Kansei Engineering Vol.7 No.1 pp.87-98 (2007)

ORIGINAL ARTICLES

DOES NUMBER OF LEVEL-2 UNITS IN MULTILEVEL STRUCTURAL EQUATION MODELING MATTER?

Ren-Hau Li*, Min-Ning Yu and Yueh-Luen Hu

* Department of Education, National Chengchi University, Taipei, Taiwan, R.O.C.

Abstract: How to determine the number of level-2 units in multilevel structural equation modeling (MSEM) as a standard applied to nested or hierarchical data structure was still unknown. This research used Canada data in the large database "Programme for International Student Assessment 2003" (PISA 2003) to check the model-fit indexes and parameters stability in our proposed empirical example processed by MSEM under different numbers of level-2 units. Our proposed example model was first be handled to fit Canada data (26884 students, 948 schools), and then the stabilities of the estimated parameters in the example model under 120, 240, 360, 480, 600, 720, 840 level-2 units were compared. Level-1 units in each school less than 10 students will be crossed out in advance. Besides, intraclass correlations of all variables were controlled in a specified range in different numbers of level-2 units. Finally, we found the ratio of the number of level-2 units relative to the number of estimated parameters of between-level in the multilevel model were 8:1.

Keywords: *Multilevel structural equation modeling, Intraclass correlation, PISA 2003*

1. INTRODUCTION

"Multilevel" is an important concept in survey data collection and analyses. When research data are collected from hierarchical sampling design, or when nested data structure are obtained due to cluster sampling or multistage sampling, traditional statistical analysis methodology would be improper for these data [1, 2, 3]. This kind of data derived from clustered or hierarchical sampling designs should be better analyzed by the statistical methods considering data property with clustered, hierarchical or multilevel characteristics. When multilevel characteristics of data are dealt with traditional statistical analysis, the chi-square test of model fit is often inflated, particularly for data with large intraclass correlation (ICC), large group sizes, and highly correlated variables; therefore better fit statistics can not be provided [4, 5, 6, 7].

In this study, we would focus on the number of level-2 units in multilevel analysis. When number of between-level groups gradually increased, the inadmissible solutions gradually decreased [8]. Although more level-2 units could be beneficial to obtain admissible solutions and to reduce biases of estimates and standard errors [8], there were no guidelines for us to follow. As a matter of fact, even the appropriate sample size in the traditional structural equation modeling analysis thus far has been inconclusive. An exhaustive examination of the effects on structural equation modeling based on maximum likelihood estimator by Monte Carlo simulation showed that samples fewer than 100 subjects were destructive to ML estimator and larger than 200 subjects were suggested [9]. Tanaka pointed out that there was some agreement on sample-size appropriateness by considering the ratio of the number of subjects to the number of parameters estimated in structural equation modeling with latent variables[10]. Although he did not offer a suggestion about the ratio, he actually explained why the transformation from concerning the ratio of the number of subjects to the number of variables in multiple regression analysis to concerning the ratio of the number of subjects to the number of parameters estimated. Kline indicated that although no absolute standards in the literature of structural equation modeling were offered on the ratio, he suggested the ratio 20:1 be a desirable goal and the ratio of 10:1 be a more realistic target [11].

In regard with the number of level-2 units in multilevel structural equation modeling, in general, though no conclusive suggestion is followed, a larger sample size was usually recommended and preferred [12] that ICC calculated from a small number of groups might not produce reliable estimates and it would be most useful when calculated based on beyond 30 groups. In addition, some studies suggested 50 to 100 groups with at least two individuals nested within each group for multilevel covariance structure modeling [3, 4], but the complexity of model was not taken into account in their suggestions. Hence, even groups less than 50 may be enough to get a good model fit. For example, in a study [6] where multilevel confirmatory factor analysis model was used to extract one factor from four measurement items on motivation, and a surprising good model fit of group-level structure based on only 39 groups. However, in another study where two factors with six measurement indexes were modeled in within-level model and only one factor in between-level model, the finding showed that inadmissible estimate problem occurred in the between-level

Received 2007.05.22
Accepted 2007.07.30

model when group-level sample size was small (50 groups in his research) and ICC was low [8]. Hence, A conclusion was made that the group-level sample size at least 100 would be a better way to deal data with unbalanced groups under Muthén's pseudo-balanced solution [8].

In sum, it seemed that the number of level-2 groups at least larger than 30 and the number that did not result in failure in iteration procedures were necessary conditions for multilevel data analysis, but the complexity of model was still not be taken into account to guide reasonable number of level-2 units. Maybe larger level-2 units were good suggestion, nonetheless, just as the chi-square value would be inflated in conventional structural equation model when the number of subjects was too large. Rationally, too large number of level-2 units might also result in inflated chi-square value. Therefore, to formulate an easy regulation for researchers to execute sampling work was an important and meaningful event.

Before checking the number of level-2 units, we had to decide a multilevel model. We adopted "the direct consensus model" based on variable type [15] to extend the same relationships from within-level structure to between-level structure. This was consistent to "homologous multilevel model" [16], which meant that both constructs and functional relations of these constructs in different level were identical. The structure part of the multilevel model was shown in Figure 1, symbol "TEACHSU" represented teacher support perceived from students, "INTMAT" represented interest and enjoyment for learning mathematics, "MATHEFF" represented mathematics self-efficacy, "INSTMOT" represented instrumental motivation to learn mathematics, and "MATH" represented mathematics performance.

2. METHOD

Subjects

All subjects in this study were obtained from Canada in PISA 2003 [17]. Totally, 26,884 15-year-old students from 948 schools were used in multilevel analysis. We deleted some schools with students fewer than 10 in advance to avoid some outlier cases. Hence, there were at least 11 students in each retrieved school. The reason why we chose Canada data was there were enough level-2 units to process between-level structural equation modeling.

We used all 26,884 subjects from 948 schools for multilevel structural equation modeling analysis to fit the proposed model. Next, we randomly sampled seven different samples from 948 schools with replacement. These seven samples were of different number of schools as 120, 240, 360, 480, 600, 720, 840 schools with 3358, 6959, 10542, 13440, 17160, 20583, and 23900 subjects, respectively.

Instrument

The measurement indicators were retrieved from student questionnaires in PISA 2003 database. We used all items involving five main factors in my proposed structural equation models, and then exhibited their intraclass correlation (ICC) as shown in Table 1. All analyses were handled with statistical software Mplus 4.0 [18].

3. RESULTS

The multilevel structural equation modeling analysis was processed simultaneously based on S_{PW} and Σ_B matrixes in respective level. The important parameter estimates were presented in Figure 2. Note that all parameter estimates were under admissible solutions in

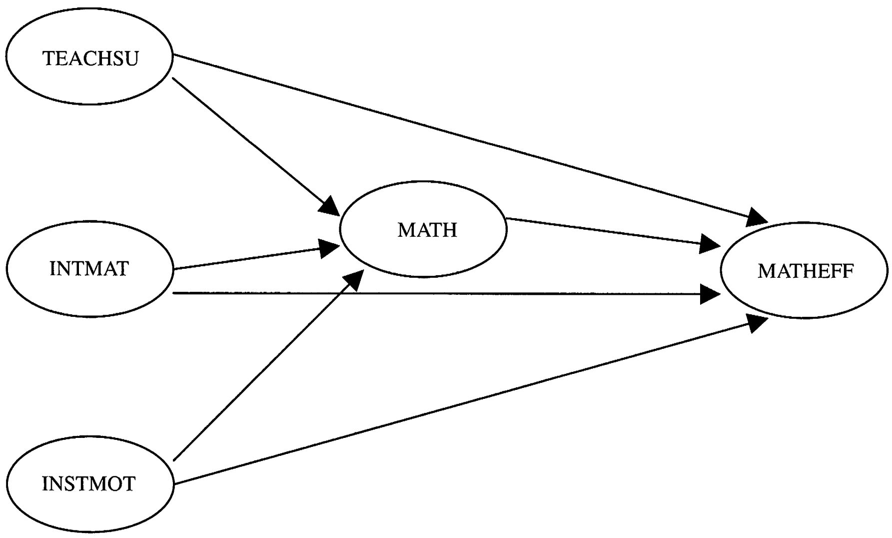

Figure 1: Two-level structural model of the homologous multilevel model.

student-level model and school-level model except that two of four mathematics grades indicators in school-level model were fixed as 0.002 for identification purposes. The degrees of freedom of the multilevel model could be calculated with respective level and then they were combined together. That is, [25×(25+1)/2-60]+[25×(25+1)/2-60+2]=265+267=532.

Since the multilevel structural equation model using 948 schools gave a good model fit as shown in Table 2, the effects of level-2 sample sizes were in turn assessed under seven different sets of school-level samples, which were randomly derived with replacement from 948 schools of Canada from PISA 2003 database. These seven sets of samples were arranged to suit the multilevel structural equation model respectively. We would first focus on the comparisons of model-fit indexes in these different level-2 sample sizes, and then the parameter estimates and their estimated standard errors of the multilevel structural equation model were compared in different level-2 sample sizes. In order to have a reference we also listed the results from 948 schools.

As school-level sample sizes increased from 120 to 948, the total sample sizes inevitably became larger gradually. But the average school size and average ICC of the twenty-five variables were controlled within a limited range. As shown in Table 2, as the school-level sample sizes increased, the average school sizes were almost the same at 28 students or so and the average ICC values were similar with few changes from 0.064 to 0.071. In addition, as expected, the χ^2 values gradually increased as total sample sizes or school-level sample sizes increased. However, the model-fit indexes CFI, TLI, RMSEA and

Table 1: ICC for each indicator of the five factors

Mathematics		INTMAT		INSTMOT		MATHEFF		TEACHSU	
content	ICC	item	ICC	item	ICC	item	ICC	item	ICC
G1	0.133	I1	0.037	I2	0.033	E1	0.060	T1	0.078
G2	0.171	I3	0.050	I5	0.042	E2	0.062	T2	0.072
G3	0.166	I4	0.041	I7	0.032	E3	0.047	T3	0.074
G4	0.142	I6	0.049	I8	0.035	E4	0.045	T4	0.079
						E5	0.055	T5	0.059
						E6	0.055		
						E7	0.045		
						E8	0.043		

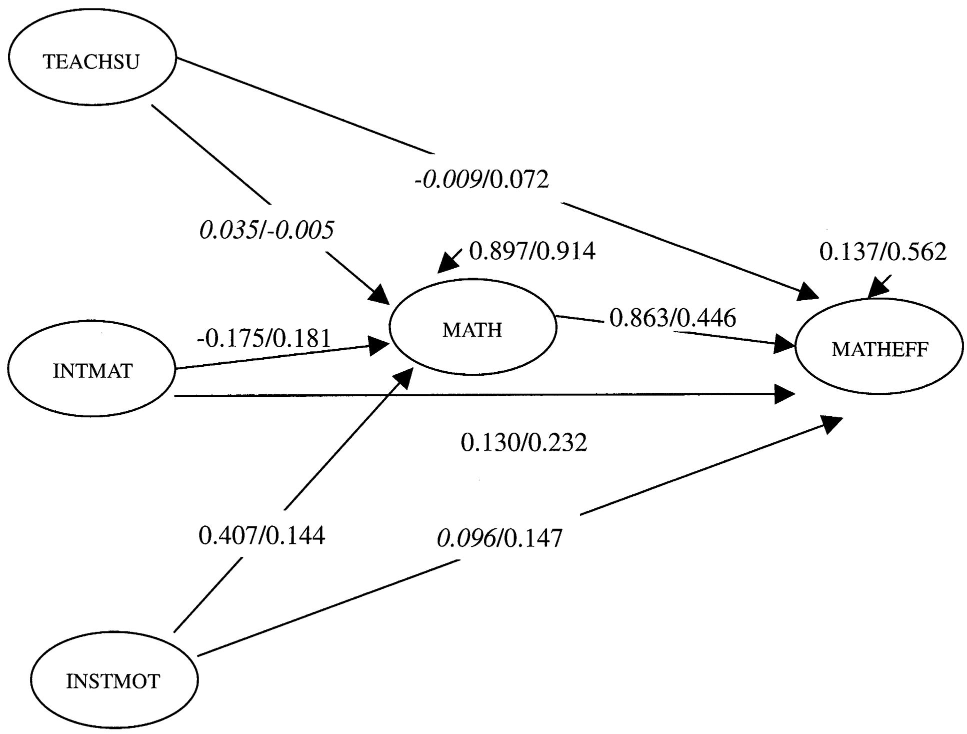

Figure 2: Two level structural model.
Note. Parameter estimates denote: between-level figure/within-level figure

within-level SRMR almost did not change as school-level sample sizes increased; namely, they were almost comparable in different school-level sample sizes. The exception was SRMR model-fit index for the between-level model. As school-level sample sizes increased, the school-level SRMR gradually decreased.

Moreover, although total sample sizes were large, the CFI, TLI, RMSEA and student-level SRMR were almost no change as school-level sample sizes increased, this might suggest that these model-fit indexes were insensitive not only to the different school-level sample sizes but also to the total sample sizes. Besides, since different school-level sample sizes were manipulated and some confounding sources were controlled in the experiment, and the school-level SRMR with obvious changes was the only corresponding outcome, the school-level sample sizes actually determined the degree of goodness-of-fit of between-level model. Hence, only school-level SRMR could help to determine how many level-2 units were sufficient to process the level-2 structural equation modeling under multilevel structural equation modeling condition. Note that if the SRMR below 0.08 was considered to be regular standard guideline, as mentioned earlier, a sample of level-2 units at least about 480 schools was necessary in this study. From the regular standard of the school-level SRMR of 0.08, the ratio of school-level units to the number of parameter estimates would be 480/(58 or 60), equal around to 8 in this study. That is, there is at least eight times or so for the level-2 units relative to the number of parameter estimates. The ratio is close to the "ten times" regular standard guideline in conventional structural equation modeling as suggested in Kline's book [11].

Thus far the degree of goodness-of-fit in multilevel structural equation modeling had been checked. In what follows, the parameter estimates and their estimated standard errors were compared in different school-level sample sizes. The stability of parameter estimates was emphasized especially for between-level model part.

In regard with within-level model, the estimates of factor loadings and residual variances, and their estimated standard errors for the within model of multilevel structural equation modeling were presented in Table 3 and Table 4. All parameter estimates in Table 3 and Table 4 were significant. The factor loadings and residual variances in within model were almost the same across the different school-level sample sizes, and their estimated standard errors, as expected, gradually decreased as school-level sample sizes increased. Similar findings also occurred to other parameter estimates in within model, such as structural paths, factor variances, and their estimated standard errors, as shown in Table 5. Although there existed one or two estimates with relatively change across the different school-level sample sizes in Table 5, such as the relationship of MATHEFF regressed on TEACHSU, the significances of the estimates were of consistence. In sum, the stabilities of parameter estimates in within-level model represented that the number of level-2 units actually did not affect level-1 parameter estimates even in the size of 120 schools.

As for between-level model, the estimates of factor loadings and residual variances, and their estimated standard errors for the between-level model of the multilevel structural equation modeling were listed in Table 6 and Table 7. These estimates in Table 6 and Table 7 were all significantly different from zero at 0.05 significance level and their estimated standard errors, as expected, roughly decreased as school-level sample sizes increased. Note that these parameter estimates were a little unstable when school-level samples were lower than 240 schools, such as factor loading of item I3, I4, I6, I8, and residual variances of item G1, I6, E1, E4, E7, and T3. Besides, it was still several estimates were unstable at the 360 school-level sample size, such as factor loadings of item I3 and I8.

Next, we checked for other parameter estimates in between-level model, such as structural paths, factor variances, and their estimated standard errors in Table 8. These parameter estimates in Table 8 changed dramatically as school-level sample size varied, especially as

Table 2: Model-fit indexes of different school-level sample sizes.

Number of schools	Total sample size	Average school size	Average ICC	χ^2	CFI	TLI	RMSEA	SRMR (Bet./Win.)
120	3358	27.877	0.071	3123.947	0.953	0.948	0.038	0.092/0.035
240	6959	28.935	0.066	6123.788	0.952	0.946	0.039	0.090/0.037
360	10542	29.248	0.064	9184.205	0.952	0.946	0.039	0.084/0.036
480	13440	27.980	0.070	11220.048	0.953	0.947	0.039	0.076/0.036
600	17160	28.581	0.071	14616.772	0.952	0.946	0.039	0.073/0.036
720	20583	28.570	0.067	17378.802	0.952	0.946	0.039	0.075/0.036
840	23900	28.439	0.067	19687.186	0.953	0.947	0.039	0.069/0.036
948	26884	28.346	0.068	22150.368	0.953	0.946	0.039	0.069/0.036

school-level sample sizes were small. Specifically, most structural paths were quite unstable when school-level samples were fewer than 360 schools, and other parameter estimates were unstable when school-level samples were fewer than 240 schools. Therefore, the number of schools beyond 480 or so would be a good choice to have relatively stable parameter estimates. Take the structural path from INSTMOT to MATH for example, the path coefficient unstably changed from 3.250, 1.758, to 0.672 when corresponding school-level sample sizes were 120, 240, and 360 respectively, and then stably changed from 1.096, 1.025, 0.837, 1.241, to 1.147 when corresponding school-level sample sizes were 480, 600, 720, 840 and 948 respectively. As for all estimated standard errors of these estimates still, as expected, roughly decreased as school-level sample sizes increased.

As a matter of fact, it was not easy to determine the plausible number of level-2 units based on the stabilities of between-level parameter estimates. As found in Table 8, the changes in structural paths were rather irregular. Nonetheless, the number of 480 schools was the best choice for all. Now the ratio of between-level sample sizes to the number of parameter estimates could be calculated as 480/(58 or 60), equal to around 8. That is, at least eight times for the number of level-2 units to the number of parameter estimates was warranted for stable estimates.

In sum, from the outcomes of the model-fit indexes and the stabilities of parameter estimates in multilevel structural equation modeling analysis, the plausible ratio of the number of level-2 units to the number of parameter estimates was 8:1.

REFERENCE

1. Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
2. Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.
3. Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
4. Duncan, T. E., Alpert, A., & Duncan, S. C. (1998). Multilevel covariance structure analysis of siblings antisocial behavior. *Structural Equation Modeling, 5*, 211-228.
5. Farmer, G. L. (2000). Use of multilevel covariance structure analysis to evaluate the multilevel nature of theoretical constructs. *Social Work Research, 24(3)*, 180-189.
6. Li, F., Duncan, T. E., Duncan, S. C., Harmer, P., & Acock, A. (1997). Latent variable modeling of multilevel intrinsic motivation data. *Measurement in Physical Education and Exercise Science, 1(4)*, 223-244.
7. Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22*, 376-398.
8. Hox, J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudo-balanced groups and small samples. *Structural Equation Modeling, 8*, 198-207.
9. Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality. Unpublished doctoral dissertation, University of Groningen.
10. Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58*, 134-146.
11. Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.
12. Bliese, P. D., & Halverson, R. R. (1998). Group size and measures of group-level properties: An examination of eta-squared and ICC values. *Journal of Management, 24(2)*, 157-172.
13. Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557-585.
14. Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28(4)*, 338-354.
15. Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83(2)*, 234-246.
16. Klein, K. J., & Kozlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods, 3*, 211-236.
17. Organization for Economic Co-operation and Development [OECD] (2005). *PISA 2003 technical report*. Paris: OECD.
18. Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide (4th ed.)*. Los Angeles, CA: Muthén & Muthén.

Table 3: Estimates of factor loadings and their standard errors of the within-level model.

	Number of Schools							
	120	240	360	480	600	720	840	948
MATH								
G1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
G2	1.055 (0.010)	1.058 (0.007)	1.060 (0.006)	1.064 (0.005)	1.064 (0.005)	1.063 (0.004)	1.061 (0.004)	1.062 (0.004)
G3	1.033 (0.011)	1.040 (0.007)	1.045 (0.006)	1.042 (0.005)	1.046 (0.005)	1.044 (0.004)	1.040 (0.004)	1.043 (0.004)
G4	1.035 (0.011)	1.042 (0.008)	1.037 (0.006)	1.039 (0.006)	1.041 (0.005)	1.040 (0.005)	1.039 (0.004)	1.040 (0.004)
INTMAT								
I1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
I3	1.146 (0.023)	1.128 (0.015)	1.098 (0.012)	1.106 (0.011)	1.109 (0.010)	1.114 (0.009)	1.110 (0.008)	1.112 (0.008)
I4	1.243 (0.024)	1.246 (0.016)	1.235 (0.013)	1.230 (0.011)	1.222 (0.010)	1.238 (0.009)	1.229 (0.008)	1.231 (0.008)
I6	1.110 (0.023)	1.108 (0.015)	1.116 (0.012)	1.100 (0.011)	1.104 (0.010)	1.112 (0.009)	1.107 (0.008)	1.108 (0.008)
INSTMOT								
I2	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
I5	0.949 (0.018)	0.938 (0.012)	0.943 (0.010)	0.965 (0.009)	0.975 (0.008)	0.959 (0.007)	0.965 (0.007)	0.961 (0.006)
I7	1.101 (0.021)	1.108 (0.014)	1.116 (0.011)	1.123 (0.011)	1.131 (0.009)	1.114 (0.008)	1.120 (0.008)	1.120 (0.007)
I8	0.954 (0.019)	0.957 (0.013)	0.955 (0.011)	0.976 (0.010)	0.976 (0.009)	0.961 (0.008)	0.963 (0.007)	0.964 (0.007)
MATHEFF								
E1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
E2	1.137 (0.036)	1.139 (0.026)	1.121 (0.020)	1.137 (0.018)	1.157 (0.016)	1.141 (0.015)	1.137 (0.014)	1.142 (0.013)
E3	1.316 (0.040)	1.303 (0.028)	1.285 (0.022)	1.305 (0.020)	1.310 (0.017)	1.293 (0.016)	1.297 (0.015)	1.298 (0.014)
E4	1.022 (0.034)	1.039 (0.024)	1.043 (0.019)	1.053 (0.017)	1.052 (0.015)	1.045 (0.014)	1.046 (0.013)	1.046 (0.012)
E5	0.946 (0.032)	0.932 (0.022)	0.983 (0.018)	0.953 (0.016)	0.964 (0.014)	0.948 (0.013)	0.941 (0.012)	0.946 (0.011)
E6	1.202 (0.040)	1.202 (0.028)	1.217 (0.022)	1.224 (0.020)	1.223 (0.018)	1.227 (0.016)	1.239 (0.015)	1.235 (0.014)
E7	1.076 (0.038)	1.081 (0.026)	1.137 (0.021)	1.107 (0.019)	1.107 (0.017)	1.101 (0.015)	1.104 (0.014)	1.104 (0.013)
E8	1.022 (0.037)	1.039 (0.026)	1.040 (0.021)	1.067 (0.019)	1.067 (0.017)	1.051 (0.015)	1.069 (0.015)	1.066 (0.014)
TEACHSU								
T1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
T2	1.074 (0.028)	1.042 (0.019)	1.012 (0.016)	1.006 (0.014)	1.009 (0.012)	1.013 (0.011)	1.013 (0.010)	1.012 (0.010)
T3	1.005 (0.025)	1.004 (0.017)	0.995 (0.015)	0.975 (0.012)	0.981 (0.011)	0.989 (0.010)	0.990 (0.009)	0.985 (0.009)
T4	1.120 (0.030)	1.108 (0.020)	1.107 (0.017)	1.087 (0.015)	1.077 (0.013)	1.091 (0.012)	1.094 (0.011)	1.091 (0.010)
T5	0.977 (0.029)	0.964 (0.020)	0.977 (0.017)	0.958 (0.015)	0.958 (0.013)	0.959 (0.012)	0.957 (0.011)	0.957 (0.010)

[a] They were fixed 1.000 as reference indexes.

Table 4: Estimates of residual variances and their standard errors of the within-level model.

	Number of Schools							
	120	240	360	480	600	720	840	948
G1	0.171 (0.005)	0.166 (0.003)	0.167 (0.003)	0.167 (0.002)	0.169 (0.002)	0.169 (0.002)	0.168 (0.002)	0.169 (0.002)
G2	0.036 (0.002)	0.037 (0.001)	0.035 (0.001)	0.035 (0.001)	0.035 (0.001)	0.035 (0.001)	0.035 (0.001)	0.035 (0.001)
G3	0.072 (0.002)	0.071 (0.002)	0.071 (0.001)	0.070 (0.001)	0.069 (0.001)	0.071 (0.001)	0.070 (0.001)	0.070 (0.001)
G4	0.092 (0.003)	0.092 (0.002)	0.096 (0.002)	0.096 (0.001)	0.098 (0.001)	0.097 (0.001)	0.096 (0.001)	0.097 (0.001)
I1	0.268 (0.008)	0.265 (0.005)	0.266 (0.004)	0.263 (0.004)	0.258 (0.003)	0.263 (0.003)	0.260 (0.003)	0.260 (0.003)
I3	0.187 (0.006)	0.194 (0.004)	0.211 (0.004)	0.199 (0.003)	0.199 (0.003)	0.200 (0.003)	0.198 (0.002)	0.197 (0.002)
I4	0.157 (0.006)	0.150 (0.004)	0.154 (0.003)	0.146 (0.003)	0.148 (0.003)	0.147 (0.002)	0.150 (0.002)	0.149 (0.002)
I6	0.202 (0.006)	0.203 (0.005)	0.202 (0.004)	0.208 (0.003)	0.204 (0.003)	0.204 (0.003)	0.204 (0.002)	0.203 (0.002)
I2	0.179 (0.006)	0.187 (0.004)	0.186 (0.003)	0.189 (0.003)	0.189 (0.003)	0.181 (0.002)	0.183 (0.002)	0.183 (0.002)
I5	0.165 (0.005)	0.167 (0.004)	0.169 (0.003)	0.164 (0.003)	0.163 (0.002)	0.163 (0.002)	0.164 (0.002)	0.164 (0.002)
I7	0.227 (0.008)	0.224 (0.005)	0.225 (0.004)	0.226 (0.004)	0.221 (0.003)	0.224 (0.003)	0.226 (0.003)	0.224 (0.003)
I8	0.216 (0.007)	0.219 (0.005)	0.224 (0.004)	0.219 (0.003)	0.217 (0.003)	0.216 (0.003)	0.219 (0.003)	0.218 (0.002)
E1	0.378 (0.010)	0.372 (0.007)	0.359 (0.006)	0.354 (0.005)	0.353 (0.004)	0.358 (0.004)	0.359 (0.004)	0.357 (0.003)
E2	0.329 (0.009)	0.322 (0.006)	0.323 (0.005)	0.314 (0.004)	0.303 (0.004)	0.314 (0.004)	0.314 (0.003)	0.313 (0.003)
E3	0.312 (0.010)	0.308 (0.007)	0.308 (0.005)	0.298 (0.005)	0.295 (0.004)	0.301 (0.004)	0.301 (0.003)	0.300 (0.003)
E4	0.315 (0.009)	0.315 (0.006)	0.295 (0.005)	0.298 (0.004)	0.295 (0.004)	0.297 (0.003)	0.299 (0.003)	0.298 (0.003)
E5	0.300 (0.008)	0.287 (0.005)	0.277 (0.004)	0.277 (0.004)	0.272 (0.003)	0.275 (0.003)	0.272 (0.003)	0.272 (0.003)
E6	0.429 (0.012)	0.434 (0.008)	0.416 (0.007)	0.422 (0.006)	0.416 (0.005)	0.418 (0.005)	0.419 (0.004)	0.418 (0.004)
E7	0.446 (0.012)	0.419 (0.008)	0.395 (0.006)	0.404 (0.006)	0.402 (0.005)	0.400 (0.004)	0.403 (0.004)	0.402 (0.004)
E8	0.472 (0.013)	0.469 (0.009)	0.453 (0.007)	0.459 (0.006)	0.454 (0.005)	0.455 (0.005)	0.453 (0.005)	0.453 (0.004)
T1	0.361 (0.011)	0.361 (0.007)	0.372 (0.006)	0.368 (0.005)	0.363 (0.005)	0.369 (0.004)	0.371 (0.004)	0.367 (0.004)
T2	0.302 (0.010)	0.294 (0.006)	0.307 (0.005)	0.305 (0.005)	0.294 (0.004)	0.299 (0.004)	0.298 (0.004)	0.296 (0.003)
T3	0.215 (0.007)	0.195 (0.005)	0.209 (0.004)	0.198 (0.004)	0.201 (0.003)	0.199 (0.003)	0.195 (0.003)	0.197 (0.002)
T4	0.350 (0.011)	0.341 (0.007)	0.344 (0.006)	0.352 (0.006)	0.347 (0.005)	0.345 (0.004)	0.351 (0.004)	0.350 (0.004)
T5	0.474 (0.013)	0.472 (0.009)	0.468 (0.007)	0.482 (0.007)	0.484 (0.006)	0.477 (0.005)	0.483 (0.005)	0.478 (0.005)
MATH	0.621 (0.019)	0.643 (0.014)	0.648 (0.011)	0.635 (0.010)	0.638 (0.009)	0.644 (0.008)	0.639 (0.007)	0.638 (0.007)
MATHEFF	0.124 (0.007)	0.123 (0.005)	0.119 (0.004)	0.118 (0.003)	0.120 (0.003)	0.119 (0.003)	0.116 (0.003)	0.118 (0.002)

Table 5: Estimates of structural paths, factor variances, factor correlations and their standard errors of the within-level model.

	Number of Schools							
	120	240	360	480	600	720	840	948
Structural paths								
MATH on TEACHSU	-0.023[a] (0.028)	-0.020[a] (0.019)	-0.032 (0.016)	-0.015[a] (0.014)	-0.014[a] (0.012)	-0.012[a] (0.012)	-0.014[a] (0.011)	-0.007[a] (0.010)
MATH on INTMAT	0.226 (0.035)	0.211 (0.024)	0.253 (0.020)	0.243 (0.017)	0.238 (0.015)	0.234 (0.014)	0.248 (0.013)	0.242 (0.012)
MATH on INSTMOT	0.183 (0.034)	0.216 (0.023)	0.197 (0.019)	0.188 (0.017)	0.202 (0.015)	0.205 (0.014)	0.187 (0.013)	0.191 (0.012)
MATHEFF on MATH	0.255 (0.011)	0.244 (0.008)	0.252 (0.006)	0.251 (0.005)	0.249 (0.005)	0.246 (0.004)	0.243 (0.004)	0.245 (0.004)
MATHEFF on TEACHSU	0.095 (0.014)	0.073 (0.010)	0.059 (0.008)	0.053 (0.007)	0.054 (0.006)	0.053 (0.006)	0.056 (0.005)	0.055 (0.005)
MATHEFF on INTMAT	0.202 (0.018)	0.188 (0.012)	0.179 (0.010)	0.171 (0.008)	0.174 (0.008)	0.172 (0.007)	0.166 (0.006)	0.170 (0.006)
MATHEFF on INSTMOT	0.089 (0.017)	0.094 (0.012)	0.093 (0.009)	0.109 (0.008)	0.100 (0.007)	0.108 (0.007)	0.110 (0.006)	0.107 (0.006)
Factor variances								
TEACHSU	0.356 (0.016)	0.359 (0.011)	0.349 (0.009)	0.356 (0.008)	0.358 (0.007)	0.351 (0.007)	0.355 (0.006)	0.354 (0.006)
INTMAT	0.367 (0.015)	0.376 (0.010)	0.393 (0.009)	0.388 (0.008)	0.390 (0.007)	0.388 (0.006)	0.394 (0.006)	0.390 (0.005)
INSTMOT	0.384 (0.014)	0.406 (0.010)	0.413 (0.008)	0.388 (0.007)	0.389 (0.006)	0.401 (0.006)	0.397 (0.005)	0.395 (0.005)
factor correlations								
TEACHSU with INTMAT	0.117 (0.008)	0.119 (0.006)	0.124 (0.005)	0.124 (0.004)	0.124 (0.004)	0.125 (0.003)	0.127 (0.003)	0.126 (0.003)
TEACHSU with INSTMOT	0.118 (0.008)	0.127 (0.006)	0.125 (0.005)	0.123 (0.004)	0.119 (0.004)	0.125 (0.003)	0.125 (0.003)	0.124 (0.003)
INTMAT with INSTMOT	0.240 (0.010)	0.250 (0.007)	0.264 (0.006)	0.242 (0.005)	0.247 (0.004)	0.253 (0.004)	0.253 (0.004)	0.251 (0.004)

[a] They were not statistically significant at $p<0.05$; all other parameter estimates were significant.

Table 6: Estimates of factor loadings and their standard errors of the between-level model.

	Number of Schools							
	120	240	360	480	600	720	840	948
MATH								
G1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
G2	1.170	1.164	1.141	1.148	1.152	1.137	1.144	1.144
	(0.036)	(0.025)	(0.019)	(0.017)	(0.014)	(0.013)	(0.012)	(0.011)
G3	1.174	1.161	1.120	1.138	1.140	1.119	1.128	1.130
	(0.037)	(0.025)	(0.019)	(0.017)	(0.014)	(0.013)	(0.012)	(0.012)
G4	1.174	1.059	1.049	1.051	1.048	1.040	1.048	1.046
	(0.037)	(0.024)	(0.019)	(0.017)	(0.014)	(0.013)	(0.012)	(0.012)
INTMAT								
I1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
I3	0.389	0.638	0.789	0.870	0.959	0.944	0.908	0.939
	(0.236)	(0.237)	(0.131)	(0.111)	(0.079)	(0.080)	(0.078)	(0.070)
I4	1.778	1.812	1.350	1.338	1.250	1.213	1.268	1.229
	(0.307)	(0.280)	(0.124)	(0.108)	(0.070)	(0.070)	(0.070)	(0.061)
I6	2.142	2.344	1.445	1.517	1.329	1.196	1.286	1.263
	(0.397)	(0.400)	(0.146)	(0.133)	(0.081)	(0.076)	(0.078)	(0.069)
INSTMOT								
I2	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
I5	0.991	1.127	1.043	1.048	1.090	1.068	1.180	1.110
	(0.123)	(0.122)	(0.087)	(0.061)	(0.054)	(0.056)	(0.060)	(0.050)
I7	0.922	0.971	1.013	1.005	0.995	1.053	1.047	1.027
	(0.141)	(0.116)	(0.086)	(0.063)	(0.053)	(0.057)	(0.056)	(0.049)
I8	0.670	0.888	0.873	0.932	0.953	0.946	0.938	0.952
	(0.124)	(0.112)	(0.085)	(0.062)	(0.055)	(0.056)	(0.054)	(0.048)
MATHEFF								
E1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
E2	1.016	1.009	1.130	1.080	1.067	1.076	1.044	1.055
	(0.147)	(0.104)	(0.085)	(0.076)	(0.062)	(0.057)	(0.053)	(0.051)
E3	1.109	1.028	0.960	0.935	0.942	0.886	0.937	0.938
	(0.158)	(0.107)	(0.072)	(0.065)	(0.056)	(0.051)	(0.048)	(0.046)
E4	0.912	0.938	0.828	0.897	0.871	0.882	0.897	0.906
	(0.141)	(0.098)	(0.065)	(0.063)	(0.052)	(0.048)	(0.044)	(0.043)
E5	0.892	0.906	0.815	0.815	0.848	0.798	0.826	0.841
	(0.129)	(0.094)	(0.066)	(0.060)	(0.053)	(0.047)	(0.044)	(0.042)
E6	0.749	0.645	0.580	0.677	0.733	0.596	0.685	0.677
	(0.157)	(0.112)	(0.082)	(0.076)	(0.066)	(0.059)	(0.056)	(0.053)
E7	0.780	0.795	0.604	0.682	0.688	0.600	0.695	0.685
	(0.137)	(0.104)	(0.074)	(0.068)	(0.060)	(0.052)	(0.050)	(0.047)
E8	0.523	0.479	0.449	0.526	0.495	0.370	0.523	0.466
	(0.150)	(0.107)	(0.075)	(0.071)	(0.060)	(0.053)	(0.051)	(0.048)
TEACHSU								
T1	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
T2	1.259	0.960	0.972	0.873	0.890	0.936	0.900	0.903
	(0.152)	(0.076)	(0.066)	(0.053)	(0.047)	(0.046)	(0.040)	(0.038)
T3	1.152	0.911	0.923	0.877	0.900	0.921	0.911	0.898
	(0.127)	(0.065)	(0.053)	(0.044)	(0.041)	(0.039)	(0.035)	(0.033)
T4	1.309	1.199	1.102	1.182	1.156	1.111	1.095	1.104
	(0.157)	(0.086)	(0.068)	(0.059)	(0.053)	(0.049)	(0.044)	(0.041)
T5	0.961	0.905	0.873	0.887	0.885	0.885	0.905	0.904
	(0.148)	(0.082)	(0.064)	(0.055)	(0.049)	(0.047)	(0.043)	(0.040)

[a] They were fixed 1.000 as reference indexes.

Table 7: Estimates of residual variances and their standard errors of the between-level model.

	Number of Schools							
	120	240	360	480	600	720	840	948
G1	0.006 (0.002)	0.004 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.000)	0.002 (0.000)	0.002 (0.000)
G2	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]
G3	0.001 (0.001)	0.000 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.002 (0.000)	0.002 (0.000)
G4	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]	0.002[a]
I1	0.007 (0.002)	0.005 (0.002)	0.005 (0.001)	0.007 (0.001)	0.005 (0.001)	0.004 (0.001)	0.005 (0.001)	0.005 (0.001)
I3	0.018 (0.004)	0.025 (0.003)	0.020 (0.002)	0.020 (0.002)	0.020 (0.002)	0.020 (0.002)	0.020 (0.001)	0.020 (0.001)
I4	0.002 (0.002)	0.003 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)	0.003 (0.001)	0.002 (0.001)	0.002 (0.001)
I6	0.002 (0.003)	0.001 (0.002)	0.007 (0.002)	0.005 (0.001)	0.006 (0.001)	0.007 (0.001)	0.006 (0.001)	0.006 (0.001)
I2	0.001 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)	0.001 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)
I5	0.002 (0.001)	0.003 (0.001)	0.002 (0.001)	0.002 (0.001)	0.003 (0.001)	0.003 (0.001)	0.002 (0.001)	0.003 (0.001)
I7	0.006 (0.002)	0.004 (0.001)	0.002 (0.001)	0.004 (0.001)	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)
I8	0.004 (0.002)	0.004 (0.001)	0.004 (0.001)	0.005 (0.001)	0.005 (0.001)	0.005 (0.001)	0.004 (0.001)	0.004 (0.001)
E1	0.017 (0.004)	0.014 (0.003)	0.008 (0.002)	0.012 (0.002)	0.009 (0.001)	0.009 (0.001)	0.009 (0.001)	0.010 (0.001)
E2	0.009 (0.003)	0.009 (0.002)	0.014 (0.002)	0.015 (0.002)	0.011 (0.001)	0.012 (0.001)	0.012 (0.001)	0.012 (0.001)
E3	0.011 (0.003)	0.010 (0.002)	0.007 (0.002)	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)	0.009 (0.001)	0.009 (0.001)
E4	0.011 (0.003)	0.007 (0.002)	0.006 (0.001)	0.008 (0.001)	0.006 (0.001)	0.006 (0.001)	0.005 (0.001)	0.005 (0.001)
E5	0.005 (0.002)	0.006 (0.002)	0.008 (0.001)	0.009 (0.001)	0.009 (0.001)	0.009 (0.001)	0.008 (0.001)	0.008 (0.001)
E6	0.027 (0.006)	0.026 (0.004)	0.025 (0.003)	0.028 (0.003)	0.025 (0.002)	0.026 (0.002)	0.026 (0.002)	0.025 (0.002)
E7	0.010 (0.004)	0.015 (0.003)	0.018 (0.003)	0.019 (0.002)	0.018 (0.002)	0.016 (0.002)	0.016 (0.002)	0.016 (0.001)
E8	0.028 (0.006)	0.025 (0.004)	0.020 (0.003)	0.025 (0.003)	0.020 (0.002)	0.019 (0.002)	0.021 (0.002)	0.020 (0.002)
T1	0.008 (0.003)	0.008 (0.002)	0.010 (0.002)	0.011 (0.002)	0.013 (0.002)	0.012 (0.002)	0.012 (0.001)	0.012 (0.001)
T2	0.010 (0.003)	0.009 (0.002)	0.016 (0.002)	0.016 (0.002)	0.014 (0.002)	0.014 (0.002)	0.013 (0.001)	0.013 (0.001)
T3	0.001 (0.002)	0.003 (0.001)	0.004 (0.001)	0.005 (0.001)	0.006 (0.001)	0.005 (0.001)	0.005 (0.001)	0.005 (0.001)
T4	0.009 (0.004)	0.006 (0.002)	0.011 (0.002)	0.008 (0.002)	0.008 (0.002)	0.009 (0.002)	0.008 (0.001)	0.008 (0.001)
T5	0.016 (0.005)	0.012 (0.003)	0.012 (0.002)	0.013 (0.002)	0.013 (0.002)	0.012 (0.002)	0.013 (0.002)	0.012 (0.002)
MATH	0.103 (0.033)	0.119 (0.016)	0.113 (0.011)	0.115 (0.010)	0.124 (0.009)	0.120 (0.008)	0.114 (0.007)	0.117 (0.007)
MATHEFF	0.003 (0.002)	0.003 (0.001)	0.002 (0.001)	0.003 (0.001)	0.003 (0.001)	0.004 (0.001)	0.004 (0.001)	0.004 (0.001)

[a] They were fixed 0.002 for identification purposes.

Table 8: Estimates of structural paths, factor variances, factor correlations and their standard errors of the between-level model.

	Number of Schools							
	120	240	360	480	600	720	840	948
Structural paths								
MATH on TEACHSU	0.189[a] (0.419)	0.072[a] (0.196)	0.107[a] (0.131)	0.062[a] (0.117)	-0.012[a] (0.110)	0.071[a] (0.097)	0.038[a] (0.090)	0.058[a] (0.085)
MATH on INTMAT	-3.449[a] (2.303)	-2.128[a] (1.183)	-0.146[a] (0.429)	-0.797 (0.390)	-0.367[a] (0.311)	-0.183[a] (0.255)	-0.404[a] (0.239)	-0.501 (0.236)
MATH on INSTMOT	3.250[a] (1.956)	1.758 (0.824)	0.672[a] (0.391)	1.096 (0.290)	1.025 (0.293)	0.837 (0.248)	1.241 (0.243)	1.147 (0.224)
MATHEFF on MATH	0.316 (0.065)	0.364 (0.037)	0.402 (0.028)	0.399 (0.026)	0.381 (0.021)	0.395 (0.020)	0.382 (0.019)	0.385 (0.017)
MATHEFF on TEACHSU	0.130[a] (0.102)	0.093[a] (0.054)	-0.026[a] (0.041)	-0.047[a] (0.037)	0.003[a] (0.032)	0.007[a] (0.031)	-0.007[a] (0.028)	-0.007[a] (0.026)
MATHEFF on INTMAT	-0.057[a] (0.609)	0.067[a] (0.311)	0.344 (0.137)	0.333 (0.126)	0.220 (0.092)	0.097[a] (0.082)	0.193 (0.075)	0.166 (0.072)
MATHEFF on INSTMOT	0.418[a] (0.543)	0.176[a] (0.226)	0.146[a] (0.123)	0.086[a] (0.095)	0.094[a] (0.088)	0.110[a] (0.081)	0.142[a] (0.078)	0.120[a] (0.071)
Factor variances								
TEACHSU	0.028 (0.007)	0.040 (0.006)	0.044 (0.006)	0.048 (0.005)	0.048 (0.005)	0.044 (0.004)	0.046 (0.004)	0.047 (0.004)
INTMAT	0.009 (0.004)	0.005 (0.002)	0.012 (0.003)	0.012 (0.002)	0.018 (0.002)	0.016 (0.002)	0.015 (0.002)	0.016 (0.002)
INSTMOT	0.014 (0.004)	0.011 (0.003)	0.014 (0.003)	0.019 (0.003)	0.019 (0.002)	0.017 (0.002)	0.014 (0.002)	0.016 (0.002)
factor correlations								
TEACHSU with INTMAT	0.008 (0.003)	0.007 (0.002)	0.010 (0.002)	0.012 (0.002)	0.016 (0.002)	0.012 (0.002)	0.012 (0.002)	0.014 (0.002)
TEACHSU with INSTMOT	0.010 (0.004)	0.010 (0.003)	0.010 (0.002)	0.013 (0.002)	0.015 (0.002)	0.012 (0.002)	0.012 (0.002)	0.013 (0.002)
INTMAT with INSTMOT	0.009 (0.003)	0.006 (0.002)	0.009 (0.002)	0.011 (0.002)	0.014 (0.002)	0.012 (0.002)	0.010 (0.001)	0.011 (0.001)

[a] They were not statistically significant at $p<0.05$; all other parameter estimates were significant.