

References

- Achieve. (2004). *Measuring up 2004: A report on language arts literacy and mathematics standards and assessments for New Jersey*. Washington, DC: Author.
- Airasian, P. W. (2001). *Classroom assessment* (4th ed.). New York: McGraw-Hill.
- Albertson, B. (1998). *Creating effective writing prompts*. Newark: Delaware Reading and Writing Project, Delaware Center for Teacher Education, University of Delaware.
- Alexander, P. A. (1992). Domain knowledge: Evaluating themes and emerging concerns. *Educational Psychologist, 27*, 33-51.
- Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 185-202). New York: Guilford Press.
- Allen, R., Bettis, N., Kurfman, D., MacDonald, W., Mullis, I. V. S., & Salter, C. (1990). *The geography learning of high school seniors*. Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service.
- American Association on Intellectual and Developmental Disabilities. (2002). *Definition of intellectual disability*. Retrieved October 12, 2009, from http://www.aaid.org/content_96.cfm?navID=20
- American Educational Research Association. (2002). *Ethical standards of the American Educational Research Association*. Retrieved from <http://www.aera.net>
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- American Federation of Teachers et al. (1990). *Standards for Teacher Competence in Educational Assessment of Students*. American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: National Council on Measurement in Education. Available from <http://www.unl.edu/buros/bimm/html/subarts.html>
- American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Retrieved from <http://www.apa.org/ethics/homepage.html>
- Anastasi, A. (1988). *Psychological testing* (6th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*, 199-210.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete ed.). New York: Longman.
- Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research, 42*, 145-170.
- Arter, J. A. (1998, April). *Teaching about performance assessment*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom*. Thousand Oaks, CA: Corwin Press.
- Arter, J. A., & Spandel, V. (1992). NCME instructional module: Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice, 11*(1), 36-44.
- Arter, J. A., & Stiggins, R. J. (1992, April). *Performance assessment in education*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Assessment Reform Group. (2002). *Assessment is for learning: 10 principles*. Downloadable from <http://www.assessment-reform-group.org>
- Association for Assessment in Counseling and Education. (2003). *Responsibilities of Users of Standardized Tests (RUST)* (3rd ed.). Retrieved from <http://aac.nacat.edu/Resources/documents/RUST2003%20VII%20Final.pdf>
- Azwell, T., & Schmar, E. (Eds.). (1995). *Report on report cards: Alternatives to consider*. Portsmouth, NH: Heinemann.
- Baglin, R. F. (1981). Does "nationally" normed really mean nationally? *Journal of Educational Measurement, 18*, 92-107.
- Baker, E. L. (1992). Issues in policy, assessment, and equity. In *Proceedings of the National Research Symposium on Limited*

REFERENCES

- English Proficiency Student Issues: Vol. 1 and 2: Focus on Evaluation and Measurement*. Washington, DC.
- Baker, F. (2001). *The basics of item response theory*. College Park: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland. Available from <http://edres.org/irt>
- Barker, K., & Ebel, R. L. (1981). A comparison of difficulty and discrimination values of selected true-false item types. *Contemporary Educational Psychology*, 7, 35–40.
- Baron, J. B. (1991). Strategies for the development of effective performance exercises. *Applied Measurement in Education*, 4, 305–318.
- Becker, K. A. (2003). *History of the Stanford-Binet Intelligence Scales: Content and psychometrics*. Retrieved July 28, 2005, from <http://www.assess.nelson.com/pdf/sb5-asb1.pdf>
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Fifth edition manual for the Differential Aptitude Tests (Forms S and T)*. New York: Psychological Corporation.
- Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *Journal of Experimental Psychology*, 57, 247–261.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–74.
- Blank, R. K. (2002). Using surveys of enacted curriculum to evaluate quality of instruction and alignment with standards. *Peabody Journal of Education*, 77(4), 86–121.
- Blommers, P. J., & Forsyth, R. A. (1977). *Elementary statistical methods in psychology and education* (2nd ed.). Boston: Houghton Mifflin.
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1, 1–12.
- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals, Handbook I: Cognitive domain*. White Plains, NY: Longman.
- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1984). *Taxonomy of educational objectives book I: Cognitive domain*. Boston: Allyn & Bacon.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bond, L. (1989). The effects of special preparation on measures of scholastic ability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 429–444). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Boothroyd, R. A., McMorris, R. F., & Pruzek, R. (1992, April). *What do teachers know about testing and how did they find out?* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Bracken, B. A. (2002). *Bracken School Readiness Assessment: Administration manual*. San Antonio, TX: The Psychological Corporation.
- Bransford, J. D., & Stein, B. S. (1984). *The IDEAL problem solver*. New York: W. H. Freeman.
- Brelend, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). *Assessing writing skill*. (Research Monograph No. 11.). New York: College Entrance Examination Board.
- Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, 38, 295–317.
- Brookhart, S. M. (1991). Letter: Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35–36.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30, 123–142.
- Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, 18(1), 5–13.
- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education*, 8, 153–169.
- Brookhart, S. M. (2009). *Grading* (2nd ed.). Upper Saddle River, NJ: Prentice Hall/Merrill Education.
- Brown, R. S., & E. Coughlin. (2007, November). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region* (Issues & Answers Report, REL 2007–No. 017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296–322.
- Buros, O. K. (Ed.). (1938). *The nineteen thirty-eight mental measurements yearbook*. New Brunswick, NJ: Rutgers University Press.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245–281.
- Cabrera, N. L., & Cabrera, G. A. (2008). Counterbalance assessment: The chorizo test. *Phi Delta Kappan*, 89(9), 677–678.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Westport, CT: Praeger.
- Carlson, S. B. (1985). *Creative classroom testing: Ten designs for assessment and instruction*. Princeton, NJ: Educational Testing Service.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.
- Carroll, J. B. (1974). The aptitude-achievement distinction: The case of foreign language aptitude and proficiency. In D. R. Green (Ed.), *The aptitude-achievement distinction: Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement*. Monterey, CA: CTB/McGraw-Hill, Inc.
- Center for the Study of Testing, Evaluation, and Educational Policy. (1992, October). *The influence of testing on teaching math and science in grades 4–12*. Boston: Boston College.
- Champagne, A. B., & Klopfer, L. E. (1980). *Using the ConSAT: A memo to teachers*. (LRDC Reports to Educators RTE/4). Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.
- Christiansen, L. L., Lazarus, S. S., Crone, M., & Thurlow, M. L. (2008). *2007 state policies on assessment participation and accommodations for students with disabilities* (Synthesis Report 69). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Cizek, G. J. (2001). *Setting performance standards*. Mahwah, NJ: Erlbaum.
- Clarridge, P. B., & Whitaker, E. M. (1997). *Rolling the elephant over: How to effect large-scale change in the reporting process*. Portsmouth, NH: Heinemann.
- Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.

REFERENCES

- Cohen, S. A., & Hyman, J. S. (1991). Can fantasies become facts? *Educational Measurement: Issues and Practice*, 10(1), 20–23.
- Cole, N. S., & Hanson, G. R. (1975). Impact of interest inventories on career choice. In E. E. Diamond (Ed.), *Issues of sex bias and sex fairness in career interest measurement*. Washington, DC: Career Education Program, National Institute of Education, Department of Health, Education, and Welfare.
- Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 201–219). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Cole, N. S., & Nitko, A. J. (1981). Instrumentation and bias: Issues in selecting measures for educational evaluations. In R. A. Berk (Ed.), *Educational evaluation methodology: The state of the art*. Baltimore: Johns Hopkins University Press.
- Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38, 369–382.
- Collis, K. F. (1991). *Assessment of the learned structure in elementary mathematics and science*. Paper presented at the Assessment in the Mathematical Sciences Conference, Victoria, Australia.
- Committee on Education and the Workforce. (2005). *Full history of the ESEA effort: Press releases, summaries, and information related to H.R. 1, the Reauthorization of the Elementary and Secondary Education Act*. Retrieved from <http://edworkforce.house.gov/democrats/eseainfo.html>
- Connolly, A. J. (2007). *KeyMath™-3 Diagnostic Assessment*. Pearson.
- Corby, K. (2002). *Tests and testing information*. East Lansing: Michigan State University, University Library.
- Council of Chief State School Officers. (2005). *Alignment analysis*. Washington, DC: Author. Available from http://www.ccsso.org/Projects/alignment_analysis
- Covington, M. V. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. Cambridge: Cambridge University Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672–683.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), *Test validity* (pp. 3–17). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 147–171). Urbana: University of Illinois Press.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- CTB/McGraw-Hill. (2008). *TerraNova, third edition, technical report*. Monterey, CA: Author.
- Culler, R. E., & Holahan, C. J. (1980). Test anxiety and academic performance: The effects of study-related behaviors. *Journal of Educational Psychology*, 72, 16–20.
- Cyert, R. M. (1980). Problem solving and educational policy. In D. T. Tuma & F. Reif (Eds.), *Problem solving and education: Issues in teaching and research*. Hillsdale, NJ: Erlbaum.
- Darden, A. D. (2000). Thoughtful conversations: Student-teacher writing conferences. *Trade Secrets: Teaching Tips for Elementary, Middle, and High School Teachers*, 20(1), 3–5.
- Das, J. P. (2002). A better look at intelligence. *Current Directions in Psychological Science*, 11(1), 28–33.
- Davey, B., & Rindone, D. A. (1990, April). *Anatomy of a performance task*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.
- Davis, H. A., & Li, J. (2008). *The relationship between high school students' cognitive appraisals of high stakes tests and their emotion regulation and achievement*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Davis, R. V. (1980). Measuring interests. In D. A. Payne (Ed.), *New directions for testing and measurement: Recent developments in affective measurement* (No. 7). San Francisco: Jossey-Bass.
- Debra P. v. Turlington*, 474 F. Supp. 244 (M.D.Fla.1979)
- Debra P. v. Turlington*, 644 F.2d 397, 408 (5th Cir.1981) (Unit B)
- Debra P. v. Turlington*, 730 F.2d 1405 (11th Cir. 1984)
- Donlon, T. F., & Angoff, W. H. (1971). The Scholastic Aptitude Test. In W. H. Angoff (Ed.), *The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and achievement tests*. New York: College Entrance Examination Board.
- Dorans, N. J. (2002). *The recentering of SAT scales and its effects on score distributions and score interpretations*. Research Report No. 2002-11. New York: College Entrance Examination Board.
- Downing, S. M., Baranowski, R. A., Grosso, L. J., & Norcini, J. J. (1995). Stem type and cognitive ability measured: The validity evidence for multiple true-false items in medical specialty certification. *Applied Measurement in Education*, 8, 187–197.
- Dudycha, A. L., & Dudycha, L. W. (1972). Behavioral statistics: An historical perspective. In R. E. Kirk (Ed.), *Statistical issues: A reader for the behavioral sciences*. Monterey, CA: Brooks/Cole.
- Dunbar, D. A., Float, B., & Lyman, F. J. (1980, November). *Report card revision steering committee final report*. Mount Lebanon, PA: Mount Lebanon School District.
- Dunbar, S., Hoover, H. D., Frisbie, D. A., & Mengeling, M. A. (2008). *Iowa Tests of Basic Skills complete and core batteries, Form C: 2005 norms and score conversions*. Chicago: Riverside.
- Dunbar, S., Hoover, H. D., Frisbie, D. A., & Oberley, K. R. (2008). *Interpretive guide for school administrators: Iowa Tests of Basic Skills, Levels 5–14, Forms A, B, and C*. Chicago: Riverside.
- Dunbar, S. B., Koretz, D., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, 4(4), 289–303.
- Duran, R. P. (1989). Testing of linguistic minorities. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 573–587). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Duschl, R. A., & Gitomer, D. H. (1991). Epistemological perspectives on conceptual change: Implications for educational practice. *Journal of Research in Science Teaching*, 28, 839–858.
- Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliffs, NJ: Prentice Hall.
- Ebel, R. L. (1972). *Essentials of educational measurement* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Ebel, R. L. (1974). Shall we get rid of grades? *Measurement in Education*, 5(4), 1–2.
- Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

REFERENCES

- Egawa, K., & Azwell, T. (1995). Telling the story: Narrative reports. In T. Azwell & E. Schmar (Eds.), *Report on report cards: Alternatives to consider*. Portsmouth, NH: Heinemann.
- Ennis, R. H. (1985). Goals for a critical thinking curriculum. In A. Costa (Ed.), *Developing minds: A resource book for teaching thinking*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Justice, Department of Labor, & Department of the Treasury. (1979). Adoption of questions and answers to clarify and provide a common interpretation of the uniform guidelines on employee section procedures. *Federal Register*, 44 (Publication Number: 11996-12006).
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice (1978, August 25). Uniform guidelines on employee selection procedures. *Federal Register*, 43 (Publication Number: 38290-38315).
- Ergene, T. (2003). Effective interventions on test anxiety reduction: A meta-analysis. *School Psychology International*, 24, 313-328.
- Ericsson, K. A., & Simon, H. A. (1999). *Protocol analysis: Verbal reports as data*. Cambridge: Massachusetts Institute of Technology.
- Evaluation Center. (1995). *An independent evaluation of the Kentucky Instructional Results Information System (KIRIS)*. Frankfort: Kentucky Institute for Education Research.
- Feister, W. J., & Whitney, D. R. (1968). An interview with D. E. F. Linn. *Epsilon Bulletin*, 42, 17-28.
- Feldt, L. S. (1967). A note on the use of confidence bands to evaluate the reliability of a difference between two scores. *American Educational Research Journal*, 4, 139-145.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.
- Ferguson, R. L. (1970). A model for computer-assisted criterion-referenced measurement. *Education*, 81, 25-31.
- Fischer, R. J. (1994). The Americans with Disabilities Act: Implications for measurement. *Educational Measurement: Issues and Practice*, 13(4), 17-26, 37.
- Fisher, T. H. (1980). The courts and your minimum competency testing program—A guide to survival. *NCME Measurement in Education*, 11(1), 1-12.
- Flanagan, J. C. (1967). Functional education for the seventies. *Phi Delta Kappan*, 49, 27-32.
- Flanagan, J. C. (1969). Program for learning in accordance with needs. *Psychology in the Schools*, 6, 133-136.
- Flaugher, R. L. (1978). The many definitions of test bias. *American Psychologist*, 33, 671-679.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Erlbaum.
- Forster, M., & Masters, G. (2004). Bridging the conceptual gap between classroom assessment and system accountability. In M. Wilson (ed.), *Towards coherence between classroom assessment and accountability*. Chicago, IL: University of Chicago Press.
- Forsyth, R. A. (1976, March). *Describing what Johnny can do* (Iowa Testing Program, Occasional Paper, No. 17). Iowa City: University of Iowa.
- Frank, L. K. (1939). Projective methods for the study of personality. *Journal of Psychology*, 8, 389-413.
- Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem-solving. *Review of Educational Research*, 54, 363-407.
- Freeman, D. J., Kuhs, T. M., Knappen, L. B., & Porter, A. C. (1982). A closer look at standardized tests. *Arithmetic Teacher*, 29(7), 50-54.
- Frisbie, D. A. (1992). The multiple true-false format: A status review. *Educational Measurement: Issues and Practice*, 11(4), 21-26.
- Frisbie, D. A., & Becker, D. F. (1990). An analysis of textbook advice about true-false tests. *Applied Measurement in Education*, 4, 67-83.
- Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practice*, 11(3), 35-42.
- Fuchs, L. S., & Fuchs, D. (2007). *Progress monitoring in the context of responsiveness-to-intervention*. Retrieved January 15, 2009, from http://www.studentprogress.org/summer_institute/2007/RTI/RTIManual_2007.pdf
- Gagné, R. M. (1962). The acquisition of knowledge. *Psychological Review*, 69, 355-365.
- Gagné, R. M. (1970). Instructional variables and learning outcomes. In M. C. Wittrock & F. Wiley (Eds.), *Evaluation of instruction*. New York: Holt, Rinehart & Winston.
- Gagné, R. M., & Briggs, L. J. (1979). *Principles of instructional design* (2nd ed.). New York: Holt, Rinehart & Winston.
- Gagné, R. M., Briggs, L. J., & Wager, W. W. (1988). *Principles of instructional design* (3rd ed.). New York: Holt, Rinehart & Winston.
- Gagné, R. M., Major, J. R., Garstens, H. L., & Paradise, N. E. (1962). Factors in acquiring knowledge of a mathematical task. *Psychological Monographs*, 76(7, whole No. 526).
- Gagné, R. M., & Paradise, N. E. (1961). Abilities and learning sets in knowledge acquisition. *Psychological Monographs*, 75(14, whole No. 518).
- Geiger, M. A. (1997). An examination of the relationships between answer changing, testwiseness, and examination performance. *Journal of Experimental Education*, 66, 49-60. *GI Forum, et al. v. Texas Education Agency, et al.* (January 7, 2000). U.S. District Court (Civil Action SA - 97-CA-1278-EP).
- Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist*, 21, 99-120.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 18, 519-521.
- Glaser, R. (1968). Adapting the elementary school curriculum to individual performances. *Proceedings of the 1967 Invitational Conference on Testing Problems*, pp. 3-36. Princeton, NJ: Educational Testing Service.
- Glaser, R. (1977). *Adaptive education: Individual diversity and learning*. New York: Holt, Rinehart & Winston.
- Glaser, R., & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 625-670). Washington, DC: American Council on Education.
- Goldman, B. A., & Mitchell, D. F. (Eds.). (2002). *Directory of unpublished experimental measures, Volume 8*. Washington, DC: American Psychological Association.
- Gong, B. (2008, February). *Developing better learning progressions: Some issues and suggestions for research and policy*. Dover, NH: Center for Assessment.

REFERENCES

- Gong, B., Venezky, R., & Mioduser, D. (1992). Instructional assessments: Level for systemic change in science education classrooms. *Journal of Science Education and Technology, 1*, 157-175.
- Good, T. L., & Brophy, J. E. (2002). *Looking in classrooms* (9th ed.). Boston: Allyn & Bacon.
- Green, K. (1984). Effects of item characteristics on multiple-choice item difficulty. *Educational and Psychological Measurement, 44*, 551-561.
- Gronlund, N. E., & Brookhart, S. M. (2009). *Gronlund's writing instructional objectives* (8th ed.). Upper Saddle River, NJ: Pearson.
- Gulliksen, H. (1986). Perspective on educational measurement. *Applied Psychological Measurement, 10*, 109-132.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: Praeger.
- Haertel, E., & Calfee, R. (1983). School achievement: Thinking about what to test. *Journal of Educational Measurement, 20*, 119-131.
- Haladyna, T. M., & Downing, S. M. (1989a). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 2*, 37-50.
- Haladyna, T. M., & Downing, S. M. (1989b). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 2*, 51-78.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education, 15*, 309-334.
- Haladyna, T. M., Nolen, S. B., & Haas, N. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20*(5), 2-7.
- Hambleton, R. K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education, 5*, 1-16.
- Harrow, A. J. (1972). *A taxonomy of the psychomotor domain: A guide for developing behavioral objectives*. White Plains, NY: Longman.
- Hawkes, H. E., Lindquist, E. F., & Mann, C. R. (Eds.) (1936). *The construction and use of achievement examinations: A manual for secondary school teachers*. Boston: Houghton Mifflin.
- Hembree, R. (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research, 58*, 47-77.
- Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. Washington, DC: Council of Chief State School Officers.
- Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Herman, W. E. (1990). Fear of failure as a distinctive personality trait measure of test anxiety. *Journal of Research and Development in Education, 23*, 180-185.
- Herndon, E. B. (1980). *Your child and testing*. Washington, DC: National Institute of Education, U.S. Department of Education.
- Hess, K. (2007). *Developing and using learning progressions as a schema for measuring progress*. Dover, NH: National Center for Assessment.
- Higgins, K. M., Harris, N. A., & Kuehn, L. L. (1994). Placing assessment into the hands of young children: A study of self-generated criteria and self-assessment. *Educational Assessment, 2*, 309-324.
- Hoover, H. D. (1984a). The most appropriate scores for measuring educational development in the elementary schools: GE. *Educational Measurement: Issues and Practice, 3*(4), 8-14.
- Hoover, H. D. (1984b). Rejoinder to Burket. *Educational Measurement: Issues and Practice, 3*(4), 16-18.
- Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1993). *Directions for administration: Iowa Tests of Basic Skills, Levels 9-14, Forms K and L*. Chicago: Riverside.
- Hough, H. (2005). *Tests and measures in social science: Tests available in compilation volumes*. University of Texas at Arlington, Central Library, Health Sciences. Retrieved from <http://libraries.uta.edu/helen/test&meas/testmainframe.htm>
- Huynh, H. (1976). On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement, 13*, 253-264.
- Ismail, M. (1994). *Development and validation of a multicomponent diagnostic test of arithmetic word problem solving ability for sixth-grade students in Malaysia*. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.
- Iwanicki, E. F. (1980). A new generation of standardized achievement test batteries: A profile of their major features. *Journal of Educational Measurement, 17*, 155-162.
- Joint Advisory Committee. (1993). *Principles for fair student assessment practices for education in Canada*. Edmonton, Alberta: Author, Centre for Research in Applied Measurement and Evaluation, University of Alberta.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. Thousand Oaks, CA: Corwin Press. Available from <http://www.wmich.edu/evalctr/jc>
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage. Available from <http://www.wmich.edu/evalctr/jc>
- Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards: How to improve evaluations of students*. Thousand Oaks, CA: Corwin Press.
- Joint Committee on Testing Practices. (1999). *Rights and responsibilities of test takers: Guidelines and expectations*. Retrieved January 23, 2006, from <http://www.apa.org/science/ttrr.html>
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education (revised)*. Washington, DC: Science Directorate, American Psychological Association. Available from <http://www.apa.org/science/fairtestcode.html>
- Jones, R. W. (1994). *Performance and alternative assessment techniques: Meeting the challenges of alternative evaluation strategies*. Paper presented at the Second International Conference on Educational Evaluation and Assessment, Pretoria, Republic of South Africa.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527-535.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*, 319-342.
- Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21*(1), 31-41.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: Praeger.
- Karweit, N. L., & Wasik, B. A. (1992). *A review of the effects of extra-year kindergarten programs and transitional first grades* (CDS Report 41). Baltimore: Center for Research on

REFERENCES

- Effective Schooling for Disadvantaged Students, Johns Hopkins University.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II Assessment*. San Francisco: Jossey-Bass.
- Keller, F. S. (1968). Goodbye teacher. *Journal of Applied Behavior Analysis, 1*, 79–89.
- Keller, F. S., & Sherman, J. G. (1974). *PSI: The Keller Plan handbook*. Menlo Park, CA: Benjamin-Cummings.
- Kelly, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17–24.
- Kentucky Department of Education. (1993a). *KIRIS Assessment Portfolio: Mathematics, Grade 4, 1993–1994*. Frankfort: Office of Assessment and Accountability.
- Kentucky Department of Education. (1993b). *Portfolios and you*. Frankfort: Office of Assessment and Accountability.
- Kentucky Department of Education. (1993c). *Teacher's guide: Kentucky Mathematics Portfolio*. Frankfort: Office of Assessment and Accountability.
- Keyser, D. J., & Sweetland, R. C. (Eds.) (2005). *Test critiques: Volume 11*. Austin, TX: PRO-ED.
- Khattari, N., Reeve, A. L., & Adamson, R. J. (1997). *Studies of education reform: Assessment of student performance*. Washington, DC: Office of Educational Research and Improvement.
- King, K. V., Gardner, D. A., Zucker, S., & Jorgensen, M. A. (2004, July). *The distractor rationale taxonomy: Enhancing multiple-choice items in reading and mathematics* (Pearson Assessment Report). Retrieved January 21, 2009, from http://pearsonassess.com/NR/rdonlyres/D7E62EC6-CC3F-47B6-B1CB-A83F341AD768/0/Distractor_Rationales.pdf
- Klopfer, L. E. (1969). *An operational definition of "understand"*. Unpublished manuscript, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA.
- Klopfer, L. E. (1971). Evaluation of learning in science. In B. S. Bloom, J. T. Hastings, & G. F. Madaus (Eds.), *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Klotz, M. B. & Canter, A. (2006). *Response to Intervention (RTI): A primer for parents*. Retrieved February 5, 2009, from <http://www.nasponline.org/resources/factsheets/rtiprimer.aspx>
- Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531–578). Westport, CT: Praeger.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont Portfolio Assessment Program: Findings and implications. *Educational Measurement: Issues and Practice, 13*(3), 5–16.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: Book 2. Affective domain*. White Plains, NY: Longman.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, 151–160.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*(2), 16–25, 46.
- La Marca, P. M., Redfield, D., Winter, P. C., Bailey, A., & Despriet, L. H. (2000). *State standards and state assessment systems: A guide to alignment*. Washington, DC: Council of State School Officers.
- Landau, D., & Lazarsfeld, P. F. (1968). Quetelet, Adolphe. *International encyclopedia of the social sciences, 13*, 247–257.
- Lane, S. (1992). The conceptual framework for the development of a mathematics performance assessment instrument. *Educational Measurement: Issues and Practice, 12*(2), 16–23.
- Lane, S., Parke, C., & Moskal, B. (1992). *Principles for developing performance assessments*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice, 21*(1), 23–30.
- Langenfeld, T. E., & Crocker, L. M. (1994). The evaluation of validity theory: Public school testing, the courts, and incompatible interpretations. *Educational Assessment, 2*, 149–165.
- Lindvall, C. M. (1976). Criteria for stating IPI objectives. In D. T. Gow (Ed.), *Design and development of curricular materials: Instructional design articles* (Vol. 2). Pittsburgh, PA: University Center for International Studies, University of Pittsburgh.
- Lindvall, C. M., & Bolvin, J. O. (1967). Programmed instruction in the schools: An application of programming principles in "Individually Prescribed Instruction." In P. Lange (Ed.), *Programmed instruction: 66th yearbook, Part II* (pp. 217–254). Chicago: National Society for the Study of Education.
- Lindvall, C. M., & Nitko, A. J. (1975). *Measuring pupil achievement and aptitude* (2nd ed.). New York: Harcourt Brace Jovanovich.
- Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis, 15*, 1–16.
- Linn, R. L. (1994). Performance assessment: Policy, promises, and technical measurement standards. *Educational Researcher, 23*(4), 4–14.
- Linn, R. L., & Baker, E. (1997, Summer). CRESST conceptual model for assessment. *Evaluation Comment, 7*(1), 1–22.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 5–21.
- Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and Practice, 9*(3), 5–14.
- Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Lord, F. M. (1953). The relation of test scores to the trait underlying the test. *Educational and Psychological Measurement, 13*, 517–549.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maddox, T. (Ed.). (2008). *Tests: A comprehensive reference for assessments in psychology, education, and business* (6th ed.). Austin, TX: PRO-ED.
- Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology, 47*, 166–173.
- Marshall, J. C. (1967). Composition errors and essay examinations grades reexamined. *American Educational Research Journal, 4*, 375–385.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the Dimensions*

REFERENCES

- of Learning Model. Alexandria, VA: Association for Supervision and Curriculum Development.
- Matter, K. K. (1989). Putting test scores in perspective: Communicating a complete report card for your schools. In L. M. Rudner, J. C. Conoley, & B. S. Plake (Eds.), *Understanding achievement tests: A guide for school administrators* (pp. 121–129). Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation.
- Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, NJ: Erlbaum.
- McDonnell, L. (1997). *The politics of state testing: Implementing new student assessments*. CSE Technical Report 424. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, UCLA.
- McKeachie, W. J., Pollie, D., & Spiesman, J. (1985). Relieving anxiety in classroom examinations. *Journal of Abnormal and Social Psychology*, 50, 93–98.
- Mealey, D. L., & Host, T. R. (1992). Coping with test anxiety. *College Teaching*, 40, 147–150.
- Mehrens, W. A. (1991). Facts about samples, fantasies about domains. *Educational Measurement: Issues and Practice*, 10(2), 23–25.
- Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice*, 8(1), 14–22.
- Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- Messick, S. (1989b). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Upper Saddle River, NJ: Prentice Hall.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Miller, M. D., & Seraphine, A. E. (1993). Can test scores remain authentic when teaching to the test? *Educational Assessment*, 1, 119–129.
- Millman, J., Bishop, C. H., & Ebel, R. L. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement*, 25, 707–726.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62, 229–258.
- Moss, P. A., Beck, J. S., Ebbs, C., Matson, B., Muchmore, J., Steele, D., Taylor, C., & Herter, R. (1992). Portfolios, accountability, and an interpretive approach to validity. *Educational Measurement: Issues and Practice*, 11(3), 12–21.
- Murphy, L. L., Spies, R. A., & Plake, B. S. (2006). *Tests in print VII*. Lincoln: University of Nebraska Press.
- National Center on Educational Outcomes. (2009). *Accommodations for students with disabilities*. Retrieved from <http://education.umn.edu/NCEO>
- National Council on Measurement in Education (NCME). (1995). *Code of professional responsibilities in educational measurement (CPR)*. Retrieved from <http://www.unl.edu/buros/bimm/html/article2.html>
- Naveh-Benjamin, M., McKeachie, W. J., & Lin, Y. G. (1987). Two types of test-anxious students: Support for an information processing model. *Journal of Educational Psychology*, 79, 131–136.
- Newman, R. (1997–1998). Parent conferences: A conversation between you and your child's teacher. *Childhood Education*, 74, 100–101.
- Nichols, S. L., & Berliner, D. C. (2008). Why has high-stakes testing so easily slipped into contemporary American life? *Phi Delta Kappan*, 89(9), 672–676.
- Nitko, A. J. (1989). Designing tests that are integrated with instruction. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 447–474). New York: Macmillan.
- Nitko, A. J. (1995). Curriculum-based continuous assessment: A framework for concepts, politics, and procedures. *Assessment in Education: Principles, Policy, and Practice*, 2, 321–337.
- Nitko, A. J. (2005a). *Using an MMY review and other test materials to evaluate a test*. Retrieved from <http://www.unl.edu/buros/bimm/html/lesson02.html>
- Nitko, A. J. (2005b). *Using an MMY review to evaluate a test*. Retrieved from <http://www.unl.edu/buros/bimm/html/lesson01.html>
- Nitko, A. J., Al-Sarimi, A., Amedahe, F. K., Wang, S., & Wingert, M. (1998, April). *How well are the Kentucky Academic Expectations matched to the KIRIS assessments, the CTBS, and the CAT?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED420677)
- Nitko, A. J., & Hsu, T.-C. (1974). Using domain-referenced tests for student placement, diagnosis, and attainment in a system of adaptive, individualized instruction. In W. Hively (Ed.), *Domain-referenced testing*. Upper Saddle River, NJ: Educational Technology Publications.
- Nitko, A. J., & Hsu, T.-C. (1987). *Teacher's guide to better classroom testing: A judgmental approach*. Pittsburgh, PA: Institute for Practice and Research in Education, School of Education, University of Pittsburgh.
- Niyogi, N. S. (1995). *Capturing the power of classroom assessment* (Focus 28). Princeton, NJ: Educational Testing Service.
- No Child Left Behind Act of 2001. Pub. L. No. 107–110, 115 Stat. 1425 (2002).
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications, Critical Thinking Press.
- Northwest Regional Educational Laboratory. (1998). *Improving classroom assessment: A toolkit for professional developers* (2nd ed.). Portland, OR: Author.
- Parkes, J., & Giron, T. (2006). *Making reliability arguments in classrooms*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.
- Parkes, J., & Stevens, J. J. (2002, April). Could school accountability systems be challenged in court? *Educational Assessment Insider: Testing and Accountability*, 1(1), 1–2.
- Pearson. (1992). *Integrated Assessment System: Science Performance Assessment*. San Antonio, TX: Author.
- Pearson. (2003). *Otis-Lennon School Ability Test (8th ed.) Levels E/F/G. Directions for administering*. San Antonio, TX: Author.
- Pearson. (2007). *Stanford Achievement Test Series (10th ed.): 2007 fall supplemental multilevel norms book*. San Antonio, TX: Author.
- Pearson. (2008). *WISC-IV and WIAT-II test scores. Report to parents/guardians*.
- Pearson. (2009). *Otis-Lennon School Ability Test (8th ed.) Assessing the Abilities That Relate to Success in School*. Retrieved

REFERENCES

- April 10, 2009, from http://pearsonassess.com/hai/images/dotCom/olsat8/OLSAT_Brochure.pdf
- Pearson Education. (1991). *The Differential Aptitude Tests* (5th ed., Fall norms booklet). San Antonio, TX: Author.
- Pearson, K. (1924). Historical note on the origin of the normal curve of errors. *Biometrika*, 16, 402–404.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know*. Washington, DC: National Academy Press.
- Perie, M., Marion, S., & Gong, B. (2007). *A Framework for Considering Interim Assessments*. National Center for the Improvement of Educational Assessment. Dover, NH: NCIEA. Available at <http://www.nciea.org>
- Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 18, 16–25.
- Perl, J. (1995). *Improving relationship skills for parent conferences*. *Teaching Exceptional Children*, 28(1), 29–31.
- Peterson, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221–262). Phoenix, AZ: Oryx Press.
- Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education*, 7, 93–120.
- Phillips, S. E. (2005, June). Legal corner: Reconciling IDEA and NCLB. *NCME Newsletter*, 13(2). Retrieved from http://www.ncme.org/pubs/vol13_2_June2005.pdf
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–12, 39.
- Popham, W. J. (1991). Appropriateness of teacher's test-preparation practices. *Educational Measurement: Issues and Practice*, 10(4), 12–15.
- Popham, W. J. (2005, March). Wyoming's instructionally supportive NCLB Tests. *NCME Newsletter*, 13(1). Retrieved from http://www.ncme.org/pubs/vol13_1_Mar2005.pdf
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Porter, A. C., & Smithson, J. L. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states—One hundredth yearbook of the National Society for the Study of Education, Part II* (pp. 60–80). Chicago: University of Chicago Press.
- Power, B. M., & Chandler, K. (1998). *Well-chosen words*. York, ME: Stenhouse.
- Purves, A. C. (1971). Evaluation of learning in literature. In B. S. Bloom, J. T. Hastings, & G. F. Madaus (Eds.), *Handbook on formative and summative evaluation of student learning* (pp. 697–766). New York: McGraw-Hill.
- Quellmalz, E. S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education*, 4, 319–331.
- Reschly, D. J. (1993). Consequences and incentives: Implications for inclusion/exclusion decisions regarding students with disabilities in state and national assessment programs. In J. E. Ysseldyke & M. L. Thurlow (Eds.), *Views on inclusion and testing accommodations for students with disabilities* (pp. 35–46). Minneapolis: National Center on Educational Outcomes, University of Minnesota.
- Resnick, L. B. (1989). *Tests as standards of achievement in schools*. Paper presented at the ETS Invitational Conference: The Uses of Standardized Tests in American Education, New York.
- Robertson, G. J. (1990). A practical model for test development. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 62–85). New York: Guilford Press.
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13.
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (2002). Influences on student cognitions about evaluation. *Assessment in Education*, 9, 81–95.
- Royer, J. M., Cisero, C. A., & Carlo, M. S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research*, 63, 201–243.
- Rozeboom, W. W. (1966). Scaling theory and the nature of measurement. *Synthese*, 16, 170–233.
- Rudner, L. M., & Boston, C. (1994). Performance assessment. *The ERIC Review*, 3(1), 2–12.
- Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split halves. *Harvard Educational Review*, 9, 99–103.
- Ryan, R. M., Connell, J. P., & Deci, E. L. (1985). A motivational analysis of self-determination and self-regulation in the classroom. In C. Ames and R. Ames (Eds.) *Research on motivation in education: Vol. 2. The classroom milieu* (p.13-51). Orlando, FL: Academic.
- Sadler, D. R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, 54, 60-79.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11, 1–31.
- Salend, S. (2009). Using technology to create and administer accessible tests. *TEACHING Exceptional Children*, 41(3), 40–51.
- Salvia, J., & Ysseldyke, J. E. (2004). *Assessment in special and inclusive education* (9th ed.). Boston: Houghton Mifflin.
- Sanders, N. M. (1966). *Classroom questions: What kinds?* New York: Harper & Row.
- Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology*, 46, 929–938.
- Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. *Review of Educational Research*, 49, 252–279.
- Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Author.
- Sattler, J. M. (1992). *Assessment of children: WSC-III and WPPSI-R supplement*. San Diego, CA: Author.
- Schmeiser, C. B. (1992). Ethical codes in the professions. *Educational Measurement: Issues and Practice*, 11(4), 5–11.
- Schutz, P. A., Distefano, C., Benson, J., & Davis, H. A. (2004). The Emotional Regulation During Test-taking scale. *Anxiety, Stress, and Coping*, 17, 253–269.
- Scriven, M. (1967). *The methodology of evaluation*. AERA monograph series on curriculum evaluation (Publication No. 1). Chicago: Rand McNally.

REFERENCES

- Shalaway, L. (1998). *Learning to teach . . . Not just for beginners*. New York: Scholastic Professional Books.
- Shavelson, R. J., & Baxter, G. P. (1991). Performance assessment in science. *Applied Measurement in Education*, 4, 347-362.
- Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51, 455-498.
- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9(3), 15-22.
- Shepard, L. A. (1991). Interview on assessment issues with Lorrie Shepard. *Educational Researcher*, 20(3), 21-23.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.
- Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 623-646). Westport, CT: Praeger.
- Shuell, T. J. (1990). Phrases of meaning learning. *Review of Educational Psychology*, 60, 531-548.
- Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence*, 4, 181-201.
- Sireci, S. G. (2005). Unlabeling the disabled: A perspective on flagging scores from accommodated test administrations. *Educational Researcher*, 34(1), 3-12.
- Slavin, R. E. (1988). Cooperative learning and student achievement. *Educational Leadership*, 46(2), 31-33.
- Smith, J., & Walker, J. (2002). Using electronic gradebooks. *Principal*, 82(2), 64-65.
- Smith, M. L. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona Student Assessment Program (ASAP): Equity and teacher capacity building* (CSE Technical Report 425). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, UCLA.
- Snow, R. E. (1980). Aptitudes and achievement. In W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade. Proceedings of the 1979 ETS Invitational Conference. New Directions for Testing and Measurement* (No. 5). San Francisco: Jossey-Bass.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D., (2005). *Vineland Adaptive Behavior Scales: Second Edition*. Circle Pines, MN: American Guidance Service.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.
- Spearman, C. E. (1927). *The abilities of man, their nature and measurement*. New York: Macmillan.
- Stiggins, R. J., Conklin, N. F., & Associates. (1992). *In teachers' hands: Investigating the practice of classroom assessment*. Albany: SUNY Press.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14.
- Stiggins, R. J., Rubel, E., & Quellmalz, E. (1986). *Measuring thinking skills in the classroom*. Washington, DC: National Education Association.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Student perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30, 325-341.
- Subkoviak, M. J. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement*, 13, 265-276.
- Subkoviak, M. J. (1980). Decision consistency approaches. In R. A. Berk (Ed.), *Criterion-referenced measurement: The state of the art*. Baltimore: Johns Hopkins University Press.
- Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement*, 11, 263-267.
- Swiderek, B. (1997). Parent conferences. *Journal of Adolescent and Adult Literacy*, 40, 580-581.
- Tallmadge, G. K., & Wood, C. T. (1976). *User's guide* (ESEA Title I Evaluation and Reporting System). Mountain View, CA: RMC Research Corporation.
- Taylor, C. S. (1998). An investigation of scoring methods for mathematics performance-based assessments. *Educational Assessment*, 5, 195-224.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved January 9, 2009, from: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>
- Thorndike, E. L. (1910). Handwriting. *Teachers College Record*, 11, 1-93.
- Thorndike, R. L. (1951). Reliability. In E. F. Lindquist (Ed.), *Educational measurement*. Washington, DC: American Council on Education.
- Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education* (5th ed.). New York: Macmillan.
- Tippets, E., & Benson, J. (1989). The effect of item arrangement on test anxiety. *Applied Measurement in Education*, 2, 289-296.
- Tittle, C. K. (1989). Validity: Whose construct is it in the teaching and learning context? *Educational Measurement: Issues and Practice*, 8(1), 5-13, 34.
- Tittle, C. K., Hecht, D., & Moore, P. (1993). Assessment theory and research for classrooms: From taxonomies to constructing meaning in context. *Educational Measurement: Issues and Practice*, 12(4), 13-19.
- Touchstone Applied Science Associates. (1995a). *Degrees of reading power and degrees of word meaning: An overview*. Brewster, NY: Author.
- Touchstone Applied Science Associates. (1995b). *DRP catalog*. Brewster, NY: Author.
- Turner, J. C., Thorpe, P. K., & Meyer, D. K. (1998). Students' reports of motivation and negative affect: A theoretical and empirical analysis. *Journal of Educational Psychology*, 90, 758-771.
- United States Supreme Court. (1971). *Griggs et al., Petitioners v. Duke Power Company* (Publication No. 125, 401 U.S. 424, decided March 8, 1971).
- U. S. Department of Education. (2005, May 10). *New flexibility for states raising achievement for students with disabilities*. No Child Left Behind. Retrieved from <http://www.ed.gov/policy/elsec/guid/raising/disab-factsheet.pdf>
- U. S. Department of Education. (n.d.). *Accountability*. Retrieved from <http://www.ed.gov/nclb/accountability/index.html>
- U. S. Department of Health and Human Services (USDHHS). (2005, June). *Protection of Human Subjects* (CFR 45, Part 46). Retrieved from <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm>

REFERENCES

- Valencia, S. W., & Place, N. A. (1994). Literacy portfolios for teaching, learning, and accountability: The Bellevue Literacy Assessment Project. In S. W. Valencia, E. H. Hiebert, & P. P. Afferbach (Eds.), *Authentic reading assessment: Practices and possibilities*. Newark, DE: International Reading Association.
- Viadero, D. (1995). New assessments have little effect on contract, study finds. *Education Week*, 14(40), 6.
- Wainer, H., & Thissen, D. (1994). On examinee choice in educational testing. *Review of Educational Research*, 64, 159–195.
- Waltman, K. K., & Frisbie, D. A. (1994). Parents understanding of their children's report cards. *Applied Measurement in Education*, 2, 223–240.
- Wang, X., Wainer, H., & Thissen, D. (1995). On the viability of some untestable assumptions in equating exams that allow examinee choice. *Applied Measurement in Education*, 8, 211–225.
- Waples, D., & Tyler, R. W. (1930). *Research methods and teacher problems*. New York: Macmillan.
- Webb, N. L. (1999). *Summary report: Alignment analyses of standards and assessments for four states in science and mathematics* (Council of Chief State School Officers and National Institute for Science Education). Madison: Wisconsin Center for Education Research, University of Wisconsin–Madison.
- Wesman, A. G. (1971). Writing the test item. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- Whitney, D. R., & Sabers, D. L. (1970, May). *Improving essay examinations III: Use of item analysis* (Technical Bulletin No. 11). Iowa City: University Evaluation and Examination Service, University of Iowa.
- Wiggins, G. (1990). *The case for authentic assessment* (EDD-TM-9010). Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation, American Institutes for Research.
- Wiliam, D. (2007). *Content then process: Teacher learning communities in the service of formative assessment*. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 183–204). Bloomington, IN: Solution Tree.
- Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.
- Wilson, M., & Draney, K. (2004). Some links between large-scale and classroom assessments: The case of the BEAR assessment system. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability* (pp. 132–154). Chicago, IL: 103rd Yearbook of the National Society for the Study of Education, Volume II.
- Wise, S. L. (1996, April). *A critical analysis of the arguments for and against item review in computerized adaptive testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.
- Wiser, B., & Lenke, J. M. (1987, April). *The stability of achievement test norms over time*. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.
- Woolfolk, A. (2005). *Educational Psychology* (9th ed.). Boston: Allyn & Bacon.
- Wormeli, R. (2006). *Fair isn't always equal: Assessing and grading in the differentiated classroom*. Portland, ME: Stenhouse, and Westerville, OH: National Middle School Association.
- Young, M. J., & Zucker, S. (2004). *The standards-referenced interpretive framework: Using assessments for multiple purposes* (Pearson Assessment Report). San Antonio, TX: Pearson. Available from <http://pearsonassess.com/NR/rdonlyres/1AD36406-3B2A-491C-A280-D4A47D3121E8/0/InterpretiveFrameworks.pdf>
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum Press.
- Zucker, S., Sassman, C., & Case, B. J. (2004). *Cognitive labs* (Pearson Technical Report). San Antonio, TX: Pearson. Available from http://pearsonassess.com/NR/rdonlyres/E5CD33E6-D234-46F3-885A-9358575372FB/0/CognitiveLabs_Final.pdf