

Sample Size Requirements in HLM: An Empirical Study

The Relationship Between the Sample Sizes at Each Level of a Hierarchical Model and the Precision of the Outcome Model

Sabine Meinck and Caroline Vandenplas

IEA Data Processing and Research Center, Hamburg, Germany

This report was funded by the National Center for Education Statistics under Contract No. ED-08-CO-0117 with the International Association for the Evaluation of Educational Achievement (IEA). Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

IERI Monograph Series
Issues and Methodologies in Large-Scale Assessments

Special Issue 1

Sample Size Requirements in HLM: An Empirical Study

The Relationship Between the Sample Sizes at Each Level of a Hierarchical Model and the Precision of the Outcome Model

Sabine Meinck and Caroline Vandenplas

TABLE OF CONTENTS

Foreword	3
List of Tables and Figures	9
Acknowledgements	16
Abstract	17
1. Introduction	19
2. Literature Review	23
2.1 The Concept of Hierarchical Models and Their Use in Educational Research	23
2.2 Precision of the Estimates in Multilevel Models for Complex Sample Survey Data	26
2.3 Sample Size Requirements and HLM—Knowledge at Hand	28
2.4 Cost Implications of Sample Size Considerations	31
3. Research Questions	33
4. Data and Methods	35
4.1 Fixed Population and Sample Parameters	35
4.1.1 Number of replicates	36
4.1.2 Achievement variable scale and socioeconomic status indicator	36
4.1.3 Covariance between the SES indicator and the achievement variable	37
4.1.4 Within- and between-schools variance of the SES indicator	37
4.2 Varied Population and Sample Parameters	37
4.2.1 Sample size of clusters and within clusters	37
4.2.2 Intraclass correlation coefficients (ICCs)	38
4.2.3 Distribution of covariance between the SES indicator and the achievement variable between the hierarchical levels	38
4.2.4 Weights	39

4.3 Hierarchical Models	39
4.4 Outcomes	42
4.4.1 Coefficients of variation	42
4.4.2 Curve estimation and equations	45
4.5 Software Used	47
5. Results and Discussion	49
5.1 Coefficients of Variation of the Different Models' Parameters	49
5.1.1 Residual variance	50
5.1.2 Mean of random intercepts	51
5.1.3 Variance of random intercepts	54
5.1.4 Fixed slope	56
5.1.5 Slope of random intercepts	58
5.1.6 Mean of random slopes	60
5.1.7 Variance of random slopes	61
5.2 Effects of Variable Population and Sample Parameters	62
5.2.1 Sample sizes	62
5.2.2 Intraclass correlation coefficients	64
5.2.3 Covariance distributions	65
5.2.4 Weights	66
5.2.5 Model complexity	67
5.3 Practical Use of Outcome Equations	70
6. Limitations and Need for Further Research	75
7. Summary and Conclusions	77
Appendix	81
References	167

LIST OF TABLES AND FIGURES

Main Text

Figure 4.1: Distribution of design weights of schools (over all participating countries, after z-transformation at country level)	40
Figure 4.2: Exemplary comparison of two methods of SE estimation: Model 1, SE of residual variance, mean over four ICC levels	43
Figure 4.3: Exemplary comparison of two methods of SE estimation: Model 3, SE of slope of random intercepts, mean over two cases of covariance distribution and four ICC levels	43
Figure 4.4: Comparison of two methods of SE estimation: Model 4, SE of variance of random slope, mean over two cases of covariance distribution and four ICC levels	44
Figure 4.5: Residual variance, its SE and the CV (%) (y-axes) by cluster size (x-axes): Model 1, average over all sampling scenarios	45
Figure 4.6: Estimated variance of the school-level variance component by within-cluster sample size	46
Table 5.1: Explored model parameters and their model allocation	49
Figure 5.1: CV (%) of the residual variance by weight status and sample size at both levels	50
Figure 5.2: CV (%) of the mean of random intercepts by weight status, ICC, and sample size at both levels: Models 1 and 2	51
Figure 5.3: CV (%) of the mean of random intercepts by covariance distribution, ICC, and sample size at both levels: Models 3 and 4, unweighted data	52
Figure 5.4: CV (%) of the variance of random intercepts by weight status, ICC, and sample size at both levels: Model 1	54
Figure 5.5: CV (%) of the variance of random intercepts by ICC, model, and sample size at both levels: Covariance Distribution Case 1 (20 at within and 10 at between level), unweighted data	55
Figure 5.6: CV (%) of the variance of intercepts by ICC, model, and sample size at both levels: Covariance Distribution Case 2 (10 at within and 20 at between level), unweighted data	56
Figure 5.7: CV (%) of the fixed slope by model, covariance distribution, and sample size at both levels: unweighted data	57
Figure 5.8: CV (%) of the slope of random intercepts by ICC, covariance distribution, and sample size at both levels: Model 3, unweighted data	58
Figure 5.9: CV (%) of the slope of random intercepts by ICC, covariance distribution, and sample size at both levels: Model 4, unweighted data	59
Figure 5.10: CV (%) of the mean of random slopes by covariance distribution, ICC, and sample size at both levels: Model 4, unweighted data	60
Figure 5.11: Variance of random slopes (means over all replicates) by covariance distribution, ICC, and sample size at both levels: Model 4	61

Figure 5.12: CV (%) of the fixed slope by covariance distribution, ICC, and total sample size, averaged over Models 2 and 3 and weight status	63
Figure 5.13: CV (%) of the fixed slope by covariance distribution, ICC, and sample size at both levels, averaged over Models 2 and 3 and weight status	64
Figure 5.14: CV (%) of the mean of random intercepts by ICC, model, and cluster size: average over weight status and numbers of sampled clusters	68
Figure 5.15: CV (%) of the variance of random intercepts by ICC, model, and cluster size: average over weight status and numbers of sampled clusters	68
Figure 5.16: CV (%) of the fixed slope by ICC, model, and cluster size: average over weight status and numbers of sampled clusters	69
Figure 5.17: CV (%) of the slope of random intercepts by ICC, model, and cluster size: average over weight status and numbers of sampled clusters	70
Figure 5.18: Excerpt from Appendix Table 17	72

Appendix

Table A1: Overview of sampling scenarios explored	82
Table A2: Results of curve estimation for the CV (%) of the residual variance: quadratic equations (model summary and parameter estimates, average over all models and both covariance cases; the independent variable is cluster size)	84
Figure A1: CV (%) of the residual variance: average over all models and both covariance cases (graphical representation of Table A2)	85
Table A3: Results of curve estimation for the CV (%) of the residual variance: quadratic equations (model summary and parameter estimates, average over all models and both covariance cases; the independent variable is number of sampled clusters)	86
Figure A2: CV (%) of the residual variance: average over all models and both covariance cases (graphical representation of Table A3)	87
Table A4: Results of curve estimation for the CV (%) of the mean of random intercepts: quadratic equations (model summary and parameter estimates; average over Models 1 and 2 and both covariance cases; the independent variable is cluster size)	88
Figure A3: CV (%) of the mean of random intercepts: average over Models 1 and 2 and both covariance cases (graphical representation of Table A4)	89
Table A5: Results of curve estimation for the CV (%) of the mean of random intercepts: quadratic equations (model summary and parameter estimates; average over Models 1 and 2 and both covariance cases; the independent variable is number of sampled clusters)	90
Figure A4: CV (%) of the mean of random intercepts: average over Models 1 and 2 and both covariance cases (graphical representation of Table A5)	91
Table A6: Results of curve estimation for the CV (%) of the mean of random intercepts: quadratic equations (model summary and parameter estimates, average over Models 3 and 4; the independent variable is cluster size)	92

Figure A5: CV (%) of the mean of random intercepts: average over Models 3 and 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A6)	93
Figure A6: CV (%) of the mean of random intercepts: average over Models 3 and 4, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A6)	94
Table A7: Results of curve estimation for the CV (%) of the mean of random intercepts: quadratic equations (model summary and parameter estimates, average over Models 3 and 4; the independent variable is number of sampled clusters)	95
Figure A7: CV (%) of the mean of random intercepts: average over Models 3 and 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A7)	97
Figure A8: CV (%) of the mean of random intercepts: average over Models 3 and 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of second part of Table A7)	98
Table A8: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 1; the independent variable is cluster size)	99
Figure A9: CV (%) of the variance of random intercepts: Model 1 (graphical representation of Table A8)	100
Table A9: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 1; the independent variable is number of sampled clusters)	101
Figure A10: CV (%) of the variance of random intercepts: Model 1 (graphical representation of Table A9)	102
Table A10: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 2; the independent variable is cluster size)	103
Figure A11: CV (%) of the variance of random intercepts: Model 2, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A10)	105
Figure A12: CV (%) of the variance of random intercepts: Model 2, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A10)	106
Table A11: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 2; the independent variable is number of sampled clusters)	107
Figure A13: CV (%) of the variance of random intercepts: Model 2, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A11)	109

Figure A14: CV (%) of the variance of random intercepts: Model 2, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A11)	110
Table A12: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 3; the independent variable is cluster size)	111
Figure A15: CV (%) of the variance of random intercepts: Model 3, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A12)	113
Figure A16: CV (%) of the variance of random intercepts: Model 3, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A12)	114
Table A13: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 3; the independent variable is number of sampled clusters)	115
Figure A17: CV (%) of the variance of random intercepts: Model 3, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A13)	117
Figure A18: CV (%) of the variance of random intercepts: Model 3, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A13)	118
Table A14: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 4; the independent variable is cluster size)	119
Figure A19: CV (%) of the variance of random intercepts: Model 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A14)	121
Figure A20: CV (%) ratio of the variance of random intercepts: Model 4, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A14)	122
Table A15: Results of curve estimation for the CV (%) of the variance of random intercepts: quadratic equations (model summary and parameter estimates, Model 4; the independent variable is number of sampled clusters)	123
Figure A21: CV (%) of the variance of random intercepts: Model 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A15)	125
Figure A22: CV (%) of the variance of random intercepts: Model 4, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A15)	126
Table A16: Results of curve estimation for the CV (%) of the fixed slope: quadratic equations (model summary and parameter estimates, Model 2; the independent variable is cluster size)	127

Figure A23: CV (%) of the fixed slope: Model 2, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A16)	128
Figure A24: CV (%) of the fixed slope: Model 2, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A16)	129
Table A17: Results of curve estimation for the CV (%) of the fixed slope: quadratic equations (model summary and parameter estimates, Model 2; the independent variable is number of sampled clusters)	130
Figure A25: CV (%) of the fixed slope: Model 2, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A17)	132
Figure A26: CV (%) of the fixed slope: Model 2, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A17)	133
Table A18: Results of curve estimation for the CV (%) of the fixed slope: quadratic equations (model summary and parameter estimates, Model 3; the independent variable is cluster size)	134
Figure A27: CV (%) of the fixed slope: Model 3, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A18)	136
Figure A28: CV (%) of the fixed slope: Model 3, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A18)	137
Table A19: Results of curve estimation for the CV (%) of the fixed slope: quadratic equations (model summary and parameter estimates, Model 3; the independent variable is number of sampled clusters)	138
Figure A29: CV (%) of the fixed slope: Model 3, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A19)	140
Figure A30: CV (%) of the fixed slope: Model 3, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A19)	141
Table A20: Results of curve estimation for the CV (%) of the slope of random intercepts: quadratic equations (model summary and parameter estimates, Model 3; the independent variable is cluster size)	142
Figure A31: CV (%) of the slope of random intercepts: Model 3, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A20)	144
Figure A32: CV (%) of the slope of random intercepts: Model 3, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A20)	145

Table A21: Results of curve estimation for the CV (%) of the slope of random intercepts: quadratic equations (model summary and parameter estimates, Model 3; the independent variable is number of sampled clusters)	146
Figure A33: CV (%) of the slope of random intercepts: Model 3, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A21)	148
Figure A34: CV (%) of the slope of random intercepts: Model 3, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A21)	149
Table A22: Results of curve estimation for the CV (%) of the slope of random intercepts: quadratic equations (model summary and parameter estimates, Model 4; the independent variable is cluster size)	150
Figure A35: CV (%) of the slope of random intercepts: Model 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A22)	152
Figure A36: CV (%) of the slope of random intercepts: Model 4, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A22)	153
Table A23: Results of curve estimation for the CV (%) of the slope of random intercepts: quadratic equations (model summary and parameter estimates, Model 4; the independent variable is number of sampled clusters)	154
Figure A37: CV (%) of the slope of random intercepts: Model 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A23)	156
Figure A38: CV (%) of the slope of random intercepts: Model 4, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A23)	157
Table A24: Results of curve estimation for the CV (%) of the mean of random slopes: quadratic equations (model summary and parameter estimates, Model 4; the independent variable is cluster size)	158
Figure A39: CV (%) of the mean of random slopes: Model 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A24)	160
Figure A40: CV (%) of the mean of random slopes: Model 4, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A24)	161
Table A25: Results of curve estimation for the CV (%) of the mean of random slopes: quadratic equations (model summary and parameter estimates, Model 4; the independent variable is number of sampled clusters)	162
Figure A41: CV (%) of the mean of random slopes: Model 4, Covariance Case 1 (20 at within and 10 at between level; graphical representation of first part of Table A25)	164

Figure A42: CV (%) of the mean of random slopes: Model 4, Covariance Case 2 (10 at within and 20 at between level; graphical representation of second part of Table A25)	165
--	-----

Acknowledgements

The authors would like to express their warmest thanks to several people who helped them to complete this project. We thank Dirk Hastedt for encouraging us to take on this study. We thank Heiko Sibberns, Leslie Rutkowski, Wolfram Schulz, Eugenio Gonzalez, and the anonymous reviewer for their thorough evaluation of the paper and their valuable feedback. We are also grateful for the help of our colleagues in the Sampling Unit and the Research and Analysis Unit at the IEA Data Processing and Research Center, who helped us with various details throughout the study. Without the generous help of these individuals, this investigation would not have been possible. Finally, we thank the National Center for Education Statistics (NCES), U.S. Department of Education, Institute of Education Sciences, for funding this research project.

Abstract

This study focused on the properties of data collected in large-scale assessments (LSA) in order to explore the relationships between sample sizes at different levels of clustered data and the sampling precision of the results derived from hierarchical linear models (HLM). A Monte Carlo simulation study was used in order to explore various population and sample conditions. The varied conditions were sample sizes of and within clusters, intraclass correlation coefficients, covariance distribution, use of sampling weights, and model complexity. As expected, the precision of all explored parameters increased as sample sizes increased. The dependency took a nonlinear format—a general observation that held true for all settings. The magnitude of the increase, and whether the effect became more pronounced as sample size increased on either of the hierarchical levels, could depend, however, on all explored sample and population conditions and could also vary across the different model parameters. In conclusion, the results showed that required sample sizes depend heavily on the parameter of interest. In particular, sampling precision differed widely for fixed model parameters versus variance estimates. For certain model parameters, the effect of how the covariance was distributed between the hierarchical levels appeared to be even more pronounced than the effect of varying sample sizes. The inclusion of sampling weights in the model decreased the sampling precision of all explored parameters consistently by approximately 10%. The model complexity had an influence on the sampling precision of all observed parameters except the residual variance. The influence thus varied according to the parameter of interest as well as the considered case of covariance distribution.

1. Introduction

Beyond controversy is the premise that education is an important factor influencing the development of national economies worldwide (Brown & Lauder, 1996; Decker, Rice, & Moore, 1997). National assessments exploring the quality and outcomes of education systems have consequently become popular in recent decades, while accretive levels of globalization have led to education increasingly being viewed from within a broader context (Dale, 2000; Suárez-Orozco & Qin-Hilliard, 2004). These developments have heightened interest in international comparative studies of education, many of which include large-scale assessments (LSA). The increasing number of educational surveys conducted by the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD) are evidence of this growing interest.¹

When analyzing data collected in large-scale educational surveys, researchers still tend to use (or to suggest the use of) simple linear regression models (Foy & Olson, 2009; Olson, Martin, & Mullis, 2008). While the application of these models is appropriate for certain types of analyses or data structures, limitations regarding their usefulness become apparent when the data have a nested structure, that is, follow specific hierarchies (Aitkin, Anderson, & Hinde, 1981; Robinson, 1950). Simple linear regression models do not consider the effects of multiple factors on different levels of the hierarchy or on their interactions. These limitations can be avoided by using hierarchical linear modeling (HLM) (e.g., Bryk & Raudenbush, 1992; Hox, 1995; Snijders & Bosker, 1999). HLM takes the multilevel structure of a comparison problem into account and allows predictors to be introduced at different levels, thereby making it possible to study the effect of the variables at the specific level in which they occur.

HLM is usually excellently suited for analyzing data collected in educational surveys. The education systems with students embedded in classes, classes embedded in schools, schools in districts, and districts in countries display the data structure for which HLM techniques were developed. In addition, general sampling strategies in international LSA generally imply the same hierarchical approach (see, for example, Martin, Mullis, & Kennedy, 2007; Olson et al., 2008).

1 <http://www.iea.nl>; <http://www.oecd.org/edu>

The first stage of the approach involves, in each participating country, selecting a sample of schools and stratifying them according to certain organizational criteria (e.g., public versus private, or regions comprising different strata). The second stage sees classes and/or students sampled from within each participating school. The hierarchical data structure also opens a window into broadly defined concepts of student achievement in relation to some correlates of learning, such as socioeconomic (SES) background and school resources.

Given these advantages, it is not surprising that more and more researchers want to employ HLM analysis in this field of research. However, this desire has to be taken into account when developing the general study design of an educational assessment. Researchers need to be aware at this time of an important problem associated with designing studies suitable for multilevel modeling, namely the required sample sizes at the different levels of the hierarchy (see, for example, Maas & Hox, 2005; Scherbaum & Ferreter, 2009; Snijders & Bosker, 1999).

In recent years, a number of researchers have tried to address the problem by conducting (mostly) simulation studies with certain conditions in order to produce rules of thumb or even software that enable users to determine the optimal survey design. However, the literature available on the subject tends to be highly technical, hard to apply, and not easily procured. Most importantly, existing simulation studies are based on assumptions that do not fully apply to data collected in educational LSA, either because they fail to or only partially address the features typical of these datasets.

But what are the characteristics of typical LSA survey designs? In general, minimum sample sizes in LSA are predetermined by multiple factors, such as the requested precision of population estimates, the number of items and the item rotation design (connected to the need to have minimum response numbers per item), minimum cell assignments in cross tables, and so on. For example, most IEA surveys specify a minimum sample size of 150 schools to ensure that certain precision requirements are met. To give another example, the item rotation design applied in studies such as TIMSS² calls for a sample size of at least 4,000 tested students per education system because each tested student takes only one-seventh of the whole assessment (Olson et al., 2008). In this second case, the total student sample size is dictated by the item rotation design while the total cluster (school) sample size is dictated by the precision requirements and the design effect. Furthermore, cluster sampling of classes often dictates within-cluster sample sizes of about 20 to 30 individuals per cluster.

In addition, data originating from complex surveys carry weights that reflect the multiple selection probabilities of each unit, adjusted for non-response. Although general sampling designs usually aim for self-weighted samples (e.g., Joncas, 2008),³ estimation weights always vary due to stratification, practical constraints associated with implementation of the ideal sampling design, and non-response adjustments,

2 Trends in International Mathematics and Science Study, conducted by IEA: <http://timss.bc.edu/>

3 Samples that lead to equal selection probabilities of the units of interest are called self-weighted samples.

a situation that can lead to increased sampling variance. Since the development of multilevel analysis techniques, the need to consider sampling weights when engaged in multilevel modeling, as well as the influence of that modeling on estimates, has attracted attention (albeit limited) in the literature (see, for example, Asparouhov, Muthén, & Muthén, 2006; Chantala, Blanchette, & Suchindran, 2006; Korn & Graubard, 1995; Pfeffermann, Skinner, Holmes, Goldstein, & Rabash, 1998; Rabe-Hesketh & Skrondal, 2006; Stapleton, 2002; Zaccarin & Donati, 2008). However, no mention seems to have been made in this body of work of relationships between sampling weights, the statistical precision of the models, and required sample sizes.

All these constraints suggest the desirability of an evaluation of the sample sizes required to achieve a predetermined level of precision when applying multilevel modeling oriented toward the specific structure of data collected from educational large-scale assessments. Our aim, therefore, in this paper is to extend knowledge about the association between sample sizes and precision of the estimates under varying population and sample conditions and relative to model complexity.