

## References

- ACT Board of Senior Secondary Studies. (2009). *Policy and Procedures Manual*. Australian Capital Territory, Canberra. Retrieved from [http://www.bsss.act.edu.au/publications/policies\\_and\\_procedures](http://www.bsss.act.edu.au/publications/policies_and_procedures)
- Ainley, J., Fleming, M., & McGregor, M. (2002). *Three years on- literacy advance in the early and middle primary years*. Melbourne, Australia: Catholic Education Commission of Victoria.
- Alagumalai, S., Keeves, J.P., & Hungi, N. (1996). *Disattenuated Correlation and Unidimensionality*. Paper presented to Australian Association for Research in Education. Retrieved from <http://www.aare.edu.au/96pap/alags96453.txt>
- Alexander, K., Entwisle, D., & Olson, L. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, 23(2), 171-191.
- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.
- Altman, D.G. (1991). *Practical statistics for medical research*. London, England: Chapman and Hall.
- Andersen, E.B., & Olsen, L.W. (2001). The life of Georg Rasch. In A. Boomsma, M. van Duijn, T. Snijders (Eds.). *Essays on Item Response Theory*. New York, NY: Springer.
- Arnold, M. (1889). *Reports on Elementary Schools 1852-1882*. London, England: MacMillan.
- Assessment and Evaluation Unit, University of Leeds. (2004). *Evaluation of the trial assessment arrangements for key stage 1: Report to QCA*. The University of Leeds Assessment and Evaluation Unit. Retrieved from [http://www.qca.org.uk/qca\\_10082.aspx](http://www.qca.org.uk/qca_10082.aspx).
- Assessment Reform Group-ARG. (2006). *The role of teachers in the assessment of learning*. Retrieved from <http://k1.ioe.ac.uk/tlrp/arg/index.html>
- Aunola, K., Leskinen, E., Lerkkanen, M., & Nurmi, J. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, 96, 699-713.
- Ayres, L.P. (1912). *A Scale for Measuring the Quality of Handwriting of School Children*. New York, NY: Russell Sage Foundation.
- Ayres, L.P. (1918). History and present status of educational measurements. *Seventeenth Yearbook of the National Society for the Study of Education*, 17(2) 11-12. Cited in Cadenhead, K., & Robinson, R (1987). Fisher's "Scale-Book": An early attempt at educational measurement. *Educational Measurement: Issues and Practice*, 6(4), 15-17.
- Bangert-Drowns, R.L., Kulik, C.C., Kulik, J.A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238.
- Barnhart, H.X., Haber, M.J. & Lin, L.I. (2007). An overview on assessing agreement with continuous measurements, *Journal of Biopharmaceutical Statistics*, 17, 529-569.
- Bates, C., & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21(2), 177-187.
- Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal Of Economics*, 121, 1437-1472.

- Bennett, R.E., Gottesman, R.L., Rock, D.A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgements of students' academic skill. *Journal of Educational Psychology*, 85, 347–356.
- Biggs, J., & Collis, K. (1982). *Evaluating the Quality of Learning: the SOLO Taxonomy*. New York, NY: Academic Press.
- Black, P. (2003). (with the King's College London Assessment for Learning Group Harrison, C., Lee, C., Marshall, B., Williams, D.). *Formative and summative Assessment: Can They Serve Learning Together?* Paper presented at American Educational Research Association Classroom Assessment SIG. Retrieved from <http://www.kcl.ac.uk/depsta/education/hpages/pblackpubs.html>
- Black, P.J.; Harrison, C.; Lee, C.; Marshall, B., & Wiliam, D (2002). *Working inside the black box: Assessment for learning in the classroom*. London, England: King's College London School of Education.
- Black, P., & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education*, 5, 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139-148.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham, England: Open University Press.
- Block, J.H. (Ed.). (1971). *Mastery learning: Theory and practice*. New York, NY: Holt, Rinehart & Winston.
- Bloom, B.S. (1968). Learning for mastery. *Evaluation Comment*, 1, 2, 1–12. Retrieved from ERIC database. (ED053419)
- Bloom, B.S. (1971). Mastery learning. In J.H. Block (Ed.), *Mastery learning: Theory and practice* (p. 47–63). New York, NY: Holt, Rinehart & Winston.
- Bond, T.G., & Bond, M.L. (2003, November). *Measure for measure: Curriculum requirements and children's achievement in music education*. Paper presented at the Annual Conference of the New Zealand and Australian Associations for Research in Education, Auckland, New Zealand.
- Bond, T.G., & Fox, C.M. (2001). Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Erlbaum.
- Bond, T.G., & Fox, C.M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences. (2nd ed.) Mahwah, NJ: Erlbaum.
- Boomer, G. (1990). *Attainment levels: The starting orders*, unpublished internal directive to members of the South Australian Education Department's Directorate of Curriculum cited by Jenkin (1996).
- Boston, K. (1992). Working for the possible. *Curriculum Perspectives*, 12(4), p. 15-16.
- Boston, K. (1993). Introducing profiles. *Curriculum Perspectives*, 13(2), p. 12-13.
- Boston, K. (1994). A perspective on the so-called 'National Curriculum'. *Curriculum Perspectives*, 14(1), p. 43-45.
- Brookhart, S.M. (2004). Classroom assessment: tensions and intersections in theory and practice. *Teachers College Record*, 106, 429-458.
- Brown, C.R., Moor, J.L., Silkstone, B.E., & Botton, C. (1996). The construct validity and context dependency of teacher assessment of practical skills in some pre-university level science examinations. *Assessment in Education*, 3, 377–391.

- Brown, C.R., Moor, J.L., Silkstone, B.E., & Botton, C. (1998). An evaluation of two different methods of assessing independent investigations in an operational pre-university level examination in biology in England. *Studies in Educational Evaluation*, 24, 87–98.
- Burgess, M.A. (1937). The construction of two height charts. *Journal of the American Statistical Association*, 32, 198, 290-310.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1-14.
- Butts, R.F.(1978). *Public Education in the United States: From Revolution to Reform*. New York, NY: Holt, Rinehart and Winston.
- Cadenhead, K, & Robinson, R (1987). Fisher’s “Scale-Book”: An early attempt at educational measurement. *Educational Measurement: Issues and Practice*, 6(4), 15-17.
- Cahan, S., & Davis, D. (1987). A between-grade-levels approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal*, 24, 1-12.
- Caldwell, O.W., & Curtis, S.A. (1925). *Then & now in education 1845:1923: A message of encouragement from the past to the present*. Yonkers-On-Hudson, NY: World Book Company.
- Callingham, R. (2003, November). *Establishing the validity of a performance assessment in numeracy*. Paper presented to the Australian Association for Research in Education Conference, Auckland. Retrieved from <http://www.aare.edu.au/03pap/cal03240.pdf>
- Camilli, G. (1999). Measurement error, multidimensionality, and scale shrinkage: A reply to Yen and Burket. *Journal of Educational Measurement*, 36, 73–78.
- Camilli, G., Yamamoto, K., & Wang, M. (1993). Scale shrinkage in vertical equating. *Applied Psychological Measurement*, 17(4), 379-388.
- Canto, A.I. (2006). Predicting third grade students’ FCAT reading achievement and oral reading fluency using student demographic, academic history, and performance indicators. Ph D dissertation, Florida State University. Retrieved from [Etd.Lib.Fsu.Edu/Theses/Available/Etd-04102006-114229/Unrestricted/Dissertation\\_Angel\\_Canto.pdf](http://Etd.Lib.Fsu.Edu/Theses/Available/Etd-04102006-114229/Unrestricted/Dissertation_Angel_Canto.pdf)
- Care, E., Griffin, P., Thomas, A., Pavlovic, M. (2007). *Early years assessment of English*. Melbourne, Australia: Assessment Research Centre, University of Melbourne.
- Caroline Springs College (2008). *Annual Report to the School Community*. Retrieved from [www.carolinesprings.vic.edu.au](http://www.carolinesprings.vic.edu.au)
- Center for Disease Control (2000). *2 to 20 years: Girls/boys stature weight-for-age percentiles*. Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion. Retrieved from <http://www.cdc.gov/growthchart>
- Chadwick, E.B.(1864). The Museum, a Quarterly Magazine of Education, Literature and Science. Vol. II. Reprinted in the *Journal of the Statistical Society of London*, 27(2) (Jun., 1864), 261-266.
- Clay, M.M. (1972). *Reading: the patterning of complex behaviour*. Auckland, New Zealand: Heinemann Educational Books.
- Code of Conduct for Using Student Achievement Information (1995). *An Appendix to the assessment and reporting for schools policy statement*. Adelaide, Australia: Department of Education and Children’s Services.

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78, 141–146.
- Collins, C. (1994a). Curriculum and pseudo-science: Is the Australian national curriculum project built on credible foundations? *Occasional Paper No. 2*. Canberra: Australian Curriculum Studies Association.
- Collins, C. (1994b). Is the National Curriculum Profiles brief valid? *Curriculum Perspectives*, 14(1), p. 45-48.
- Collins, L.M.(2006). Analysis of longitudinal data: The integration of theoretical models, temporal design, and statistical models. *Annual Review of Psychology*, 57, 505–528.
- Condry, K.F., & Spelke, E.S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology*, 137, 22–38.
- Cooper, H., & Tom, D. (1984). Teacher expectation research: A review with implications for classroom instruction. *The Elementary School Journal*, 85, 76-89.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66, 227-268.
- Courtis, S.A. (1909a). Measurement of growth and efficiency in arithmetic. *The Elementary School Teacher*, 10, 58-74.
- Courtis, S.A. (1909b). Measurement of growth and efficiency in arithmetic (continued). *The Elementary School Teacher*, 10, 177-199.
- Courtis, S.A. (1910). Measurement of growth and efficiency in arithmetic (continued). *The Elementary School Teacher*, 11, 171-185.
- Courtis, S.A. (1911a). Measurement of growth and efficiency in arithmetic (continued). *The Elementary School Teacher*, 11, 360-370.
- Courtis, S.A. (1911b). Measurement of growth and efficiency in arithmetic (continued). *The Elementary School Teacher*, 11, 528-539.
- Courtis, S.A. (1911c). Standard scores in arithmetic. *The Elementary School Teacher*, 12, 127-137.
- Courtis, S.A. (1913). The reliability of single measurements with standard tests. *The Elementary School Teacher*, 13, 326-345.
- Courtis, S.A. (1914). Standard tests in English. *The Elementary School Teacher*, 14, 374-392.
- Courtis, S.A. (1916). Courtis tests in arithmetic: Value to superintendents and teachers. *The Fifteenth Yearbook of the National Society For The Study of Education*. (pp. 91-106). Chicago, IL: University of Chicago Press, .
- Courtis, S.A. (1917). *Courtis standard practice tests in arithmetic: Teacher's manual*. Yonkers-On-Hudson, NY: World Book Company.
- Courtis, S.A. (1919). The uses of the Hillegas scale. *The English Journal*, 8(4), 203-217.
- Courtis, S.A. (1929). Maturation units for the measurement of growth. *School and Society*, 30, 683-690.
- Crawford, C., Dearden, L., & Meghir, C. (2007). When you are born matters: The impact of date of birth on child cognitive outcomes in England. The Institute for Fiscal Studies. Retrieved from [http://www.ifs.org.uk/publications.php?publication\\_id=4073](http://www.ifs.org.uk/publications.php?publication_id=4073)

- Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Cross, A. (1917). Weighing the scales. *The English Journal*, 6(3), 183-191.
- Cudeck, R., & Harring, J.R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology*, 58, 616-637.
- Cumming, J.J., Wyatt-Smith, C., Elkins, J., & Neville, M. (2006). *Teacher Judgment: Building An Evidentiary Base For Quality Literacy And Numeracy*. Griffith University. Retrieved from [www.qsa.qld.edu.au/downloads/publications/research\\_qsa\\_teacher\\_judgment.pdf](http://www.qsa.qld.edu.au/downloads/publications/research_qsa_teacher_judgment.pdf)
- Curriculum Corporation. (1994a). *The statements and profiles for Australian schools*. (16 vols.) Melbourne, Australia: Author.
- Curriculum Corporation. (1994b). *CURASS guidelines papers*, Melbourne, Australia: Author.
- Curriculum Corporation. (1994c). *English –a curriculum profile for Australian schools*. Melbourne, Australia: Author.
- Darwin, C. (1860). A naturalist's voyage round the world: Journal of researches into the natural history and geology of the countries visited during the voyage round the world of H.M.S. Beagle under the command of Captain Fitz Roy, R.N. (1913 edition). London, England: John Murray.
- Daugherty, R. (1997). National curriculum assessment: The experience of England and Wales. *Educational Administration Quarterly*, 33, 198-218.
- De Ayala, R.J. (2008). A commentary on historical perspectives on invariant measurement: Guttman, Rasch and Mokken. *Measurement: Interdisciplinary Research and Perspectives*, 6(3), 209-212.
- Demaray, M.K., & Elliot, S.N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13, 8-24.
- Department for Children, Schools and Families. (2009). *The national strategies: AfL with APP: Developing collaborative school-based approaches: Guidance for senior leaders*. Retrieved from <http://nationalstrategies.standards.dcsf.gov.uk/>.
- Department of Education and Early Childhood Development, Victoria. (2003). *Years Prep - 10 Curriculum and Standards Framework II - Benchmarks, 2002*. Spreadsheet CSF\_BENCHMARKS\_02\_(v1.1)-1.xls. Retrieved from <http://www.education.vic.gov.au/management/schoolimprovement/performance/pubs/publications.htm>
- Department of Education and Early Childhood Development, Victoria. (2006). *Years Prep - 10 Curriculum and Standards Framework II - Benchmarks, 2005*. Spreadsheet CSF\_BENCHMARKS\_05\_(v1.0).xls. Retrieved from <http://www.education.vic.gov.au/management/schoolimprovement/performance/pubs/publications.htm>
- Department of Education and Early Childhood Development, Victoria. (2009a). *English Online Interview School User Guide*. Retrieved from <http://www.education.vic.gov.au/studentlearning/teachingresources/english/englishonline.htm#2>
- Department of Education and Early Childhood Development, Victoria. (2009b). *Mathematics Online Interview School User Guide*. Retrieved from <http://www.eduweb.vic.gov.au/MathematicsOnline/>

- Department of Education, Employment and Training, Victoria. (1996). *Assessment and Reporting Support Materials*. Retrieved August 2004 from <http://www.sofweb.vic.edu.au>
- Department of Education and Training, Victoria. (2002). *Teacher Judgement Benchmarks 2001: Years Prep-10 CSF Benchmarks*. Standards and Accountability Division, Office of School Education, Department of Education and Training. Victoria. Retrieved from <http://www.education.vic.gov.au/management/schoolimprovement/performance/publications.htm>
- Department of Education, Victoria. (1997). *Benchmarks 96: Years Prep–10: curriculum and standards framework*. Retrieved from <http://www.education.vic.gov.au/management/schoolimprovement/performance/publications.htm>
- Department of Education, Victoria. (1998). *Benchmarks 97: Years Prep–10: curriculum and standards framework*. Retrieved from <http://www.education.vic.gov.au/management/schoolimprovement/performance/publications.htm>
- Department of Education, Victoria. (1999). *Years Prep –10 curriculum and standards framework benchmarks 1998*. Office of Review. Retrieved from [www.education.vic.gov.au/management/schoolimprovement/performance/publications.htm](http://www.education.vic.gov.au/management/schoolimprovement/performance/publications.htm)
- Donnelly, K. (2007). Australia's adoption of outcomes based education. *Issues In Educational Research*, 17. Retrieved from <http://www.iier.org.au/iier17/donnelly.html>
- Du Bois, P.H. (1970). *A history of psychological testing*. Boston, MA: Allyn & Bacon.
- Duncan, A (2009). *Teacher preparation: Reforming the uncertain profession*—Remarks of Secretary Arne Duncan at Teachers College, Columbia University. October 22. Retrieved from <http://www.ed.gov/print/news/speeches/2009/10/10222009.html>
- Dunn, G. (2007). Regression models for method comparison data, *Journal of Biopharmaceutical Statistics*, 17, 739-756.
- Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2004). *The student assessment handbook*. London, England: Routledge & Falmer.
- Durant, D. (2003, September). *A comparative analysis of Key Stage tests and teacher assessments*. Paper presented to the British Education Research Association Annual Conference. Retrieved from [www.leeds.ac.uk/educol/documents/00003153.htm](http://www.leeds.ac.uk/educol/documents/00003153.htm)
- Early Childhood Longitudinal Study. (2004). *Data files and electronic code book*. US Dept of Education, Institute of Education Sciences. NCES 2004-089.
- Education Department of South Australia (1992). *Attainment levels*. Adelaide, Australia: Author.
- Education Department of South Australia (1993). *Monitoring Student Achievement Resource Paper 2: Attainment Levels and National Profiles*. Adelaide, Australia: Author.
- Egan, O., & Archer, P. (1985). The accuracy of teachers' ratings of ability: A regression model. *American Educational Research Journal*, 22, 25-34.
- Endler, L.C., & Bond, T.G. (2008). Changing science outcomes: Cognitive acceleration in a US setting. *Research in Science Education*, 38(2), 149-166.
- Engelhard, G. (1984). Thorndike, Thurstone, and Rasch: A comparison of their methods of scaling psychological tests. *Applied Psychological Measurement*, 8, 21-38.
- Engelhard, G. (1991a). Thorndike, Thurstone, and Rasch: A comparison of their approaches to item-invariant measurement. *Journal of Research and Development in Education*, 24(2), 45-60.

- Engelhard, G. (1991b). Thorndike and Wood. *Rasch Measurement Transactions*, 5(2), 146.
- Engelhard, G. (1992a). Historical views of invariance: Evidence from the measurement theories of Thorndike, Thurstone, and Rasch. *Educational and Psychological Measurement*, 52, 275-291.
- Engelhard, G. (1992b). Thorndike's scaling vs. Wood's scoring. *Rasch Measurement Transactions*, 5(4), 182.
- Farley, T. (2009). *Making the grades: My misadventures in the standardized testing industry*. Sausalito, CA: PoliPointPress.
- Feinberg, A.B. & Shapiro, E.S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, 18, 52-65.
- Fisher, G. (1862). *On the numerical mode of estimating educational qualifications, as pursued at the Greenwich Hospital School*. Paper presented to 32<sup>nd</sup> Meeting of the British Association for the Advancement of Science, Section F –Economic Science and Statistics, October 1862, listed in the *Journal of the Statistical Society of London*, 25(4), Republished by Cadenhead & Robinson, 1987.
- Fitz-Gibbon, C.T. (1992). Peer and cross-age tutoring. In M.C. Alkin (Ed.) *Encyclopedia of Educational Research*. (pp. 980-984). New York, NY: Macmillan.
- Forster, M. (2009, August). *Informative assessment: Understanding and guiding learning*. Paper presented to the Australian Council for Educational Research Assessment and Student Learning Conference, Perth.
- Forster, M., & Masters, G. (1996). *Portfolios assessment resource kit (ARK Portfolios)*. Melbourne, Australia: Australian Council for Educational Research.
- 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*. 103<sup>rd</sup> Yearbook of the National Society for the Study of Education. (pp. 51-73). Chicago, IL: University of Chicago Press.
- Frederiksen, J.R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Freund, P.A., Holling, H., & Preckel, F. (2007). A multivariate, multilevel analysis of the relationship between cognitive abilities and scholastic achievement. *Journal of Individual Differences*, 28(4), 188-197.
- Frigo, T. (1997, November). *Using curriculum statements and profiles in South Australian schools*. Paper presented at the AARE Annual Conference, Brisbane. Retrieved from [www.aare.edu.au/97pap/frigt213.htm](http://www.aare.edu.au/97pap/frigt213.htm)
- Frigo, T. (1998). *Refining the curriculum statements and profiles: report of focus group sessions with teachers*. Melbourne, Australia: Australian Council for Educational Research.
- Fuchs, L.S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199-208.
- Fullan, M., Hill, P., & Crevola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin.
- Fuller, M. (2000). Teacher judgment as formative and predictive assessment of student performance on Ohio's fourth and sixth grade proficiency tests. Paper presented to the American Educational Research Association Annual Meeting, April 2000, Retrieved from ERIC database. (ED 441015).
- Garner, R. (2009, April 12). Chanting teachers welcome vote to boycott primary tests. *The Independent*. Retrieved from

<http://www.independent.co.uk/news/education/education-news/chanting-teachers-welcome-vote-to-boycott-primary-tests-1667517.html>

- Garrett, R., & Plitz, W. (1999). A case study of curriculum control: curriculum reform in health and physical education. In B. Johnson, & A. Reid (Eds.). *Contesting the curriculum*. Katoomba, Australia: Social Science Press.
- Gilmore, J., & Vance, M. (2007). Teacher ratings of children's listening difficulties. *Child Language Teaching and Therapy*, 23, 133-156.
- Gipps, C., Brown, M, McCullum, B., & McAlister, S. (1995). *Intuition or evidence? Teachers and national assessments of seven-year-olds*. Buckingham, England: Open University Press.
- Gipps, C.V., Murphy, P. (1994). *A fair test?: Assessment, achievement and equity*. Buckingham, England: Open University Press.
- Gladwell, M. (2005). *Blink: The power of thinking without thinking*. New York, NY: Little, Brown and Co.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society*, 115, 513-583. Retrieved from <http://rstl.royalsocietypublishing.org.elibrary.jcu.edu.au/content/115.toc>
- Good, H.G. (1926). An early school surveyor, *Educational Research Bulletin*, 5, 351-353.
- Gough, J. (2006). VELS versus CSF; replacement, conflict or harmony? *Vinculum*, 43(2), 3-6.
- Green, S.K., & Mantz, M. (2002). *Classroom assessment practices: Examining impact on student learning*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. ED 464 920.
- Griffin, P.E. (1990). Profiling literacy development: Monitoring the accumulation of reading skills. *Australian Journal of Education*, 34, 290-311.
- Griffin, P.E. (1998). Outcomes and profiles: Changes in teachers' assessment practices. *Curriculum Perspectives*, 18(1), 9-19.
- Griffin, P.E. (2004, March). *The comfort of competence and the uncertainty of assessment*. Paper presented at the Hong Kong School Principal's Conference, Hong Kong Institute of Education.
- Griffin, P.E. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation*, 33, 87-99.
- Griffin, P., & Callingham, R. (2006). A twenty-year study of mathematics achievement. *Journal for Research in Mathematics Education*, 37(3), 167-186.
- Grissom, J.B. (2004). Age and achievement. *Education Policy Analysis Archives*, 12(49), 1-40. Retrieved from <http://epaa.asu.edu/ojs/article/view/204>
- Gunther, C. (1919). My experience with the Hillegas scale. *The English Journal*, 8, 535-542.
- Guskey, T.R. (1981). Comparison of a Rasch model scale and the grade-equivalent scale for vertical equating of test scores. *Applied Psychological Measurement*, 5, 187-201.
- Gustafsson, J. (1979). The Rasch Model in Vertical Equating of Tests: A Critique of Slinde and Linn. *Journal of Educational Measurement*, 16, 3, 153-158.
- Haertel, E. (1991). Report on TRP analyses of issues concerning within-age versus cross-age scales for the National Assessment of Educational Progress. Washington, DC: National Center for Education Statistics. Retrieved from ERIC database. (ED404367)



- Haertel, E., & Herman, J. (2005). *A Historical Perspective on Validity Arguments for Accountability Testing*. CSE Report No. 654.. Technical Report for the National Center for Research on Evaluation, Standards, and Student Testing, UCLA.
- Hampton Primary School (2006). *Sample whole year reporting schedule and extracts from whole year assessment calendars*. Retrieved from [www.education.vic.gov.au/studentlearning/studentreports/casestudies/default.htm](http://www.education.vic.gov.au/studentlearning/studentreports/casestudies/default.htm)
- Harlen, W. (1994). Towards quality in assessment, in W. Harlen (Ed.), *Enhancing quality in assessment*. London, England: Paul Chapman.
- Harlen, W. (2004a). A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes. London: EPPI-Centre, Social Science Research Unit, Institute of Education. Retrieved from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=149>
- Harlen, W. (2004b). *A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes*. London: EPPI-Centre, Social Science Research Unit, Institute of Education. Retrieved from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=149>
- Harlen, W. (2005a). Teachers' summative practices and assessment for learning-tensions and synergies. *The Curriculum Journal*, 16, 207-223.
- Harlen, W. (2005b). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20, 245-270.
- Harlen, W. (2007a). Criteria for evaluating systems for student assessment. *Studies In Educational Evaluation*, 33, 15-28.
- Harlen, W. (2007b). The quality of learning: assessment alternatives for primary education. *Primary Review Research Survey 3/4*, Cambridge, England: University of Cambridge Faculty of Education.
- Harlen, W. (2007c). *Assessment of learning*. London, England: Sage.
- Hattie, J. (1999). *Influences on student learning: inaugural professorial address*. August 2, 1999. Auckland, New Zealand: University of Auckland. Retrieved from [www.arts.auckland.ac.nz/edu/staff/jhattie](http://www.arts.auckland.ac.nz/edu/staff/jhattie)
- Hattie, J. (2003, October). *Teachers make a difference: What is the research evidence?* Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Hattie, J.A., & Brown, G.T. (2003). Standard setting for asTTle reading: A comparison of methods. *asTTle Technical Report 21*, University of Auckland/Ministry of Education. Retrieved from [http://www.tki.org.nz/r/asttle/tech-reports\\_e.php](http://www.tki.org.nz/r/asttle/tech-reports_e.php)
- Hauser, C. (2003, April). *So, what d'ya expect? Pursuing reasonable individual student growth targets to improve accountability systems*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago. Retrieved from <http://www.kingsburycenter.org/>
- Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. Paper prepared for the formative assessment for teachers and students. State Collaborative on Assessment and Student Standards of The Council Of Chief State School Officers. Washington DC.

- Herman, J., & Choi, K. (2008). Formative assessment and the improvement of middle school science learning: The role of teacher accuracy. *CRESST Report 740*. Retrieved from [www.cse.ucla.edu/products/reports/R740.pdf](http://www.cse.ucla.edu/products/reports/R740.pdf)
- Hieronymus, A.N. & Hoover, H.D. (1986). *Iowa Tests of Basic Skills manual for school administrators*. Chicago: Riverside.
- Hill, C.J., Bloom, H.S., Rebeck Black, A., & Lipsey, M.W. (2007). Empirical benchmarks for interpreting effect sizes in research. *MDRC Working Papers on Research Methodology*. Retrieved from [www.mdrc.org](http://www.mdrc.org)
- Hill, P.W. (1994). Putting the national profiles to use. *Unicorn*, 20(2), 36-42.
- Hillegas, M. (1912). Scale for the measurement of quality in English composition by young people. *Teachers College Record*, 13(4), 1-55.
- Hilton, T.L. & Patrick, C. (1970). Cross-sectional versus longitudinal data: an empirical comparison of mean differences in academic growth. *Journal of Educational Measurement*, 7, 1,15-24.
- Hinnant, J.B., O'Brien, M., & Ghazarian, S.R. (2009). The longitudinal relations of teacher expectations to achievement in the early years. *Journal of Educational Psychology*, 101, 662-670.
- Hinton, E.M. (1940). An analytical study of the qualities of style and rhetoric found in English compositions. *Teachers College Record*, 42(2), 157-159. Retrieved from <http://www.tcrecord.org> (ID Number: 8975)
- Hoeksma, J.B., & Kelderman, H. (2006). Commentary: On growth curves and mixture models. *Infant and Child Development*, 15, 627-634.
- Hoff, D.J. (2006). Delving into data. *Education Week Technology Counts 2006*. 25(5), 12-14 and 20-22. Retrieved from <http://www.edweek.org/>
- Hoge, R.D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of the literature. *Review of Educational Research*, 59, 297-781.
- Holmes. (1913). From Professor Holmes. *The English Leaflet*, 104, 4-5. Retrieved from <http://www.archive.org/details/hillegasscale00newerich>
- Holmes, S.E. (1982). Unidimensionality and vertical equating with the Rasch model, *Journal of Educational Measurement*, 19, 2, 139-147.
- Hornibrook, M., & Wallace, M. (2001). Report on the Independent Evaluation of the Development of the South Australian Curriculum Standards and Accountability Framework. Melbourne, Australia: Curriculum Corporation. Retrieved from <http://www.sacsa.sa.edu.au>
- Hubbard, D.W. (2007). *How to measure anything: Finding the value of intangibles in business*. New York, NY: John Wiley and Sons.
- Hudelson, E. (1916). Some achievements in the establishment of a standard for the measurement of English composition in the Bloomington, Indiana, schools. *The English Journal*, 5, 590-597.
- Hudelson, E. (1923). The development and comparative values of composition scales. *The English Journal*, 12(3), 163-168.
- Humboldt, A. von, (1811). *Atlas géographique et physique du royaume de la nouvelle-espagne, fonde sur des observations astronomiques, des mesures trigonometriques et des nivellemens barometriques*. No. 29. Paris: F. Schoell. Cited in H.G Funkhouser (1937). Historical development of the graphical representation of statistical data. *Osiris*, 3, 269-404. Saint Catherines Press. Retrieved from <http://www.jstor.org/stable/301591>

- Hungi, N. (2003). *Measuring school effects across grades*. Institute of International Education Research Collection, Number 6. Adelaide, Australia: Flinders University.
- Huxley, J. (Ed.). (1936). *T.H. Huxley's diary of the voyage of H.M.S. Rattlesnake*. Garden City, NY: Doubleday, Doran and Company.
- Hyams, D. (2001). *CurveExpert Version 1.38 A curve fitting system for Windows*. Copyright 1995-2001. Retrieved from <http://curveexpert.webhop.net/>
- Jenkin, R. (1996). *Australian national collaborative curriculum development- a tangled web*. Contribution to M. Dilena and C.E. van Kraayenoord website: Whole school approaches to assessing and reporting literacy. Retrieved 27 September 2004 (link now broken) from <http://www.gu.edu.au/school/cls/clearinghouse/>
- Johanningmeier, E.V. (2004). The transformation of Stuart Appleton Courtis: Test maker and progressive, *American Educational History Journal*, 31(2), 202-210.
- Johanningmeier, E.V. & Richardson, T.R. (2008). *Educational Research, The National Agenda, and Educational Reform: A History*. Charlotte, NC: Information Age Publishing.
- Johnson, B., & Reid, A. (Eds.). (1999). *Contesting the curriculum*. Katoomba, Australia: Social Science Press.
- Johnson, F.W. (1913). The Hillegas-Thorndike scale for measurement of quality in English composition by young people. *The School Review*, 21(1), 39-49.
- Johnson, J.A., Dupuis, V.L., Musial, D., Hall, G.E., & Gollnick, D. (2002). *Introduction to the foundations of American Education* (12th ed.). Boston: Allyn, & Bacon.
- Jorgensen, M.A. (2004). And there is much left to do. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability*. 103<sup>rd</sup> Yearbook of the National Society for the Study of Education. (pp. 203-208). Chicago, IL: University of Chicago Press.
- Jussim, L., & Eccles, J. (1995). Naturalistic studies of interpersonal expectancies. *Review of Personality and Social Psychology*, 15, 74-108.
- Justice, L.M., Pence, K., Bowles, R.B., & Wiggins, A. (2006). An investigation of four hypotheses concerning the order by which 4-year-old children learn the alphabet letters. *Early Childhood Research Quarterly*, 21, 374-389.
- Keats, J.A. (1983). Ability measures and theories of cognitive development. In H. Wainer & S. Messick (Eds.) *Principles of modern psychological measurement: A festschrift for Frederick M. Lord*. Hillsdale, NJ: Erlbaum.
- Keeves, J.P. (Chair). (1982). *Education and change in South Australia: Final Report of the Committee of Enquiry into Education in South Australia*. Adelaide, Australia: Government Printer.
- Keeves, J.P. & Marjoribanks, K. (Eds.). (1999). *Australian education: Review of research 1965-1998*. Melbourne, Australia: Australian Council for Educational Research.
- Kerbow, D., & Bryk, A. (2005). *STEP literacy assessment technical report of validity and reliability*. Center for Urban School Improvement (USI) at the University of Chicago. Retrieved from [http://www.iisrd.org/software\\_step/index.shtml](http://www.iisrd.org/software_step/index.shtml)
- Kingsbury, G.G.(2003). *A Long-Term Study Of Item Parameter Estimates*. Paper presented to the Annual Meeting of the American Educational Research Association, April, Chicago.
- Kissane, B.V. (1982). The Measurement of Change as the Study of the Rate of Change. *Education Research and Perspectives*, 9(1), 55-72.

- Klein, G. (1999). *Sources of power. How people make decisions*. Cambridge, MA: MIT Press.
- Klein, G. (2009). *Streetlights and shadows. Searching for the keys to adaptive decision making*. Cambridge, MA: MIT Press.
- Kolen, M.J., & Brennan, R.L.(2004). *Test equating, scaling, and linking-Methods and practices*. (2nd ed.). New York, NY: Springer-Verlag.
- Laidra, K., Allik, J., Harro, M., Merenäkk, L., & Harro, J. (2006). Agreement among adolescents, parents, and teachers on adolescent personality. *Assessment, 13*(2), 187-196.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- Leahy, S., & Wiliam, D. (2009). *From teachers to schools: scaling up professional development for formative assessment*. Paper presented at American Educational Research Association conference, Retrieved from <http://www.dylanwiliam.net/>
- Learned, W.S. (1913). From Dr. Learned. *The English Leaflet, 104*, 7-8. Retrieved from <http://www.archive.org/details/hillegassscale00newerich>
- Lee, O.K. (1993). Absolute zeroes of reading and mathematics. *Rasch Measurement Transactions, 6*(4), 245-246.
- Lee, O.K. (2003), Rasch simultaneous vertical equating for measuring reading growth. *Journal of Applied Measurement, 4*(1), 10-23.
- Ligon, G.D. (2009). Growth models – finding real gains. Extraordinary insight into today’s education information topics. Retrieved from <http://www.espsolutionsgroup.com>
- Linacre J.M., & Wright B.D (1989). The "length" of a logit. *Rasch Measurement Transactions, 3*(2), 54-55.
- Linacre, J.M. (1998). Wright: The measure of the man. *Popular Measurement, Spring*, 23-25.
- Linacre, J.M. (1999). Explorations into local independence with T-Rasch. *Rasch Measurement Transactions, 13*(3), 710.
- Linacre, J.M. (2006). A User's Guide to WINSTEPS MINISTEP Rasch-model computer programs. Retrieved from [www.winsteps.com](http://www.winsteps.com)
- Lokan, J. (Ed) (1997). *Describing Learning; Implementation of Curriculum Profiles in Australian Schools 1986-1996*. Melbourne, Australia: Australian Council for Educational Research.
- Lokan, J., & Wu, M. (1997). Summary of the 1992-93 ACER calibration studies. In J. Lokan (Ed.). (1997). *Describing Learning; Implementation of Curriculum Profiles in Australian Schools 1986-1996*. (pp. 8-23). Melbourne, Australia: Australian Council for Educational Research.
- Macken, E., Suppes, P., & Zanotti, M. (1980). Considerations in evaluating individualised instruction. *Journal of Research and Development in Education, 14*(1), 79-83.
- Malone, T.W., Suppes, P., Macken, E., Zanotti, M., & Kanerva, L. (1979 ). Projecting student trajectories in a computer-assisted instruction curriculum. *Journal of Educational Psychology, 71*, 74-84.
- Mann, H. (1845). Boston Grammar and Writing Schools. *The Common School Journal*. Reprinted in O.W. Caldwell & S.A. Courtis. (1925). *Then & now in education 1845:1923: A message of encouragement from the past to the present*. (pp. 237-272). Yonkers-On-Hudson, NY: World Book Company.
- Mantel, H. (2005), Is the particle there? Review of *A Game with Sharpened Knives* by Neil Belton. *London Review of Books, 27*(13), 13-14.

- Marist-Sion College. (2008). *School Annual Report*. Retrieved from <http://www.vrqa.vic.gov.au/>.
- Marsh, C.J. (1994). *Producing a national curriculum: Plans and paranoia*. Sydney Australia: Allen and Unwin.
- Marsh, J.A., Pane, J.F., & Hamilton, L.S. (2006). *Making sense of data-driven decision making in education: evidence from recent RAND research*. Retrieved from [http://www.rand.org/pubs/occasional\\_papers/OP170/](http://www.rand.org/pubs/occasional_papers/OP170/)
- Marzano, R. (2000a). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. (2000b). *Designing a new taxonomy of educational objectives*. Experts In Assessment Series. Thousand Oaks, CA: Corwin Press.
- Marzano, R.J. (2007). *The art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Masters, G.N. (1990). *Subject profiles as frameworks for assessing and reporting student achievement*. Unpublished discussion paper, commissioned by the Australasian Cooperative Assessment Program, Australian Council for Educational Research, Melbourne.
- Masters, G.N. (1992). Towards a national framework for assessment and reporting. *Unicorn*, 18(1), 66-77.
- Masters, G.N. (1999). Research in educational measurement. In J. Keeves & K. Majoribanks (Eds.). (1999). *Australian education: Review of research 1965-1998*. Melbourne, Australia: Australian Council for Educational Research.
- Masters, G.N. (2005). Against the grade: In search of continuity in schooling and learning. *Professional Educator*, 4(1), 12-22.
- Masters, G.N., & Forster, M. (1996). *Developmental assessment: Assessment resource kit*. (ARK developmental assessment). Melbourne, Australia: Australian Council for Educational Research.
- Masters, G.N., & Forster, M. (1996). *Progress maps: Assessment resource kit*. Melbourne, Australia: Australian Council for Educational Research.
- Masters, G.N., & Forster, M. (1997). *Mapping literacy achievement, results of the 1996 national school English literacy survey*. Canberra, Australia: Commonwealth of Australia. Retrieved from <http://www.dest.gov.au/mla/mla.pdf>
- Masters, G.N., & Forster, M.(undated). *The assessments we need*. Australian Council for Educational Research website. Retrieved from [www.acer.edu.au/research/documents/Theassessmentsweneed.pdf](http://www.acer.edu.au/research/documents/Theassessmentsweneed.pdf)
- Masters, G.N., & Keeves, J.P. (Eds.). (1999). *Advances in measurement in educational research and assessment*. New York, NY: Pergamon (Elsevier Science).
- Masters, G. N., Lokan, J., Doig, B., Khoo, S.T., Lindsey, J., Robinson, L. & Zammit, S. (1990). *Profiles of Learning: The Basic Skills Testing Program in New South Wales 1989*. Melbourne, Australia: Australian Council for Educational Research. Retrieved from ERIC database. (ED327276)
- Masters, G.N., & Mossenson, L. T. (1983). *Using a Latent Trait Model to Vertically Equate Educational Tests*. Paper presented at Australian Association for Research in Education Conference, Canberra.
- Masters, G.N., Rowley, G., Ainley, & J. Khoo, S. T. (2008). *Reporting and Comparing School Performances*. Paper prepared for the MCEETYA Expert Working. Commissioned by the Reporting and Accountability Branch, National Education

Systems Group, Commonwealth Department of Education, Employment and Workplace Relations (DEEWR).

- Maxwell, G.S. (2004). *Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland*. Paper presented at the third conference of the Association of Commonwealth Examination and Assessment Boards, March, Nadi, Fiji. Retrieved from [http://www.spbea.org.fj/aceab\\_conference.html](http://www.spbea.org.fj/aceab_conference.html)
- Mazyck, M. (2002). Integrated learning systems and students of color: Two decades of use in K-12 education. *TechTrends*, 46(2), 33.
- McCall, M., Hauser, C., Cronin, J., Kingsbury, G., & Houser, R. (2006). Achievement gaps: An examination of differences in student achievement and growth. Technical report from the NWEA Growth Research Database. Retrieved from <http://www.kingsburycenter.org/our-research/research-reports-publications>
- McCall, M.S. (2006). *Item Response Theory And Longitudinal Modeling: The Real World Is Less Complicated Than We Fear*. Presented to the MSDE/MARCES Conference Assessing & Modeling Cognitive Development In School: Intellectual Growth And Standard Setting October. [PowerPoint slides] Retrieved from <http://www.education.umd.edu/EDMS/MARCES/conference/cognitive/MARCESPre sMcCalltiny.ppt>
- McCall, W.A. (1922). *How to Measure in Education*. New York: Macmillan.
- McGaw, B. (1994). Standards from a curriculum and assessment perspective. *Queensland Researcher*, 10(2), 1-18. Retrieved from <http://education.curtin.edu.au/iier/qjer/qr10/mcgaw.html>
- McKinsey & Company. (2007). *How the World's Best-performing School Systems Come Out on Top*. Retrieved from [http://www.mckinsey.com/client-service/social-sector/resources/pdf/Worlds\\_School\\_Systems\\_Final.pdf](http://www.mckinsey.com/client-service/social-sector/resources/pdf/Worlds_School_Systems_Final.pdf)
- Meisels, S., Bickel, D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment from Kindergarten to Grade 3. *American Educational Research Journal*; 38, 73-95.
- Meisels, S., Dorfman, A., & Steele, D. (1994). *Equity and excellence in group-administered and performance-based assessments*. In M. Nettles & A. Nettles (Eds.), *Equity in educational assessment and testing* (pp. 195—211). Boston, MA: Kluwer Academic Publishers.
- Meisels, S.J., Liaw, F., Dorfman, A., & Nelson, R. (1995). The Work Sampling System: reliability and validity of a performance assessment for young children. *Early Childhood Research Quarterly*, 10, 277-296.
- Merrell, C., & Tymms, P. (2007). What children know and can do when they start school and how this varies between countries. *Journal of Early Childhood Research*, 5, 115-134.
- Merton, R. (1948). The self-fulfilling prophecy. *Antioch Review*, 8, 193-210.
- Messick, S. (1993). *Foundations of validity: Meaning and consequences in psychological assessment*. Princeton, NJ: Educational Testing Service.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Miller, K.F., Smith, C. M. Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science*, 6, 56-60.

- Ministry of Education, Québec. (2002). *Evaluation Of Learning At The Preschool And Elementary Levels: Framework*, Gouvernement du Québec, Ministère de l'Éducation.
- Molenaar, P.C.M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2, 201–218.
- Molenaar, P.C.M., & Campbell, C.G. (2009). The new person-specific paradigm in psychology, *Current Directions in Psychological Science*, 18, 112-117.
- Molenaar, P.C.M., Sinclair, K.O., Rovine, M.J., Ram, N., & Corneal, S.E. (2009). Analyzing developmental processes on a individual level using non-stationary time series modeling. *Developmental Psychology*, 45, 260–271.
- National Assessment Program Literacy and Numeracy (NAPLAN). (2008). *Achievement in reading, writing, language conventions and numeracy*. Melbourne, Australia: Ministerial Council on Education, Employment, Training and Youth Affairs.
- National Center for Education Statistics (2005). *A First Look at the Literacy of America's Adults in the 21st Century*. NCES Number: 2006470
- National Curriculum Board. (2009). *The shape of the Australian curriculum: English*. Canberra, Australia: Commonwealth of Australia 2009. Retrieved from <http://www.acara.edu.au>.
- National Curriculum Council (UK). (1991). Report on monitoring the implementation of the national curriculum core subjects 1989-1990. York, England: Author.
- National Report on Schooling in Australia Preliminary Paper. (2007). *National Benchmark Results Reading, Writing and Numeracy Years 3, 5 and 7*. Melbourne, Australia: Ministerial Council on Education, Employment, Training and Youth Affairs.
- Neilson. (1913). From Professor Neilson. *The English Leaflet*, 104, 5. Retrieved from <http://www.archive.org/details/hillegassscale00newerich>
- Northwest Evaluation Association. (2002). *RIT scale norms*. Retrieved from <http://www.nwea.org/>
- Northwest Evaluation Association. (2005). *2005 normative data: Monitoring growth in student achievement*. Retrieved from <http://www.nwea.org/>
- Northwest Evaluation Association. (2005). *Normative data, RIT point monitoring, growth in student achievement*. Retrieved from <http://www.nwea.org/>
- Olsen, L.W. (2003). *Essays on Georg Rasch and his contributions to statistics*. PhD thesis-part, Institute of Economics, University of Copenhagen. Retrieved from <http://www.rasch.org/olsen.htm>
- Ouchi, W.G. (2009). *The Secret of TSL: The Revolutionary Discovery That Raises School Performance*. New York, NY: Simon & Schuster.
- Oxford University Archives (n.d.). Retrieved from <http://www.oua.ox.ac.uk/holdings/Local%20Examinations%20Delegacy%20LE.pdf>
- Parrila, R., Aunola, K., Leskinen, E., Nurmi, J., & Kirby, J.R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology*, 97, 299–319.
- Participation and Achievement Series: Nos 2-5. (1998, 1999). Broadsheet pamphlets reporting profile data for English, science, mathematics, health, physical education. Adelaide, Australia: Department of Education, Training and Employment.
- Patz, R.J. (2007). *Vertical Scaling in Standards-Based Educational Assessment and Accountability Systems*. Prepared for the Technical Issues in Large Scale Assessment

- (TILSA) State Collaborative on Assessment and Student Standards (SCASS) of the Council of Chief State School Officers (CCSSO). CTB/McGraw-Hill.
- Pedulla, J.J., Airasian, & Madaus, G.F. (1980). Do Teacher Ratings and Standardized Test Results of Students Yield the Same Information? *American Educational Research Journal*, 17, 3, 303-307.
- Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Research Council. 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. Retrieved from [http://www.nciea.org/cgi-bin/pubspage.cgi?sortby=pub\\_date](http://www.nciea.org/cgi-bin/pubspage.cgi?sortby=pub_date)
- Perry, N., & Meisels, S. (1996). *How accurate are teacher judgments of students' academic progress?* National Center for Educational Statistics, Working Paper 96-08. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=9608>
- Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). *Scaling, norming, and equating*. In R.L. Linn, *Educational measurement* (3rd ed. pp. 221-262). New York: Macmillan.
- Piper, K.R. (1997). *Riders in the chariot: Curriculum reform and the national interest: 1965-1995*. Melbourne, Australia: Australian Council for Educational Research.
- Pollack, J.M., Atkins-Burnett, S., Najarian, M., & Rock, D.A. (2005). Early Childhood Longitudinal Study, Kindergarten class of 1998–99 (ECLS–K): Psychometric report for the fifth grade. (NCES 2006–036). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006036rev>
- Popham, W.J. (2003). *Crafting curricular aims for instructionally supportive assessment*. Paper prepared for American Association of School Administrators, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the National Education Association, and the National Middle School Association.
- Popham, W.J. (2007). The lowdown on learning progressions, *Educational Leadership*, 64(7), 83-84.
- Popham, W.J., & Husek, T.R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6, 1–9.
- Popham, W.J., Cruse, K.I., Rankin, S.C., Sandifer P.D., & Williams P.L. (1985). Measurement-driven instruction: It's on the road. *Phi Delta Kappan*, 66, 628–634.
- Qualifications and Curriculum Authority. (2008). *Assessment and reporting arrangements 2009*. Retrieved from [http://www.naa.org.uk/naa\\_21551.aspx](http://www.naa.org.uk/naa_21551.aspx)
- Qualifications and Curriculum Authority. (2009a). *Evaluation of the assessing pupils' progress in year 1 pilot project 2007/8 Final report*, September 2008 QCA/09/4016. Retrieved from <http://www.qcda.gov.uk/resources/publication.aspx?id=5079a52a-e09b-4001-be36-19e66fd13075>
- Qualifications and Curriculum Authority. (2009b). *Single level tests: Report of the first three test sessions: December 2007, June 2008 and December 2008*. QCA/09/4303. Retrieved from <http://www.qcda.gov.uk/resources/> [link broken]
- Qualifications and Curriculum Development Agency. (2010). *Single level tests arrangements*. Retrieved from <http://testsandexams.qcda.gov.uk/>
- Quantile Framework for Mathematics. (2010). *Metametrics*. Retrieved from [www.Quantiles.com](http://www.Quantiles.com)



- Quinlan, M., & Scharaschkin, A. (1999, September). National curriculum testing: problems and practicalities. Paper presented at the British Educational Research Association Annual Conference, Brighton. Cited by Tymms (2004) and Durant (2003).
- Raudenbush S.W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, 52, 501–525.
- Rao, C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14, 1-17.
- Rogosa, D.R., and Willett, J.B. (1985). Understanding Correlates Of Change By Modeling Individual Differences In Growth. *Psychometrika*, 50, 203-228.
- Raudenbush, S.W. (2009). The Brown legacy and the O'Connor challenge: Transforming schools in the images of children's potential, *Educational Researcher*, 38, 169-180.
- Ravitch, D. (2010). The death and life of the great American school system: How testing and choice are undermining education. New York, NY: Basic Books.
- Reid, A. (1991). *Attainment levels: A critique*. Paper presented at a seminar, Underdale Campus University, Centre for Studies In Educational Leadership.
- Reid, A. (1992a). Reshaping education: National agenda - national curriculum problems and possibilities, *SAIT Journal*, 24 June, 1992 p.13 and p.22.
- Reid, A. (1992b). Accountability and education: The great profile debate. *Curriculum Perspectives*, 12(1), 55-57.
- Reid, A. (1995). Profiles: Real problems or real gains-from whose perspective? *Curriculum Perspectives*, 15(3), 76-80.
- Reid, A. (1999). The national education agenda and its curriculum effects. In B. Johnson & A. Reid (Eds.). *Contesting the curriculum*. Katoomba, Australia: Social Science Press.
- Reid, A. (2005). Rethinking national curriculum collaboration: Towards an Australian curriculum. Canberra, Australia: DEST.
- Rice, J.M. (1893). *The public-school system of the United States*. New York, NY: The Century Company. Retrieved from [www.archive.org/details/scientificmanag00ricegoog](http://www.archive.org/details/scientificmanag00ricegoog).
- Rice, J.M. (1897). The futility of the spelling grind. *Forum*, 23, 163-172, 409-419. Republished in J.M. Rice (1913). *Scientific management in education*. New York, NY: Publishers Printing Company. Retrieved from [www.archive.org/details/scientificmanag00ricegoog](http://www.archive.org/details/scientificmanag00ricegoog).
- Rice, J.M. (1902). Educational research: a test in arithmetic. *Forum*, 34, 281-297. Republished in J.M. Rice (1913). *Scientific management in education*. New York, NY: Publishers Printing Company. Available at [www.archive.org/details/scientificmanag00ricegoog](http://www.archive.org/details/scientificmanag00ricegoog).
- Rice, J.M. (1913). *Scientific management in education*. New York, NY: Publishers Printing Company. Available at [www.archive.org/details/scientificmanag00ricegoog](http://www.archive.org/details/scientificmanag00ricegoog).
- Rogoff, B. (1990). Apprenticeship in thinking: Cognitive development in social context. New York, NY: Oxford University Press.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom: teacher expectation and pupils' intellectual development. New York, NY: Holt, Rinehart, & Winston.
- Rosenthal, R., & Rubin, D.B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, 3, 377–386.

- Rothman, S. (1998). *Factors influencing assigned student achievement levels*. Paper presented at Australian Association for Research in Education. Retrieved from <http://www.aare.edu.au/98pap/rot98337.htm>
- Rothman, S. (1999). *Factors influencing assigned student achievement levels II: Mathematics, The Arts, and Health and Physical Education*. Paper presented at Australian Association for Research in Education. Retrieved from <http://www.aare.edu.au/99pap/rot99019.htm>
- Rowe, K.J., & Hill, P.W. (1996). Assessing, recording and reporting students' educational progress: The case for 'subject profiles'. *Assessment in Education*, 3, 309–352.
- Rudner, L.M., & Boston, C. (2003). Data warehousing: Beyond disaggregation. *Educational Leadership*, 60(5), 62-65.
- Rupp, A., & Templin, J. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219-262.
- Ryan, J., & Williams, J. (2007). *Children's mathematics 4-15: Learning from errors and misconceptions*. Buckingham, England: Open University Press.
- Sadler, D.R. (1986). *Subjectivity, objectivity and teachers' qualitative judgements*. Assessment Unit Discussion Paper 5. Brisbane, Australia: Board of Secondary School Studies.
- Sadler, D.R. (1987). Specifying and Promulgating Achievement Standards, *Oxford Review of Education*, 13(2), 191-209.
- Sadler, D.R. (1998). Formative assessment: revisiting the territory. *Assessment in Education*, 5, 77-84.
- Schulz, E.M., & Nicewander, W.A. (1997). Grade equivalent and IRT representations of growth. *Journal of Educational Measurement*, 34(4), 315-331.
- Shepard, L.A. (2000) *The Role of Classroom Assessment in Teaching and Learning*. CSE Technical Report 517. CRESST/University of Colorado at Boulder.
- Shepard, L., Hammerness, K., Darling-Hammond, L., Rust, F., Snowden, J.B., Gordon, E., Gutierrez, C., & Pacheco, J. (2005). Chapter 8 in Darling-Hammond, L. and Bransford, J. (Eds) *Preparing teachers for a changing world: What teachers should know and be able to do*. San Francisco, CA: Jossey-Bass.
- Sharpley, C.F., & Edgar, E. (1986). Teachers' ratings vs standardized tests: an empirical investigation of agreement between two indices of achievement. *Psychology in the Schools*, 23, 106–111.
- Shinn, M.R. (Ed.) (1989). *Curriculum-based measurement: Assessing special children*. New York, NY: The Guilford Press.
- Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153-189.
- Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Slinde, J.A., & Linn, R.L. (1978). An exploration of the adequacy of the Rasch model for the problem of vertical equating. *Journal of Educational Measurement*, 15, 23-35.
- Smith, N.J. (2008). Next generation state data system: What is needed to support the next generation assessment and accountability systems. Data Quality Campaign. Retrieved from [www.DataQualityCampaign.org](http://www.DataQualityCampaign.org)

- Sneyd-Kynnersley, E.M. (1913). *HMI: Some passages in the life of one of H.M. inspectors of schools*. London, England: MacMillan.
- South Australian Curriculum, Standards and Accountability Framework. (2000). *Implementation plan 2001 to 2002*. Adelaide, Australia: Department of Education, Training and Employment.
- Spady, W. (1993). *Outcomes-based education*. Workshop Report Number 5. Canberra, Australia: ACSA,.
- Stake, R.E. (Ed). (1991). *Using assessment policy to reform education: Advances in program evaluation*. Greenwich, CT: JAI Press.
- Star Reading. (2005). *Star reading computer adaptive test and database technical manual*. Wisconsin Rapids, WI: Renaissance Learning, Inc.
- Starch, D. (1916). *Educational Measurements*. New York, NY: The Macmillan Company.
- Starch, D., & Elliott, E.C. (1912). Reliability of grading high school work in English, *School Review*, 20, 442-457.
- State Educational Technology Directors Association (SETDA). (2008). *Technology-Based Assessments Improve Teaching and Learning*. Retrieved from <http://www.setda.org/web/guest/2020>
- Statistical First Release 29/1999 (1999). *National Curriculum Assessments of 7, 11 and 14 year olds by Local Education Authority 1999*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 43/2000 (2000). *National Curriculum Assessments of 7, 11 and 14 year olds by Local Education Authority 2000*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 37/2001 (2001). *National Curriculum Assessments for 7, 11 and 14 year olds in England 2001*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 21/2002 (2002). *National Curriculum assessments of 7 and 11 year olds in England 2002 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 20/2003 (2003). *National Curriculum Assessments of 7, 11 and 14 year olds in England, 2003 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 30/2004 (2004). *National Curriculum Assessments of 11 year olds in England, 2004 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 31/2005 (2005). *National Curriculum Assessments of 11 year olds in England, 2005 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 31/2006 (2006). *National Curriculum Assessments at Key Stage 2 in England, 2006 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 24/2007 (2007). *National Curriculum Assessments at Key Stage 2 in England, 2007 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>

- Statistical First Release 19/2008 (2008). *National Curriculum Assessments at Key Stage 2 in England, 2008 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 20/2008. (2008). *National curriculum assessments at Key Stage 3 in England, 2008 (Provisional)*. Department for Children, Schools and Families. Retrieved from <http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000836/index.shtml>
- Statistical First Release 21/2008. (2008). *National curriculum assessments at Key Stage 1 in England, 2008*. Department for Children, Schools and Families. Retrieved from <http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000806/index.shtml>
- Statistical First Release 06/2009 (2009). *National curriculum assessments at Key Stage 2 in England, 2008 (Revised)*. Department for Children, Schools and Families. Retrieved from <http://www.education.gov.uk/rsgateway/DB/SFR/index.shtml>
- Statistical First Release 19/2009. (2009). *National curriculum assessments at Key Stage 2 in England, 2009*. Department for Children, Schools and Families. Retrieved from <http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000865/index.shtml>
- Stehn, J. (1997). Renewing the curriculum cycle in South Australia. In J. Lokan (Ed.). (1997). *Describing Learning; Implementation of Curriculum Profiles in Australian Schools 1986-1996*. (pp. 166-194). Melbourne, Australia: Australian Council for Educational Research.
- Stenner, A.J., & Stone, M.H. (2004, May), Does the reader comprehend the text because the reader is able or because the text is easy? Paper presented at International Reading Association, Reno-Tahoe, Nevada. Retrieved from Metametrics website [www.lexiles.com](http://www.lexiles.com)
- Stenner, A.J., Burdick, H., Sanford, E.E., & Burdick, D.S. (2007). The Lexile framework for reading technical report. Retrieved from Metametrics website [www.lexiles.com](http://www.lexiles.com).
- Stenner, J. (1996). *Measuring reading comprehension with the Lexile framework*. Paper presented at the Fourth North American Conference on Adolescent/Adult Literacy. Washington, D.C. February 1996.
- Stiggins, R. (2002). Leadership for Excellence in Assessment, A Powerful New school District Planning Guide. Portland, Oregon: Assessment Training Institute.
- Stiggins, R.J. (2008). *Assessment Manifesto: A Call for the Development of Balanced Assessment Systems*. Princeton, N.J.: Educational Testing Service.
- Stobart, G. (2001). The validity of national curriculum assessment. *British Journal of Educational Studies*, 49(1), 26–39.
- Stone, C.W. (1908) *Arithmetical Abilities, and Some Factors Determining Them*. *Teacher's College Series No 19*, Columbia University.
- Strøm, B. (2004). *Student achievement and birthday effects*. Mimeo, Norwegian University of Science and Technology. Retrieved from <http://www.hks.harvard.edu/pepg/PDF/events/Munich/PEPG-04-24Strom.pdf>
- Suppes, P., Fletcher, J.D., & Zanotti, M. (1976). Models of individual trajectories in computer-assisted instructions for deaf students. *Journal of Educational Psychology*, 68, 117-127.
- Test Statistics. (2007). *Key stage 2 – English, mathematics and science test statistics*. Retrieved from [http://www.naa.org.uk/naa\\_19225.aspx](http://www.naa.org.uk/naa_19225.aspx)
- National Council of Teachers of English (1917). Proceedings of the sixth annual meeting, New York City, November 30 and December 1-2, 1916. *The English Journal*, 6(1), 40-68.

- Thomas, C.S. (1913). The Hillegas Scale, *The English Leaflet*, 104, 1-4. Retrieved from <http://www.archive.org/details/hillegasscale00newerich>
- Thomson, P. (1999). How doing justice got boxed in: a cautionary curriculum tale for policy activists. In B. Johnson & A. Reid (Eds.). *Contesting the curriculum*. Katoomba, Australia: Social Science Press.
- Thorndike, E.L. (1910). Handwriting. *Teachers College Record*, 11(2), 1-81. Retrieved from <http://www.tcrecord.org>
- Thorndike, E.L. (1911). A scale for measuring the merit of English writing. *Science*, 33(859), 935-938.
- Thorndike, E.L. (1912). The measurement of educational products. *The School Review*, 20(5), 289-299.
- Thorndike, E.L. (1913). Notes on the Significance and Use of the Hillegas Scale for Measuring the Quality of English Composition. *The English Journal*, 2(9), 551-561.
- Thorndike, E.L. (1914a). Educational psychology volume 3: Mental work and fatigue and individual differences and their causes. New York, NY: Teachers College, Columbia University.
- Thorndike, E.L. (1914b). The measurement of ability in reading: Preliminary scales and tests. *Teachers College Record*, 15(4), 1-71. Retrieved from <http://www.tcrecord.org>
- Thorndike, E.L. (1915). An improved scale for measuring ability in reading. *Teachers College Record*, 16(1), 31-53. Retrieved from <http://www.tcrecord.org>
- Thorndike, E.L. (1916a). *An introduction to the theory of mental and social measurements*. (2nd Ed.). New York, NY: Teachers College, Columbia University.
- Thorndike, E.L. (1916b). Measurement of achievement in reading: Word knowledge. *Teachers College Record*, 17(5), 430-454. Retrieved from <http://www.tcrecord.org>
- Thorndike, R. L. (1971). *Educational measurement*. (2nd ed.) Washington, DC: American Council on Education.
- Thorndike, R.M., & Lohman, D. (1990). *A century of ability testing*. Chicago, IL: Riverside.
- Thurber. (1913). From Mr. Thurber. *The English Leaflet*, 104, 5-6. Retrieved from <http://www.archive.org/details/hillegasscale00newerich>
- Thurstone, L.L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433-451.
- Thurstone, L.L. (1928). Scale construction with weighted observations. *Journal of Educational Psychology*, 19, 441-453.
- Timperley, H. (2009, August). *Using assessment data for improving teaching practice*. Paper presented to the Australian Council for Educational Research Assessment and Student Learning Conference, Perth.
- Tourangeau, K., Nord, C., Lê T., Pollack, J.M., and Atkins-Burnett, S. (2006). Early Childhood Longitudinal Study, Kindergarten class of 1998–99 (ECLS-K), Combined user's manual for the ECLS-K Fifth-grade data files and electronic codebooks (NCES 2006–032). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Trabue, M.R. (1916). Trabue's dissertation on completion tests. *Teachers College Record*, 17(1), 85-91. Retrieved from <http://www.tcrecord.org>
- Trabue, M.R. (1917) Supplementing the Hillegas Scale. *Teachers College Record*, 18(1), 51-84. Retrieved from <http://www.tcrecord.org>

- Training and Development Agency for Schools. (2008). *Professional Standards for Qualified Teacher and Requirements for Initial Teacher Training*. Retrieved from <http://www.tda.gov.uk/teachers/professionalstandards/downloads.aspx>
- Triga, A. (2004). An analysis of teachers' rating scales as sources of evidence for a standardised Greek reading test. *Journal of Research in Reading*, 27, 311–320.
- Tymms, P., Merrell, C., & Jones, P. (2003, October). *The problems of creating a one-entry assessment across culture and languages*. Paper presented to the 29th International Association for Educational Assessment (IAEA) Annual Conference.
- Tymms, P. (2004). Are standards rising in English primary schools? *British Educational Research Journal*, 30, 477-494.
- Tymms, P., & Merrell, C. (2007). *Standards and quality in English primary schools over time: the national evidence*. Primary Review Research Survey 4/1. Cambridge, England: University of Cambridge Faculty of Education.
- Tymms, P., & Wylde, M. (2003, April). *Baseline assessment and monitoring in primary schools*. Paper presented at the Symposium Connectable Processes in Elementary and Primary Section. Bamberg. Retrieved from [www.cemcentre.org/publications/downloads/germansymposium.doc](http://www.cemcentre.org/publications/downloads/germansymposium.doc)
- Tymms, P., Merrell, C., & Jones, P. (2004). Using baseline assessment data to make international comparisons, *British Educational Research Journal*, 30, 673-689.
- Victorian Auditor-General. (2009). *Literacy and numeracy achievement*. Melbourne, Australia: Victorian Government Printer.
- Victorian Curriculum and Assessment Authority. (2004). *Victorian curriculum reform 2004 consultation paper*. Retrieved from [www.vcaa.vic.edu.au](http://www.vcaa.vic.edu.au)
- Victorian Curriculum and Assessment Authority. (2006a). *VELS Standards and Progression Points: English*. Retrieved from <http://vels.vcaa.vic.edu.au>
- Victorian Curriculum and Assessment Authority. (2006b). *VELS Standards and Progression Points: Mathematics*. Retrieved from <http://vels.vcaa.vic.edu.au>
- Victorian Curriculum and Assessment Authority. (2009). *National Assessment Program Literacy and Numeracy (NAPLAN) reference guide – Analysing NAPLAN data*. Retrieved from [www.vcaa.vic.edu.au/prep10/naplan/schools/analysingnaplandata.pdf](http://www.vcaa.vic.edu.au/prep10/naplan/schools/analysingnaplandata.pdf)
- Victorian Institute of Teaching. (2009). *Standards For Graduating Teachers*. Retrieved from [http://www.vit.vic.edu.au/content.asp?Document\\_ID=23](http://www.vit.vic.edu.au/content.asp?Document_ID=23)
- von Davier, A.A., Holland, P.W., & Thayer, D.T. (Eds.). (2004). *The kernel method of test equating*. New York, NY: Springer-Verlag.
- Vygotsky, L.S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Walston, J., Rathbun, A., & Germino Hausken, E. (2008). *Eighth grade: First findings from the final round of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99*. (NCES 2008-088). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Wang, S., & Jiao, H. (2009). Construct equivalence across grades in a vertical scale for a K-12 large-scale reading assessment. *Educational and Psychological Measurement*, 69, 760-777.
- Wang, S., Jiao, H., Brooks, T., & Young, M.J. (2004). *Construct equivalence between customized and original Stanford Achievement Reading Comprehension Tests* (10th ed.) Research Report. San Antonio, TX: Harcourt Assessment. Cited in Wang, S. & Jiao, H. (2009). Construct equivalence across grades in a vertical scale for a K-12

- large-scale reading assessment. *Educational and Psychological Measurement*, 69, 760-777.
- Wasik, B.H., & Loven, M.D. (1980). Classroom observational data: Sources of inaccuracy and proposed solutions. *Behavioral Assessment*, 2, 211-227.
- Wiliam, D. (1994). *Towards a philosophy for educational assessment*. Revised version of paper originally presented to the British Educational Research Association's 20th annual conference in Oxford in 1994. Retrieved from <http://www.kcl.ac.uk//depsta/education/hpages/dwliam.html>
- Wiliam, D. (1998, September). Enculturating learners into communities of practice: Raising achievement through classroom assessment. Paper presented at European Conference on Educational Research, Ljubljana, Slovenia. Cited by Forster (2009).
- Wiliam, D. (1999). Standards: Annex 4 in Weighing the baby: the report of the independent scrutiny panel on the 1999 Key Stage 2 national curriculum tests in English and mathematics. London, Department for Employment and Education. Cited by Stobart, G. (2001). The validity of national curriculum assessment. *British Journal of Educational Studies*, 49, 26-39.
- Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, 34, 254-284.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C.A. Dwyer.(Ed.) *The future of assessment: shaping teaching and learning*. Mahwah, NJ: Erlbaum.
- Willett, J.B., & Sayer, A.G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, 116, 363-381.
- Williams, D., Johnson, B., Peters, J., & Cormack, P. (1999). Assessment: from standardised to authentic approaches. In B. Johnson & A. Reid (Eds.). *Contesting the curriculum*. Katoomba, Australia: Social Science Press.
- Williams, J., Wo, L., & Lewis, S. (2007). Mathematics progression 5-14: Plateau, curriculum/age and test year effects. *Research in Mathematics Education*, 9, 127-142.
- Williamson, G.L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. *Journal of Educational Measurement*, 28, 61-76.
- Williamson, G.L. (2006). *What is expected growth?* A white paper from The Lexile Framework for Reading. Retrieved from Metametrics website [www.lexiles.com](http://www.lexiles.com)
- Wilson, B. (1993). *National curriculum and national profiles: Development, implementation and use*. Incorporated Association of Registered Teachers of Victoria Seminar Series No 23. Melbourne, Australia.
- Wilson, B. (1994). Profiles meet post structuralism. *Curriculum Perspectives*, 14(2), 20-21.
- Wilson, M. (2004). Assessment, Accountability and the Classroom: A Community of Judgment. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability*. 103<sup>rd</sup> Yearbook of the National Society for the Study of Education. (pp. 1-19). Chicago, IL: University of Chicago Press.
- Wilson, M. (Ed.). (2004). *Towards coherence between classroom assessment and accountability*. 103<sup>rd</sup> Yearbook of the National Society for the Study of Education. Chicago, IL: University of Chicago Press.
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13, 181-208.

- Withers, G. (1999). *Supporting the implementation of statements and profiles in South Australian schools: Report of a survey conducted in November – December 1998*. Melbourne, Australia: Australian Council for Educational Research.
- Wright, B.D. (1968). Sample-free test calibration and person measurement. Proceedings 1967 Invitational Conference on Testing Princeton: Educational Testing Service, 85-101. Cited in Wright (1977).
- Wright, B.D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97-116.
- Wright, B.D., & Stone, M.H. (1979). *Best test design*. Chicago, IL: MESA Press.
- Wright, B.D., & Stone, M.H. (1999). *Measurement essentials*. (2nd ed.). Wilmington, DE: Wide Range Inc. Chapter retrieved from [www.rasch.org/measess/met-9.pdf](http://www.rasch.org/measess/met-9.pdf)
- Wright, B.D., & Stone, M.H. (2004). *Making measures*. Chicago, IL: The Phaneron Press.
- Wright, B.D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice*, 16(4), 33-45.
- Wright, B.D. (2001). Counts or measures? Which communicate best? *Rasch Measurement Transactions*, 14(4), 784.
- Wright, D. & Wiese, M.J. (1988). Teacher judgement in student evaluation: A comparison of grading methods. *Journal of Educational Research*, 82, 10-14.
- Wyatt, H.G. (1917). *Methods of school inspection in England*. Occasional Reports No. 7. , India: Bureau of Education. Retrieved from <http://www.archive.org/details/methodsofschooli00wyatrich>
- Yeh, S.S. (2006). *Raising student achievement through rapid assessment and test reform*. New York, NY: Teachers College Press.
- Yen, W.M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125-145.
- Yen, W.M. 1985. Increasing item complexity: A possible cause of scale shrinkage for unidimensional item response theory. *Psychometrika*, 50(4), 399–410.
- Yen, W.M. 1986. The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement*, 23(4), 299–325.