



UvA-DARE (Digital Academic Repository)

Cluster bias: Testing measurement invariance in multilevel data

Jak, S.

Publication date
2013

[Link to publication](#)

Citation for published version (APA):

Jak, S. (2013). *Cluster bias: Testing measurement invariance in multilevel data*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

REFERENCES

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of educational measurement, 29*, 67-91.
- Asparouhov, T., & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. *Proceedings of the 2007 Joint Statistical Meetings, Section on Statistics in Epidemiology* (pp. 2531–2535). Alexandria, VA: American Statistical Association.
- Barendse, M. T., Oort, F. J. & Garst, G. J. A. (2010). Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias; a simulation study. *Advances in Statistical Analysis, 94*, 117–127.
- Barendse, M. T., Oort, F. J., Werner, C. S., Ligtoet, R. & Schermelleh-Engel, K. (2012). Measurement bias detection through factor analysis. *Structural Equation Modeling, 19*, 561-579.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley, New York.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological review, 111*, 1061-1071.
- Browne, M. W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230-258.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.
- Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing, 10*, 107-132.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: A comparison of four approaches. *Multivariate Behavioral Research, 47*, 840-876.
- Christoffersson, A. (1975). Factor Analysis of Dichotomized Variables. *Psychometrika, 40*, 5-32.
- De Jong, M. G., Steenkamp, J. B. E., & Fox, J. P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT models. *Journal of consumer research, 34*, 260-278.

- Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *47*, 309-326.
- Dolan, C. V., Roorda, W. & Wicherts, J. M. (2004). Two failures of Spearman's hypothesis: The GATB in Holland and the JAT in South Africa. *Intelligence*, *32*, 155-173.
- Dudgeon, P. (2003). NIESEM: A computer program for calculating noncentral interval estimates (and power analysis) for structural equation modeling [Computer software].
- Duncan, T. E., Alpert, A., & Duncan, S. C. (1998). Multilevel covariance structure analysis of sibling antisocial behavior. *Structural Equation Modeling*, *5*, 211-228.
- Elffers, L. (2012). One foot out the school door? Interpreting the risk for dropout upon the transition to post-secondary vocational education. *British Journal of Sociology of Education*, *33*, 41-61.
- Engle, R. F. (1983). Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In Intriligator, M. D.; and Griliches, Z.. *Handbook of Econometrics*. Elsevier. pp. 796–801.
- Flora, D. B. & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491.
- Fox, J.-P., & Verhagen, A. J. (2010). Random item effects modeling for cross-national survey data. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 467–488). London: Routledge Academic.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. mvtnorm: Multivariate normal and t distributions. 2012. URL <http://CRAN.R-project.org/package=mvtnorm>. R package version 0.9-9992.
- Goldstein, H. (1995). *Multilevel statistical models*. New York: Halstead Press.
- Grilli, L., & Rampichini, C. (2007). Multilevel factor models for ordinal variables. *Structural Equation Modeling*, *14*, 1-25.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. *Test validity*, 129-145
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental aging research*, *18*, 117-144.
- Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum.
- Hox, J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64*, 157-170.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

- Jackson, S. E., & Joshi, A. (2004). Diversity in social context: a multi-attribute, multilevel analysis of team diversity and sales performance. *Journal of Organizational Behavior*, *25*, 675-702.
- Jak, S. & Oort, F.J. (under review). On the power of the test for cluster bias.
- Jak, S., Oort, F.J. & Dolan, C.V. (2010). Measurement bias and multidimensionality; an illustration of bias detection in multidimensional measurement models. *Advances in Statistical Analysis*, *94*, 129-137.
- Jak, S., Oort, F.J. & Dolan, C.V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling*, *20*, 265-282.
- Jak, S., Oort, F.J. & Dolan, C.V. (in press). Measurement bias in multilevel data. *Structural Equation Modeling*.
- Jak, S., Oort, F.J. & Dolan, C.V. (under review). Using two-level ordinal factor analysis to test for cluster bias in ordinal data.
- Jones, A. (2004). Social anxiety, sex, surveillance and the 'safe' teacher. *British Journal of Sociology of Education*, *25*, 53-66.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.
- Jöreskog, K. G. & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*, 347-387.
- King-Kallimanis, B. L., Oort, F. J., & Garst, G. J. A. (2010). Using structural equation modelling to detect measurement bias and response shift in longitudinal data. *ASTA Advances in Statistical Analysis*, *94*, 139-156.
- Koman, E. S., & Wolff, S. B. (2008). Emotional intelligence competencies in the team and team leader: A multi-level examination of the impact of emotional intelligence on team performance. *Journal of Management Development*, *27*, 55-75.
- Koomen, H. M. Y., Verschueren, K., & Pianta, R. C. (2007). *Leerling-Leerkracht Relatie Vragenlijst (LLRV): Handleiding*. [Student-Teacher Relationship Scale: Manual.] Houten, The Netherlands: Bohn Stafleu van Loghum.
- Koomen, H. M., Verschueren, K., van Schooten, E., Jak, S., & Pianta, R. C. (2012). Validating the Student-Teacher Relationship Scale: Testing factor structure and measurement invariance across child gender and age in a Dutch sample. *Journal of School Psychology*, *50*, 215-234.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53-76.
- Longford, N. T. (1993). *Random coefficient models*. Oxford: Clarendon Press.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. London: Addison-Wesley.

- Lubke, G. H., Dolan, C. V., Kelderman, H. & Mellenbergh, G. J. (2003). Weak measurement invariance with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology*, *56*, 231–248.
- Marsh, H. W., & Hocevar, D. (1984). The factorial invariance of students' evaluations of college teaching. *American Educational Research Journal*, *21*, 341-366.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Statistics*, *13*, 127-143.
- Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, *29*, 223-236.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525-543.
- Meredith, W. & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*, 69-77.
- Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparison using latent means. *Multivariate Behavioral Research*, *26*, 479-497.
- Millsap, R. E., & Yun - Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*, 479-515.
- Molenaar, D., Dolan, C. V., & de Boeck, P. (2012). The heteroscedastic graded response model with a skewed latent trait: Testing statistical and substantive hypotheses related to skewed item category functions. *Psychometrika*, *77*, 455-478.
- Molenaar, D., Dolan, C. V., & Verhelst, N. D. (2010). Testing and modelling non-normality within the one-factor model. *British Journal of Mathematical and Statistical Psychology*, *63*, 293-317.
- Molenaar, D., Dolan, C. V., Wicherts, J. M. & van der Maas, H. L. J. (2010). Modeling Differentiation of Cognitive Abilities within the Higher-Order Factor Model using Moderated Factor Analysis. *Intelligence*, *38*, 611-624.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557-585.
- Muthén, B. (1990). *Mean and covariance structure analysis of hierarchical data*. Los Angeles, CA: UCLA statistics series, NO. 62.
- Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, *22*, 376-398.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431-462.

- Muthén, B., Khoo, S.T. & Gustafsson, J.E. (1997). Multilevel latent variable modeling in multiple populations. Unpublished technical report. Retrieved from www.statmodel.com/papers.shtml, March 2, 2011.
- Muthén, L. K., & Muthén, B. O. (2007). *Mplus Users Guide*. Fifth Edition. Los Angeles, CA: Muthén & Muthén.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443-460.
- Oort, F. J. (1991). Theory of violators: assessing unidimensionality of psychological measures. In: Steyer, R., Wender, K.F., Widaman, K.F. (eds.) *Psychometric Methodology*, pp. 377–381. Stuttgart: Fischer.
- Oort, F.J. (1992). Using restricted factor analysis to detect item bias. *Methodika*, *6*, 150-166.
- Oort, F.J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling*, *5*, 107-124.
- Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, *14*, 587-598.
- Pianta, R. C. (2001). *Student-Teacher Relationship Scale: Professional Manual*. Lutz, FL: Psychological Assessment Resources.
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for multilevel mediation, *Psychological Methods*, *15*, 209-233.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. [Computer software]. Vienna, Austria.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, *69*, 167-190.
- Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, *19*, 337–350.
- Reise, S. P., Widaman, K. F. & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552–566.
- Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modelling. *Structural equation modeling*, *16*, 583-601.
- Satorra, A., & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*, 243-248.
- Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modelling in cross-cultural research: Multigroup and multilevel approaches. In F. J. R. Van de Vijver, D. A. van Hemert, & Y. Poortinga (Eds.), *Individuals and cultures in Multi-level analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Smits, I. A., Dolan, C. V., Vorst, H. C., Wicherts, J. M., & Timmerman M. E. (2011). Cohort differences in Big Five personality factors over a period of 25 years. *Journal of Personality and Social Psychology*, *100*(6), 1124-1138.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229-239.
- Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371-384.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72-101.
- Spearman, C. (1928). The sub-structure of the mind. *British Journal of Psychology*, *18*, 249-261.
- Spilt, J., Koomen, H.M.Y & Jak, S. (2012). Are boys better off with male and girls with female teachers? A multilevel investigation of measurement invariance and gender match in teacher-student relationship quality. *Journal of School Psychology*, *50*, 363 - 378.
- Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Stoel, R. D., Garre, F.G., Dolan, C. V., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, *11*, 439-455.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational measurement*, *27*, 361-370.
- Thoonen, E. E. J., Slegers, P. J. C., Peetsma, T. T. D., & Oort, F. J. (2011). Can teachers motivate students to learn? *Educational Studies*, *37*, 345-360.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *2*, 4-69.
- Verhagen, A. J. & Fox, J.-P. (2012). Bayesian Tests of Measurement Invariance. *British Journal of Mathematical and Statistical Psychology*. DOI: 10.1111/j.2044-8317.2012.02059.
- Voorpostel, M., & Blieszner, R. (2008). Intergenerational solidarity and support between adult siblings. *Journal of Marriage and Family*, *70*, 157-167.
- Webb, M. L., & Neuharth-Pritchett, S. (2010). Examining factorial validity and measurement invariance of the Student-Teacher Relationships Scale. *Early Childhood Research Quarterly*, *26*, 205-215.

- Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., & Zhou, X. (2012). Gender differences in children's arithmetic performance are accounted for by gender differences in language abilities. *Psychological science, 23*, 320-330.
- Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., Baal, G. C. M. van, Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32*, 509-537.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.
- Woods, C. M. (2009). Evaluation of MIMIC-model methods to DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research, 44*, 1-27.
- Yuan, K. H. & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M.E. Sobel & M.P. Becker (eds.), *Sociological Methodology 2000*. Washington, D.C.: ASA.