References

Adams, R. J., Wilson, M., and Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, *22*, 47–76.

Afshartous, D. and De Leeuw, J. (2005). Prediction in multilevel models. *Journal of Educational and Behavioral Statistics*, *30*, 109–139.

Aitkin, M. (1997). The calibration of p-values, posterior Bayes factors and the AIC from the posterior distribution of the likelihood. *Statistics and Computing*, *7*, 253–261.

Aitkin, M. and Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, *149*, 1–43.

Albers, W., Does, R. J. M. M., Imbos, T., and Janssen, M. P. E. (1989). A stochastic growth model applied to repeated tests of academic knowledge. *Psychometrika*, *54*, 451–466.

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, *17*, 251–269.

Albert, J. H. and Chib, S. (1993). Bayesian analysis for binary and polychotomous response data. *Journal of the American Statistical Association*, *88*, 669–679.

Albert, J. H. and Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika*, *82*, 747–769.

Anderson, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. Amsterdam: North-Holland.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Hoboken, NJ: Wiley.

American Educational Research Association, American Psychological Association, and National Council of Measurement in Education (2000). *Standards for Educational and Psychological Testing 1999,* 2nd ed. Washington, DC: AERA.

Baker, F. B. and Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques,* 2nd ed. New York: Marcel Dekker.

Barnard, J., McCullogh, R. E., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Bartholomew, D. J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis,* 2nd ed. London: Arnold.

Bayarri, M. J. and Berger, J. O. (1999). Quantifying surprise in the data. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6* (pp. 53–82). New York: Oxford University Press.

Bayarri, M. J. and Berger, J. O. (2000). P values for composite null models. *Journal of the American Statistical Association*, *95*, 1127–1142.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 370–418.

Béguin, A. A. and Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*, 541–561.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis,* 2nd ed. New York: Springer.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *2*, 317–335.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for linear models. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 5* (pp. 25–44). New York: Oxford University Press.

Berger, J. O. and Selke, T. (1987). Testing a point null hypothesis: The irreconcilability of P values and evidence. *Journal of the American Statistical Association*, *82*, 112–122.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* New York: Wiley.

Best, N. G., Cowles, M. K., and Vines, K. (2010). CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, version 0.5-1 [computer software and manual]. Retrieved from `http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml`.

Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology*, *6*, 258–276.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, MA: The Massachusetts Institute of Technology.

Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practices*, *16*, 21–33.

Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.

Bock, R. D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, *35*, 179–197.

Böckenholt, U. and van der Heijden, P. G. M. (2007). Item randomized-response models for measuring noncompliance: Risk-return perceptions, social influences, and self-protective responses. *Psychometrika*, *72*, 245–262.

Boscardin, W. J. and Zhang, X. (2004). Modeling the covariance and correlation matrix of repeated measures. In A. Gelman and X.-L. Meng (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (pp. 215–226). New York: Wiley.

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Series A*, *143*, 383–430.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

Bradlow, E. T., Wainer, H., and Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, *64*, 153–168.

Braun, H. I., Jones, D. H., Rubin, D. B., and Thayer, D. T. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. *Psychometrika*, *48*, 171–181.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Browne, W. J. (2006). MCMC algorithms for constrained variance matrices. *Computational Statistics and Data Analysis*, *50*, 1655–1677.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman and Hall.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.

Casella, G. and Berger, R. L. (2002). *Statistical Inference,* 2nd ed. Pacific Grove, CA: Duxbury Thomson Learning.

Chaloner, K. and Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika*, *75*, 651–659.

Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, *8*, 69–92.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, *49*, 327–335.

Clark, S. J. and Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods*, *3*, 160–168.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*, 249–253.

Commenges, D. and Jacqmin, H. (1994). The intraclass correlation coefficient: Distribution-free definition and test. *Biometrics*, *50*, 517–526.

Congdon, P. (2001). *Bayesian Statistical Modelling*. Chichester: Wiley.

Cowles, M. K. (1996). Accelerating Monte Carlo Markov Chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, *6*, 101–111.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Jounal of the American Statistical Association*, *91*, 883–904.

Cruyff, M. J. L. F., van den Hout, A., van der Heijden, P. G. M., and Böckenholt, U. (2007). Log-linear randomized-response models taking self-protective response behavior into account. *Sociological Methods and Research*, *36*, 266–282.

Davis, S. F., Grover, C. A., Becker, A. H., and McGregor, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology*, *19*, 16–20.

De Boeck, P. and Wilson, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.

De Jong, M. G., Pieters, R., and Fox, J.-P. (2010). Reducing social desirability bias through item randomized response: An application to measure underreported desires. *Journal of Marketing Research*, *47*, 14–27.

De Jong, M. G. and Steenkamp, J. B. E. M. (2009). Finite mixture multilevel multidimensional ordinal IRT models for large scale cross-cultural research. *Psychometrika*, *(online)*.

De Jong, M. G., Steenkamp, J. B. E. M., and Fox, J.-P. (2007). Relaxing cross-national measurement invariance using a hierarchical IRT model. *Journal of Consumer Research*, *34*, 260–278.

De Jong, M. G., Steenkamp, J. B. E. M., Fox, J.-P., and Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104–115.

De Leeuw, J. and Kreft, I. G. G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational and Behavioral Statistics*, *11*, 57–85.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: Wiley.

DeIorio, M. and Robert, C. P. (2002). Discussion of Spiegelhalter et al. *Journal of the Royal Statistical Society, Series B*, *64*, 629–630.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, *76*, 341–353.

Dickey, J. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Statistics*, *42*, 204–223.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B*, *56*, 363–375.

Donders, F. C. (1868). Over de snelheid van psychische processen [On the speed of mental processes]. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868–1869, Tweede reeks, II, 92–120.

Doolaard, S. (1999). *Schools in Change or Schools in Chains?* PhD dissertation, University of Twente.

Edgell, S. E., Himmelfarb, S., and Duchan, K. L. (1982). Validity of forced responses in a randomized response model. *Sociological Methods and Research*, *11*, 89–100.

Edwards, A. W. F. (1963). The measure of association in a 2x2 table. *Journal of the Royal Statistical Society, Series A*, *126*, 109–114.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalization. *Journal of the American Statistical Association*, *70*, 311–319.

Embretson, S. E. and Reise, S. P. (2000). *Item Response Theory for Psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Emons, W. H. M., Sijtsma, K., and Meijer, R. R. (2005). Global, local, and graphical person-fit analysis using person-response functions. *Psychological Methods*, *10*, 101–119.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models,* 2nd ed. New York: Springer.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-mental/state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Fox, J. A. and Tracy, P. E. (1986). *Randomized Response: A Method for Sensitive Surveys.* Beverly Hills, CA: Sage.

Fox, J.-P. (2001). *Multilevel IRT: A Bayesian Perspective on Estimating Parameters and Testing Statistical Hypotheses.* PhD dissertation, University of Twente, Faculty of Behavioural Sciences.

Fox, J.-P. (2003). Stochastic EM for estimating the parameters of a multilevel IRT model. *British Journal of Mathematical and Statistical Psychology*, *56*, 65–81.

Fox, J.-P. (2004). Applications of multilevel IRT modeling. *School Effectiveness and School Improvement*, *15*, 261–280.

Fox, J.-P. (2005a). Multilevel IRT model assessment. In A. van der Ark, M. A. Croon, and K. Sijtsma (Eds.), *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences* (pp. 227–252). Mahwah, NJ: Lawrence Erlbaum.

Fox, J.-P. (2005b). Multilevel IRT using dichotomous and polytomous items. *British Journal of Mathematical and Statistical Psychology*, *58*, 145–172.

Fox, J.-P. (2005c). Randomized item response theory models. *Journal of Educational and Behavioral Statistics*, *30*, 189–212.

Fox, J.-P. (2007). Multilevel IRT modeling in practice. *Journal of Statistical Software*, *20*, Issue 5.

Fox, J.-P. and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271–288.

Fox, J.-P. and Glas, C. A. W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*, *68*, 169–191.

Fox, J.-P., Klein Entink, R. E., and van der Linden, W. J. (2007). Modeling of responses and response times with the package cirt. *Journal of Statistical Software*, *20*, Issue 7.

Fox, J.-P. and Meijer, R. R. (2008). Using item response theory to obtain individual information from randomized response data: An application using cheating data. *Journal of Applied Psychological Measurement*, *32*, 595–610.

Fox, J.-P. and Wyrick, C. (2008). A mixed effects randomized item response model. *Journal of Educational and Behavioral Statistics*, *33*, 389–415.

Fuller, W. A. (1987). *Measurement Error Models*. New York: Wiley.

Fuller, W. A. (1991). Regression estimation in the presence of measurement error. In P. P. Biemer, R. M. Groves, L. E. Lyberg, and N. A. Mathiowetz (Eds.), *Measurement Errors in Surveys* (pp. 617–635). New York: Wiley.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, *70*, 320–328.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B*, *56*, 501–514.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling based methods (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4* (pp. 147–167). Oxford: Oxford University Press.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.

Gelman, A. (1995). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 131–143). London: Chapman and Hall.

Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis*, *3*, 445–450.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gelman, A. and King, D. G. (1990). Estimating the electoral consequences of legislative redirecting. *Journal of the American Statistical Association*, *85*, 274–282.

Gelman, A., Meng, X.-L., and Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–807.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4* (pp. 169–193). Oxford: Oxford University Press.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Hoboken, NJ: Wiley.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science*, *4*, 473–483.

Ghosh, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statistics and Probability Letters*, *23*, 165–170.

Ghosh, M., Ghosh, A., Chen, M.-H., and Agresti, A. (2000). Noninformative priors for one-parameter item response models. *Journal of Statistical Planning and Inference*, *88*, 99–115.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1995). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Glas, C. A. W. and Meijer, R. R. (2003). A Bayesian approach to person fit analysis in item response theory models. *Applied Psychological Measurement*, *27*, 217–233.

Glas, C. A. W. and van der Linden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement*, *27*, 247–261.

Goldstein, H. (2003). *Multilevel Statistical Models,* 3rd ed. London: Hodder Arnold.

Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education*, *11*, 319–330.

Goldstein, H., Bonnet, G., and Rocher, T. (2007). Multilevel structural equation models for the analysis of comparative data on educational performance. *Journal of Educational and Behavioral Statistics*, *32*, 252–286.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G., and Healy, M. (1998). *A User's Guide to MLwiN*. London: Multilevel Models Project, Institute of Education.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*, 215–231.

Greenberg, B. G., Abul-Ela, A., Simmons, W. R., and Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *The American Statistician*, *64*, 520–539.

Gulliksen, H. O. (1950). *Theory of Mental Tests*. New York: Wiley.

Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit problems. *Journal of the Royal Statistical Society, Series B*, *29*, 83–100.

Hall, D. B. and Clutter, M. (2004). Multivariate multilevel nonlinear mixed effects models for timber yield predictions. *Biometrics*, *60*, 16–24.

Hambleton, R. K. and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. London: Sage.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and related problems. *Journal of the American Statistical Association*, *72*, 320–340.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97–109.

Hedeker, D. R. (1999). MIXNO: A computer program for mixed-effects nominal logistic regression. *Journal of Statistical Software*, *4*, 1–92.

Hedeker, D. R. and Gibbons, R. D. (1996). MIXOR: A computer program for mixed-effects ordinal probit and logistic regression analysis. *Computer Methods and Programs in Biomedicine*, *49*, 157–176.

Hedeker, D. R. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Hoboken, NJ: Wiley.

Heidelberg, P. and Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*, 1109–1144.

Higdon, D. M. (1998). Auxiliary variable methods for Markov Chain Monte Carlo with applications. *Journal of the American Statistical Society*, *93*, 585–595.

Hill, B. M. (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Society*, *60*, 806–825.

Hoijtink, H. (2001). Conditional independence and differential item functioning in the two-parameter logistic model. In A. Boomsma, M. A. J. van Duijn, and T. A. B. Snijders (Eds.), *Essays on Item Response Theory* (pp. 109–129). New York: Springer.

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, *55*, 577–601.

Holland, P. W. and Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, NJ: Erlbaum.

Janssen, R., Schepers, J., and Peres, D. (2004). Models with item and item group predictors. In P. de Boeck and M. Wilson (Eds.), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach* (pp. 189–212). New York: Springer.

Janssen, R., Tuerlinckx, F., Meulders, M., and De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics*, *25*, 285–306.

Jeffreys, H. J. (1961). *Theory of Probability*. New York: Oxford University Press.

Johnson, V. E. and Albert, J. H. (1999). *Ordinal Data Modeling*. New York: Springer.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, *38*, 79–93.

Karim, M. R. and Zeger, S. L. (1992). Generalized linear models with random effects; salamander mating revisited. *Biometrics*, *48*, 631–644.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kim, S.-H. (2001). An evaluation of a Markov chain Monte Carlo method for the Rasch model. *Applied Psychological Measurement*, *25*, 163–176.

Kim, S.-H., Cohen, A. S., Baker, F. B., Subkoviak, M. J., and Leonard, T. (1994). An investigation of hierarchical Bayes procedures in item response theory. *Psychometrika*, *59*, 405–421.

Klein Entink, R. H. (2009). *Statistical Models for Responses and Response Times*. PhD dissertation, University of Twente, Faculty of Behavioural Sciences.

Klein Entink, R. H., Fox, J.-P., and van der Linden, W. J. (2009a). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, *74*, 21–48.

Klein Entink, R. H., Kuhn, J.-T., Hornke, L. F., and Fox, J.-P. (2009b). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, *14*, 54–75.

Kuha, J. (1997). Estimation by data augmentation in regression models with continuous and discrete covariates measured with error. *Statistics in Medicine*, *16*, 189–201.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*. New York: Houghton Mifflin.

Lee, P. M. (2004). *Bayesian Statistics: An Introduction,* 3rd ed. New York: Wiley.

Lee, S.-Y. and Zhu, H.-T. (2000). Statistical analysis of non-linear equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *53*, 209–232.

Lehmann, E. L. and Casella, G. (2003). *Theory of Point Estimation,* 2nd ed. New York: Springer.

Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., and Maas, C. J. M. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods and Research*, *33*, 319–348.

Leonard, T. and Hsu, J. S. J. (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge: Cambridge University Press.

Levy, R. (2006). *Posterior Predictive Model Checking for Multidimensionality in Item Response Theory and Bayesian Networks*. PhD dissertation, University of Maryland.

Lindley, D. V. (1965). *An Introduction to Probability and Statistics from a Bayesian Viewpoint (Parts 1 and 2).* Cambridge: Cambridge University Press.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B*, *34*, 1–41.

Little, R. J. A. and Rubin, D. A. (2002). *Statistical Analysis with Missing Data,* 2nd ed. Hoboken, NJ: Wiley.

Liu, J. S., Wong, W. H., and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, *81*, 27–40.

Liu, L. C. and Hedeker, D. (2006). A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics*, *62*, 261–268.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*, 817–827.

Longford, N. T. (1993). *Random Coefficient Models.* New York: Oxford University Press.

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, *23*, 157–162.

Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization.* New York: Oxford University Press.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility [computer software]. *Statistics and Computing*, *10*, 325–337.

MacEachern, S. N. and Berliner, L. M. (1994). Subsampling the Gibbs sampler. *The American Statistician*, *48*, 188–190.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Maier, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, *26*, 307–330.

Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and their application as psychometric models for response times. *Psychometrika*, *58*, 445–469.

Maris, G. and Maris, E. (2002). A MCMC-method for models with continuous latent responses. *Psychometrika*, *67*, 335–350.

Masters, G. M. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.

Masters, G. N. and Wright, B. D. (1997). The partial credit model. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 101–121). New York: Springer.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd ed. New York: Chapman and Hall.

McCulloch, R. E., Polson, N. G., and Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, *99*, 173–193.

McDonald, R. P. (1967). *Nonlinear Factor Analysis* (Psychometric Society Monograph No. 15). Richmond, VA: William Byrd Press.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models.* New York: Wiley.

Meijer, R. R. (2003). Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychological Methods*, *8*, 72–87.

Meijer, R. R. and Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, *25*, 107–135.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*, 127–143.

Mellenbergh, G. J. (1994a). Generalized linear item response theory. *Psychological Bulletin*, *115*, 300–307.

Mellenbergh, G. J. (1994b). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, *29*, 223–236.

Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics*, *22*, 1142–1160.

Meng, X.-L. and van Dyk, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, *86*, 301–320.

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, *6*, 831–860.

Meredith, W. and Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, *57*, 289–311.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*, 1087–1092.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, *49*, 359–381.

Mislevy, R. J. (1986). Bayes model estimation in item response models. *Psychometrika*, *51*, 177–195.

Mislevy, R. J. and Sheehan, K. M. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika*, *54*, 661–679.

Molenaar, I. W. (1995). Some background for item response theory and the Rasch model. In G. H. Fisher and I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments and Applications* (pp. 3–14). New York: Springer.

Molenaar, I. W. and Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika*, *55*, 75–106.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, *78*, 47–55.

Moshagen, M. (2008). *Multinomial Randomized Response Models*. Phd dissertation, Heinrich-Heine-Universität Dusseldörf, Mathematisch Naturwissenschaftlichen Fakultät.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176.

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, *17*, 351–363.

Muraki, E. and Bock, R. D. (1997). *PARSCALE: IRT Based Test Scoring and Item Analysis for Graded Items and Rating Scales* [computer software]. Chicago, IL: Scientific Software International.

Muthén, B. O. (1992). Latent variable modeling in epidemiology. *Alcohol Health and Research World*, *16*, 286–292.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides and R. E. Schumacker (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1–33). New York: Lawrence Erlbaum Associates.

Muthén, B. O. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469.

Muthén, L. K. and Muthén, B. O. (1998). *Mplus: The Comprehensive Modeling Program for Applied Researchers* [computer software]. Los Angeles, CA: Muthén and Muthén.

Nandram, B. and Chen, M.-H. (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation*, *54*, 129–144.

Neal, R. M. (1997). Markov Chain Monte Carlo methods based on 'slicing' the density function. Technical Report No. 9722, University of Toronto, Department of Statistics.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B*, *56*, 3–48.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, *16*, 1–32.

Novick, M. R., Lewis, C., and Jackson, P. H. (1973). The estimation of proportions in m groups. *Psychometrika*, *38*, 19–46.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: Wiley.

OECD (Organisation for Economic Co-operation and Development) (2004). *Learning From Tomorrow's World. First Results from PISA 2003*. Paris: OECD.

O'Hagan, A. (1976). On posterior joint and marginal modes. *Biometrika*, *63*, 329–333.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society, Series B*, *57*, 99–138.

O'Hare, T. M. (1997). Measuring problem drinkers in first time offenders: Development and validation of the college alcohol problem scale (CAPS). *Journal of Substance Abuse Treatment*, *14*, 383–387.

Orlando, M. and Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous response theory models. *Applied Psychological Measurement*, *24*, 50–64.

Orlando, M. and Thissen, D. (2003). Further investigation of the performance of $s - x^2$: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, *27*, 289–298.

Ostini, R. and Nering, M. L. (2006). *Polytomous Item Response Theory Models*. Thousand Oaks, CA: Sage.

Owen, R. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, *70*, 351–356.

Patz, R. J. and Junker, B. W. (1999a). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, *24*, 342–366.

Patz, R. J. and Junker, B. W. (1999b). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*, 146–178.

Patz, R. J., Junker, B. W., Johnson, M. S., and Mariano, L. T. (2002). The hierarchical rater model for rated test items and its application to large-scale educational assessment data. *Journal of Educational and Behavioral Statistics*, *27*, 341–384.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus*. New York: Springer.

Press, S. J. (2003). *Subjective and Objective Bayesian Statistics*. Hoboken, NJ: Wiley.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rabe-Hesketh, S. and Skrondal, A. (2001). Parameterization of multivariate random effects models for categorical data. *Biometrics*, *57*, 1256–1264.

Raftery, A. E. (1995). Bayesian model selection in social research. *Social Methodology*, *25*, 111–163.

Raftery, A. L. and Lewis, S. (1992). Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, *7*, 493–497.

Rasch, G. (1960). *Probabilistic Models for some Intelligence Tests and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods,* 2nd ed. Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., and Congdon, R. T. (2000). *HLM 5: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

Raudenbush, S. W. and Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, *29*, 1–41.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, *9*, 401–412.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*, 25–36.

Reinsel, G. (1982). Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. *Journal of the American Statistical Association*, *77*, 190–195.

Reinsel, G. (1983). Some results on multivariate autoregressive index models. *Biometrika*, *70*, 145–156.

Richman, W. L., Kiesler, S., Weisband, S., and Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, *84*, 754–775.

Rigdon, S. E. and Tsutakawa, R. K. (1983). Parameter estimation in latent trait models. *Psychometrika*, *48*, 567–574.

Rijmen, F., Tuerlinckx, F., De Boeck, P., and Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, *8*, 185–205.

Ripley, B. D. (1987). *Stochastic Simulation*. New York: Wiley.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.

Roberts, G. O. (1995). Markov chain concepts related to sampling algorithms. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 45–57). London: Chapman and Hall.

Roberts, G. O. and Tweedie, R. L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and Their Applications*, *80*, 211–229; Correction *91*, 337–338.

Robins, J. M., van der Vaart, A. W., and Ventura, V. (2000). Asymptotic distribution of P values in composite null models: Rejoinder. *Journal of the American Statistical Association*, *95*, 1171–1172.

Rosenbaum, P. R. (1988). Items bundles. *Psychometrika*, *53*, 349–359.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, *90*, 558–566; Correction *91*, 1136.

Roskam, E. E. (1997). Models for speed and time-limit tests. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 187–208). New York: Springer.

Rossi, P. E., Allenby, G. M., and McCullogh, R. E. (2005). *Bayesian Statistics and Marketing*. Chichester: Wiley.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271–282.

Rost, J. and von Davier, M. (1995). Mixture distribution Rasch models. In G. Fischer and I. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments and Applications* (pp. 257–268). New York: Springer.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, *12*, 1151–1172.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Rupp, A. A., Dey, D. K., and Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling*, *11*, 424–451.

Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, *72*, 217–232.

Samejima, F. (1997). The graded response model. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 85–100). New York: Springer.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Schafer, J. L. and Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, *11*, 437–457.

Scheerens, J. (1992). *Effective Schooling: Research, Theory and Practice*. London: Cassell.

Scheerens, J., Glas, C. A. W., and Thomas, S. M. (2003). *Educational Evaluation, Assessment, and Monitoring*. Lisse: Swets and Zeitlinger.

Scheers, N. J. and Dayton, C. (1988). Covariate randomized response model. *Journal of the American Statistical Association*, *83*, 969–974.

Scheiblechner, H. (1979). Specifically objective stochastic latency mechanisms. *Journal of Mathematical Psychology*, *19*, 18–38.

Schnipke, D. L. and Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method for measuring speededness. *Journal of Educational Measurement*, *34*, 213–232.

Schnipke, D. L. and Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. In C. N. Mills, M. T. Potenza, J. J. Fremer, and W. C. Ward (Eds.), *Computer-Based Testing: Building the Foundation for Future Assessments* (pp. 237–266). Mahwah, NJ: Lawrence Erlbaum.

Schulz-Larsen, K., Kreiner, S., and Lomholt, R. K. (2007a). Mini-mental status examination: A short form of MMSE was as accurate in predicting dementia. *Journal of Clinical Epidemiology*, *60*, 260–267.

Schulz-Larsen, K., Kreiner, S., and Lomholt, R. K. (2007b). Mini-mental status examination: Mixed Rasch model item analysis derived two different

cognitive dimensions of the MMSE. *Journal of Clinical Epidemiology*, *60*, 268–279.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. New York: Wiley.

Shalabi, F. (2002). *Effective Schooling in the West Bank*. PhD dissertation, University of Twente.

Shi, J.-Q. and Lee, S.-Y. (1998). Bayesian sampling-based approach for factor analysis models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *51*, 233–252.

Sinharay, S. (2005). Assessing fit of unidimensional item response theory models using a Bayesian approach. *Journal of Educational Measurement*, *42*, 375–394.

Sinharay, S. (2006). Bayesian item fit analysis for unidimensional item response theory models. *British Journal of Mathematical and Statistical Psychology*, *59*, 429–449.

Sinharay, S., Johnson, M. S., and Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, *30*, 298–321.

Sinharay, S. and Stern, H. S. (2002). On the sensitivity of Bayes factors to the prior distributions. *The American Statistician*, *56*, 196–201.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. London: Chapman and Hall.

Smith, B. (2010). BOA: Bayesian Output Analysis Program, version 1.1.5 [computer software and manual]. Retrieved from `http://www.public-health.uiowa.edu/boa/`.

Snijders, T. A. B. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika*, *66*, 331–342.

Snijders, T. A. B. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Soares, T. M., Gonçalves, F. B., and Gamerman, D. (2009). An integrated Bayesian model for DIF analysis. *Journal of Educational and Behavioral Statistics*, *34*, 348–377.

Song, X.-Y. and Lee, S.-Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology*, *54*, 237–263.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, *64*, 583–639.

Steenkamp, J. B. E. M. and Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.

Stern, H. S. (2000). Asymptotic distribution of P values in composite null models: Comment. *Journal of the American Statistical Association*, *95*, 1157–1159.

Stone, C. A. and Hansen, M. A. (2000). The effect of errors in estimating ability on goodness-of-fit tests for IRT models. *Educational and Psychological Measurement*, *60*, 974–991.

Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L., and Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, *20*, 331–354.

Swaminathan, H. and Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, *7*, 175–192.

Swaminathan, H. and Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika*, *50*, 349–364.

Swaminathan, H. and Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika*, *51*, 589–601.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*, 528–540.

Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, *49*, 95–110.

Theil, H. (1963). On the use of incomplete prior information in regression analysis. *Journal of the American Statistical Association*, *58*, 401–414.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, *47*, 175–186.

Thissen, D. (1983). Timed testing: An approach using item response theory. In D. J. Weiss (Ed.), *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing* (pp. 179–203). New York: Academic Press.

Thissen, D. (1991). *MULTILOG: Multiple Category Item Analysis and Test Scoring Using Item Response Theory* [computer software]. Chicago, IL: Scientific Software International.

Thissen, D. and Wainer, H. (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum.

Tiao, G. C. and Tan, W. Y. (1965). Bayesian analysis of random-effects models in the analysis of variance. I: Posterior distribution of variance components. *Biometrika*, *52*, 37–53.

TIBCO Software (2009). *TIBCO Spotfire S+ 8.1: Programmer's Guide and Computer Program* [computer software]. TIBCO Software Inc.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*, 1701–1762.

Tombaugh, T. N. (1992). The mini-mental state examination: A comprehensive review. *The Journal of the American Geriatrics Society*, *40*, 922–935.

Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York: Wiley.

Tourangeau, R., Rips, L., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tourangeau, R. and Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*, 859–883.

Tracy, P. E. and Fox, J. A. (1981). The validity of the randomized response for sensitive measurements. *American Sociological Review*, *46*, 187–200.

Tsutakawa, R. K. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational Statistics*, *9*, 263–276.

Tsutakawa, R. K. and Lin, H. Y. (1986). Bayesian estimation of item response curves. *Psychometrika*, *51*, 251–267.

Tsutakawa, R. K. and Soltys, M. J. (1988). Approximation for Bayesian ability estimation. *Journal of Educational Statistics*, *13*, 117–130.

Tucker, L. R. (1952). A level of proficiency scale for a unidimensional skill. *American Psychologist*, *7*, 408 (Abstract).

Tutz, G. and Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics and Data Analysis*, *22*, 537–557.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, *92*, 351–370.

van Breukelen, G. J. P. (2005). Psychometric modeling of response speed and accuracy with mixed and conditional regression. *Psychometrika*, *70*, 359–376.

van den Hout, A. and Klugkist, I. (2009). Accounting for non-compliance in the analysis of randomized response data. *Australian and New Zealand Journal of Statistics*, *51*, 353–372.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*, 181–204.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*, 287–308.

van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics*, *33*, 5–20.

van der Linden, W. J. and Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.

van der Linden, W. J. and Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika*, *68*, 251–265.

van der Maas, H. L. J. and Wagenmakers, E.-J. (2005). A psychometric analysis of chess expertise. *The American Journal of Psychology*, *118*, 29–60.

van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, *10*, 1–50.

Vandenberg, R. J. and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70.

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, *90*, 614–618.

Verhelst, N. D., Glas, C. A. W., and Verstralen, H. H. F. M. (1995). *OPLM: One Parameter Logistic Model* [computer software]. Arnhem: Cito.

Verhelst, N. D. and Verstralen, H. H. F. M. (2001). An IRT model for multiple raters. In A. Boomsma, M. A. J. van Duijn, and T. A. B. Snijders (Eds.), *Essays on Item Response Theory* (pp. 89–106). New York: Springer.

Verhelst, N. D., Verstralen, H. H. F. M., and Jansen, M. G. H. (1997). A logistic model for time-limit tests. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 169–185). New York: Springer.

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, *17*, 33–51.

Wainer, H., Bradlow, E. T., and Wang, X. (2007). *Testlet Response Theory and Its Applications*. Cambridge: Cambridge University Press.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*, 63–69.

Wright, B. D. (1977). Misunderstanding the Rasch model. *Journal of Educational Measurement*, *14*, 219–225.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*, 187–213.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association*, *86*, 79–86.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Zellner, A. (1997). *Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers*. Cheltenham: Edward Elgar.

Zimowski, M. F., Muraki, E., Mislevy, R. J., and Bock, R. D. (1996). *BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items* [computer software]. Chicago, IL: Scientific Software International.

Index