Kendrick, S.A., & Wesman, A. (1975). Educational use of tests with disadvantaged students. *American Psychologist, 30*, 15-41.
- Cronbach, L.J., Yallow, E., & Schaffer, G. (1980). A mathematical structure for analyzing fairness in selection. *Personnel Psychology, 33*, 693-704.

- Darlington, R.B. (1971). Another look at "cultural fairness." *Journal of Educational Measurement*, 8, 71-82.
- Drasgow, F. (1987). Study of measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72, 19-29.
- Flaugher, R.L. (1978). The many definitions of test bias. *American Psychologist*, 33, 671-679.
- Guion, R.M. (1998). Analysis of bias (Chapter 10). *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahwah, NJ: Erlbaum.
- Hartigan, J.A. & Wigdor, A.K. (1989). *Fairness in employment testing*. Washington, DC: National Academy Press.
- Hernnstein, R.J. & Murray, C. (1994). *The bell curve: Intelligence and class structure in american life*. New York: Free Press.
- Humphreys, L.G. (1973). Statistical definitions of test validity for minority groups. *Journal of Applied Psychology*, 58, 1-4.
- Hunter, J.E. & Schmidt, F.L. (1976). Critical analysis of the statistical and ethical implications of various definitions of test bias. *Psychological Bulletin*, 83, 1053-1071.
- Hunter, J.E., Schmidt, F.L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721-735.
- Hunter, J.E., Schmidt, F.L., & Rauschenberger, J.M. (1977). Fairness of psychological tests: Implications of four definitions for selection utility and minority hiring. *Journal of Applied Psychology*, 62, 245-260.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Linn, R.L. (1978). Single group validity, differential validity, and differential prediction. *Journal of Applied Psychology*, 63, 507-512.
- Linn, R.L. (1984). Selection bias: Multiple meanings. *Journal of Educational Measurement*, 21, 33-47.
- Linn, R.L. & Werts, C.E. (1971). Considerations for studies of test bias. *Journal of Educational Measurement*, 8, 1-4.

- Roznowski, M. (1987). Use of tests manifesting sex differences as measures of intelligence: Implications for measurement bias. *Journal of Applied Psychology, 72*, 480-483.
- Sackett, P.R., Schmitt, N., Ellingson, J.E., & Kabin, M.B. (2001). High stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist, 56*, 302-318.
- Schmidt, F.L., Berner, J., & Hunter, J.E. (1973). Racial differences in validity of employment tests: Reality or illusion? *Journal of Applied Psychology, 58*, 5-9.
- Schmidt, F.L. & Hunter, J.E. (1974). Racial and ethnic bias in psychological tests: Divergent implications of two definitions of test bias. *American Psychologist, 29*, 1-3.
- Schmitt, N., Hattrup, K., & Landis, R.S. (1993). Item bias indices based on total test score and job performance estimates of ability, *Personnel Psychology, 46*, 593-611.
- Schmitt, N., Mellon, P.M., & Bylenga, C. (1978). Sex differences in validity for academic and employment criteria, and different types of predictors. *Journal of Applied Psychology, 63*, 145-150.
- Schmitt, N. & Noe, R.A. (1986). Personnel selection and equal employment opportunity. In C.L. Cooper & I. Robertson (Eds.), *Review of industrial and organizational psychology*. New York: Wiley.
- Thorndike, R.L. (1971). Concepts of culture fairness. *Journal of Educational Measurement, 8*, 63-70.
- Tryon, W.W. (1979). The test-trait fallacy. *American Psychologist, 34*, 402-406.
- Zwick, R. & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement, 26*, 55-66.