

References

- Afshartous, D. (1995). Determination of sample size for multilevel model design. In V. S. Williams, L. V. Jones, & I. Olkin (Eds.), *Perspectives on statistics for educational research: Proceedings of the National Institute for Statistical Sciences* (NISS), Technical Report #35 (pp. 20–22). Triangle Park, NC: National Institute of Statistical Sciences.
- Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Series A*, *144*(4), 419–461.
- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, *149*(1), 1–43.
- Anderson, J. O., Milford, T., & Ross, S. P. (2009). Multilevel modeling with HLM: Taking a second look at PISA. In M. C. Shelley, L. D. Yore, & B. Hand (Eds.), *Quality research in literacy and science education* (pp. 263–286). Dordrecht, the Netherlands: Springer Science + Business Media B. V.
- Asparouhov, T., & Muthén, B. (2006). *Multilevel modeling of complex survey data: ASA Section on Survey Research Methods*, 2718–2726. Retrieved from <http://www.statmodel.com/download/SurveyJSM1.pdf>
- Asparouhov, T., Muthén, B., & Muthén, L. (2006). General multi-level modeling with sampling weights. *Communications in Statistics: Theory and Methods*, *35*(3), 439–460.
- Baker, D. P., Goesling, B., & Letendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the “Heyneman-Loxley effect” on mathematics and science achievement. *Comparative Education Review*, *46*(3), 291–312.
- Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Loudermilk, B. L., Kromrey, J. D., & Ferron, J. M. (2010). *Dancing the sample size limbo with mixed models: How low can you go?* (SAS Global Forum 2010 Posters Paper 197-2010). Retrieved from <http://support.sas.com/resources/papers/proceedings10/197-2010.pdf>
- Braun, H., Jenkins, F., & Grigg, W. (2006). *Comparing private schools and public schools using hierarchical linear modeling* (NCES 2006-461). Washington, DC: U.S. Government Printing Office.
- Brown, P., & Lauder, H. (1996). Education, globalization and economic development. *Journal of Education Policy*, *11*(1), 1–25.

- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*, Thousand Oaks, CA: Sage.
- Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education*, 8, 158–233.
- Caro, D. H. (2010). *Measuring family socioeconomic status: A methodological proposal for PIRLS*. Manuscript submitted for publication.
- Caro, D. H., & Lehmann, R. (2009). Achievement inequalities in Hamburg schools: How do they change as students get older? *School Effectiveness and School Improvement*, 20(4), 407–431.
- Caro, D. H., McDonald, T., & Willms, J. D. (2009). Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education*, 32(3), 558–590.
- Chantala, K., Blanchette, D., & Suchindran, C. M. (2006). *Software to compute sampling weights for multilevel analysis*. Chapel Hill, NC: Carolina Population Center, University of North Carolina. Retrieved from http://www.cpc.unc.edu/restools/data_analysis/ml_sampling_weights
- Cheong, Y. F., Fotiu, R. P., & Raudenbush, S.W. (2001). Efficiency and robustness of alternative estimators for two- and three-level models: The cases of NAEP. *Journal of Educational and Behavioral Statistics*, 26(4), 411–429.
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: Wiley.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1998). Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics*, 14(3), 267–275.
- Cronbach, L. J. (1967). How can instruction be adapted to individual differences? In R. M. Gagné (Ed.), *Learning and individual differences*. Columbus, OH: Merrill Books.
- Dale, R. (2000). Globalization and education: Demonstrating a “common world educational culture” or locating a “globally structured educational agenda”? *Educational Theory*, 50(4), 427–448.
- Decker, P. T., Rice, J. K., & Moore, M. T. (1997). *Education and the economy: An indicators report* (NCES 97-269). Washington, DC: U.S. Government Printing Office.
- De Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11(1), 57–85.
- Desimone, L. M., Smith, T., Baker, D., & Ueno, K. (2005). Assessing barriers to the reform of U.S. mathematics instruction from an international perspective. *American Educational Research Journal*, 42(3), 501–535.
- Foy, P., & Olson, J. F. (Eds.). (2009). *TIMSS 2007 international database and user guide*. Chestnut Hill, MA: Boston College.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73(1), 43–56.

- Goldstein, H. (1996). *Multilevel statistical models*. London, UK: Edward Arnold.
- Graubard, B., & Korn, E. (1996). Modeling the sampling design in the analysis of health surveys. *Statistical Methods in Medical Research*, 5(3), 263–281.
- Green, J. L., Camilli, G., & Elmore, P. B. (2006). *Handbook of complementary methods in education research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Grilli, L., & Pratesi, M. (2004). Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling design. *Survey Methodology*, 30(1), 93–103.
- Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam, the Netherlands: TT-Publikaties.
- Hox, J. J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147–154). New York, NY: Springer.
- Joncas, M. (2008). TIMSS 2007 sampling design. In J. F. Olson, M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 76–93). Chestnut Hill, MA: Boston College.
- Kish, L. (1965). *Survey sampling*, New York, NY: Wiley.
- Knapp, T. R. (1977). The unit-of-analysis problem in applications of simple correlation analysis to educational research. *Journal of Educational Statistics*, 2(3), 171–186.
- Koretz, D., McCaffrey, D., & Sullivan, T. (2001). Predicting variations in mathematics performance in four countries using TIMSS. *Education Policy Analysis Archives*, 9(34), 28 pp. Retrieved from <http://epaa.asu.edu/epaa/v9n34/>
- Korn, E. L., & Graubard, B. I. (1995). Analysis of large health surveys: Accounting for the sampling design. *Journal of the Royal Statistical Society*, 158(Series A), 263–295.
- Korn, E. L., & Graubard, B. I. (2003). Estimating variance components by using survey data. *Journal of the Royal Statistical Society*, 65(Series B), 175–190.
- Kovacevic, M. S., & Rai, S. N. (2003). A pseudo maximum likelihood approach to multilevel modeling of survey data. *Communications in Statistics, Theory and Methods*, 32, 103–121.
- Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies*. Unpublished manuscript, California State University, Los Angeles.
- Lamb, S. & Fullarton, S. (2001). *Classroom and school factors affecting mathematics achievement: A comparative study of the US and Australia using TIMSS*. Camberwell, Victoria, Australia: Australian Council for Educational, Research (ACER). Available online at http://research.acer.edu.au/timss_monographs/10.
- Lohr, S. L. (1999). *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press.
- Longford, N. T. (1993). *Random coefficient models*. Oxford, UK: Clarendon Press.
- Longford, N. T. (1996). Model-based variance estimation in surveys with stratified clustered designs. *Australian Journal of Statistics*, 38, 333–352.
- Lubienski, S. T., & Lubienski, C. (2006). School sector and academic achievement: A multi-level analysis of NAEP mathematics data. *American Educational Research Journal*, 43(4), 651–698.

- Ma, X., & McIntyre, L. J. (2005). Exploring differential effects of mathematics courses on mathematics achievement. *Canadian Journal of Education*, 28(4), 827–852.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology* 2005, 1(3), 86–92.
- Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.
- Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. Van der Linden, & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101–122). New York, NY: Springer.
- Moerbeek, M., Van Breukelen, G. J. P., & Berger, M. P. F. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*, 25(3), 271–284.
- Moerbeek, M., Van Breukelen, G. J. P., & Berger, M. P. F. (2001). Optimal experimental design for multilevel logistic models. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 50(1), 17–30.
- Mok, M. (1995). Sample size requirements for 2-level designs in educational research. *Multilevel Modelling Newsletter*, 7(2), 11–15.
- Muthén, B. O. (2008, March 3). *Mplus discussion: Adjusting sampling errors for cluster sampling*. Retrieved from <http://www.statmodel.com/discussion/messages/12/305.html>
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599–620.
- Muthén, L. K., & Muthén, B. O. (2008). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
- Okumura, T. (2007). Sample size determination for hierarchical linear models considering uncertainty in parameter estimates. *Behaviormetrika*, 34(2), 79–93.
- Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: Boston College.
- Organisation for Economic Co-operation and Development (OECD). (2006). *PISA 2006 technical report*. Paris, France: Author.
- Organisation for Economic Co-operation and Development (OECD). (2009). *PISA data analysis manual: SAS, second edition. Education and skills*. Paris, France: Author.
- Pfeffermann, D., Moura, F., & Silva, P. (2006). Multilevel modeling under informative sampling. *Biometrika*, 4, 943–959.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rabash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society, Series B. Statistical Methodology*, 60, 23–56.
- Pong, S., & Pallas, A. (2001). Class size and eighth-grade math achievement in the United States and abroad. *Educational Evaluation and Policy Analysis*, 23(3), 251–273.
- Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society*, 169(4), 805–827.

- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (expanded ed.). Chicago, IL: University of Chicago Press.
- Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics, 13*(2), 85–116.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173–185.
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1–17.
- Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2005). *Optimal design for longitudinal and multilevel research (Version 1.55)* [computer software]. Chicago, IL: University of Chicago Press.
- Reise, S. P., & Duan, N. (2003). Design issues in multilevel studies. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 285–298). Mahwah, NJ: Lawrence Erlbaum.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.
- Rogosa, D. (1978). Politics, process, and pyramids. *Journal of Educational Statistics, 3*(1), 79–86.
- Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal, 32*, 583–625.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151.
- Sarndal, C.-E., Swenson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York, NY: Springer-Verlag.
- Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347–367.
- Schulz, W. (2006, April). *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Snijders, T. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, pp. 1570–1573). Chichester, UK: Wiley.
- Snijders, T. (2006). Sampling. In A. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 159–174). New York, NY: Wiley.
- Snijders, T., & Bosker, R. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18*(3), 237–259.
- Snijders, T., & Bosker, R. (1999). *An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Stapleton, L. (2002). The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modeling*, 9(4), 475–502.

Suárez-Orozco, M. M., & Qin-Hilliard, D. (2004). *Globalization: Culture and education for a new millennium*. Berkeley, CA: University of California Press.

Thomas, L. T., & Heck, R. H. (2001). Analysis of large scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education*, 42(5), 517–540.

Van der Leeden, R., Busing, F., & Meijer, E. (1997). *Applications of bootstrap methods for two-level models*. Paper presented at the Multilevel Conference, Amsterdam, the Netherlands.

Von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? *IERI Monograph Series (2): Issues and methodologies in large-scale assessments* (pp. 9–36). Hamburg, Germany: IEA-ETS Research Institute.

Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis*, 20(3), 137–156.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12). Retrieved from <http://epaa.asu.edu/epaa/v10n12/>

Willms, J. D. (2003). *Ten hypotheses about socioeconomic gradients and community differences in children's developmental outcomes*. Ottawa, Ontario, Canada: Applied Research Branch, Human Resources Development Canada.

Willms, J. D., & Shields, M. (1996). *A measure of socioeconomic status for the national longitudinal study of children*. Draft prepared as a reference for researchers conducting analysis of the first wave of data from the Canadian National Longitudinal Study of Children and Youth (NLSCY).

Wu, M. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27.

Zaccarin, S., & Donati, C. (2008). *The effects of sampling weights in multilevel analysis of PISA data* (working paper no. 119). Trieste, Italy: University of Trieste. Retrieved from http://www2.units.it/nirdses/sito_inglese/working%20papers/files%20for%20wp/wp119.pdf