Qual Quant (2012) 46:155–188
DOI 10.1007/s11135-010-9346-4

RESEARCH PAPER

Constructing measures for school process variables: the potential of multilevel confirmatory factor analysis

Ellen D'Haenens · Jan Van Damme · Patrick Onghena

Published online: 23 June 2010

Abstract This article investigates the potential of multilevel confirmatory factor analysis compared to using conventional confirmatory factor analysis, as regards the construction of measures for school process variables within educational effectiveness research. Since both techniques have not been systematically compared within educational research, we propose to contrast several aspects: factor structure, pattern of factor loadings, communalities of the indicators, variance extracted by the factors, internal consistency of the factors, and inter-factor correlations. To this end, teacher questionnaire data are used to construct measures for school process variables. Our results show no substantial differences when comparing both techniques at the teacher level. However, when contrasting the different levels within the more complex technique, we notice a fairly large discrepancy between the teacher-level results and the school-level results.

Keywords School process variables · Secondary education · Multilevel confirmatory factor analysis · Educational effectiveness research

Abbreviations

CFA	Confirmatory factor analysis
MCFA	Multilevel confirmatory factor analysis
LOSO	'Longitudinaal Onderzoek in het Secundair Onderwijs' or 'Longitudinal Research in Secondary Education'

A previous version of this study was presented at the annual meeting of the American Educational Research Association, March 2007, Chicago, Illinois.

E. D'Haenens (✉) · J. Van Damme
Department of Educational Sciences, Centre for Educational Effectiveness and Evaluation, K.U.Leuven, Dekenstraat, P.O. Box 3773, 3000 Leuven, Belgium
e-mail: Ellen.Dhaenens@ped.kuleuven.be

P. Onghena
Department of Educational Sciences, Centre for Methodology of Educational Research, K.U.Leuven, Andreas Vesaliusstraat 2, P.O. Box 3762, 3000 Leuven, Belgium

ICC	Intraclass correlation coefficient
MLR	Robust Maximum Likelihood
RMSEA	Root Mean Square Error of Approximation
SRMR	Standardized Root Mean Square Residual
TLI	Tucker-Lewis Index
CFI	Comparative Fit Index
MSA	Measure of Sampling Adequacy
StdYX	Completely standardized solution

1 Introduction

In educational effectiveness research, school process variables are commonly investigated to determine their association with diverse student outcomes. School process variables refer to particular aspects of the school environment. For instance, the extent to which teachers cooperate within a school has been found to be positively associated with student achievement. This means that in schools where teachers cooperate more, students tend to perform better in terms of achievement tests (Scheerens and Bosker 1997). Evidently, revealing the school process variables that matter for students could inspire policy makers and school heads to render schooling more effective.

1.1 Collecting and constructing information on school process variables

Several techniques can be used to capture the process variables, such as questionnaires, observations, and interviews (Freiberg 1999). Obviously, questionnaires are a relatively easy and low-cost means to attain a large group of information suppliers, and thus are often preferred to obtain information about process variables at the school level. Using these surveys, school process variables are typically described as the shared perceptions of students or teachers about their school environment (Fraser and Walberg 1991; Halpin and Croft 1962; Hoy et al. 1991). In the current study, the shared perceptions of teachers about their school environment will be considered.

One should consider though that valid and reliable measures for the school process variables still need to be constructed based on the individual teacher survey data. To date, researchers have generally used a conventional factor analysis applied to the teacher-level data. This way, the associations between the items of the survey are investigated, yielding one or more latent factors that account for a considerable proportion of common variance in the items. This can be achieved in either an exploratory or a confirmatory way. In cases where some a priori expectations exist about the factor structure, a confirmatory factor analysis (CFA) is advised. Using a CFA, one can test a single hypothesized factor structure, compare several alternative factor structures, or further refine the factor structure based on the a priori assumptions (Hair et al. 1998; Schumacker and Lomax 1996). However, the acquired factors only represent how individual teachers perceive their school environment. To obtain global measures for school process variables, in which we are typically interested as an educational effectiveness researcher, the individual factor scores or scale scores are aggregated at the school level (Rowan et al. 1991).

While this conventional strategy seems to be generally applied in most educational effectiveness studies (either exploratory or confirmatory), some major problems arise in case of hierarchically organized data. First, a technical-statistical difficulty comes into play. Regarding our teacher survey data, different teachers from the same school describe a similar school environment. This way, data are no longer independent but correlated, violating the fundamental independence assumption underlying many commonly used statistical techniques like factor analysis (Dyer et al. 2005; Hair et al. 1998). Consequently, using a conventional factor analysis at the teacher level leads to overestimated parameters and underestimated standard errors (Julian 2001; Muthén and Satorra 1995), which in turn cause interpretational problems. A conceptual issue yields a second challenge. Aggregation of the teacher-level factors into the school level assumes functional correspondence between the latent dimensions at the teacher level and the measures for school process variables. However, among the few studies applying a two-level factor analysis within educational research, some have actually demonstrated a distinct latent factor structure at the individual and at the group level (Holfve-Sabel and Gustafsson 2005; Westling Allodi 2002). Holfve-Sabel and Gustafsson have constructed latent factors both at the pupil level and at the class level based on pupils' responses regarding attitudes concerning their school, teacher, classmates, and other schooling aspects. By means of a two-level confirmatory factor analysis, they have created seven latent constructs at the pupil level and three latent constructs at the class level by collapsing highly correlated factors of the pupil level.

Thus, a solution to both the technical-statistical problem and the conceptual issue is to perform a multilevel factor analysis. Muthén (1989, 1990) has shown that a multilevel confirmatory factor analysis (MCFA) can be performed using conventional software analysis for covariance structure analysis, like M*plus* (Muthén and Muthén 1998–2006) and LISREL (Jöreskog and Sörbom 1996). Specifying the model as a two-level model, these programs automatically calculate the necessary within- and between-group covariance matrices and use multiple group modelling to model these matrices simultaneously (Stapleton 2006). This way, two separate factor analysis models are obtained: one that accounts for the structure of the items between individuals within groups and one that accounts for the structure of the items between groups (Muthén 1994). Thus, MCFA separates the collective part of the perceptions from the individual, idiosyncratic part (den Brok et al. 2004). Importantly, this means that MCFA should be distinguished from a multilevel regression analysis. While a multilevel regression analysis constitutes one analysis, MCFA actually considers two. In fact, using MCFA the total variance in the observed items is decomposed into four main categories: variance due to latent factors at the group level, residual variance at the group level, variance due to latent factors at the individual level, and residual variance at the individual level. For more technical details, see Muthén (1989, 1990).

However, despite the auspicious findings concerning multilevel factor analysis, researchers confronted with a nested data structure still do not use this alternative approach in a consistent way. In fact, its specific added value in the domain of educational effectiveness has hardly been systematically compared to the more conventional approach. Therefore, we propose to use both conventional CFA at the teacher level and MCFA at the teacher level and at the school level to construct measures for school process variables, and thoroughly compare different aspects of the solutions.

1.2 Results from a multilevel exploratory factor analysis

In a previous study (D'Haenens et al. 2008), multilevel exploratory factor analysis has been used to explore the factor structure at both the teacher level and the school level. In line with Holfve-Sabel and Gustafsson (2005), we found broader factors at the school level in comparison with the teacher level. In addition, we have observed higher internal consistencies (i.e., Cronbach's alpha) of the school-level factors, irrespective of the number of items included in the latent variables.

However, one can argue that the use of an exploratory factor analysis is to a certain degree a subjective affair due to several decisions one has to take, and therefore less appropriate to test the existence of different factor structures at different levels. A major advantage of CFA involves the possibility of testing and comparing different factor structures, while evaluating the difference in fit between the models. This makes it more straightforward and thus less subjective to choose the most appropriate model at each of both levels.

1.3 Research objectives

The overall objective of the current study is to assess the potential of MCFA to construct meaningful measures for process characteristics at the school level, based on teacher data. This method is compared to the use of CFA at the teacher level, without taking into account the hierarchical structure in our data.

Several aspects are considered when comparing the results: (a) the factor structure, (b) the pattern of factor loadings (i.e., pattern coefficients), (c) the communalities of the indicators, (d) the variance extracted by the factors, (e) the internal consistency of the factors, and (f) the inter-factor correlations.

2 Method

2.1 Sample

Data from the LOSO-project (i.e., 'Longitudinaal Onderzoek in het Secundair Onderwijs' or 'Longitudinal Research in Secondary Education') were used to carry out the analyses and address our research objectives. Starting from the beginning of the 1990s, 6411 students entering 57 schools in Flanders (i.e., the Dutch-speaking part of Belgium) were followed throughout their entire secondary educational career. In Flanders, secondary education consists of six grades that are grouped into three cycles. Due to practical reasons, almost all schools from two regions in Flanders were included. This way, students could be followed more easily whenever they changed schools. The selected schools represented to a certain degree the Flemish secondary schools in general. Among other things, attention was given to following aspects: the curriculum offered, the school size, the school type, and the school system (i.e., the percentage of Catholic and public schools). Data from several respondents were collected: students, teachers, principals, and parents (Van Damme et al. 2002).

The current study is restricted to teacher survey data. During the school year 1990–1991, teachers from 86 schools answered items concerning their background, teaching practice, educational and pedagogical framework, and daily school life. This inquiry was done separately for the teachers from the first cycle (i.e., junior high school in Flanders) and for the teachers from the last two cycles (i.e., senior high school in Flanders). To this end, representative samples of approximately 15 cycle-1 teachers, and 17 cycle-2 and -3 teachers were drawn from each school where possible. In four schools with a small number of teachers, a representative sample of approximately 17 teachers was drawn across all three cycles. When drawing the sample, the subject taught by teachers was taken into account. This way, the samples were comparable across schools and representative for the entire staff in each school (Van Damme et al. 2004).

Previous results based on our 'multilevel exploratory factor analysis'-study (D'Haenens et al. 2008) indicated differences in some school process variables between cycle 1 on the one hand, and cycles 2 and 3 on the other hand. This is not surprising, since a different organization of the study programmes exists between these cycles. Therefore, in accordance with this previous study, we split up the schools with three cycles into a ‘cycle-1 school group’ and a ‘cycle-2 and -3 school group’. The four schools with only a small number of teachers were not included in our analyses, since we performed a multiple group analysis in advance (i.e., cycle 1 versus cycles 2 and 3). Splitting up these samples into cycle 1 versus cycles 2 and 3 would make them unrepresentative regarding the subjects taught. Thus, 82 schools remained. Of these 82 schools, 22 only provided cycle-1 education, 28 only provided cycle-2 and -3 education, and 32 provided education in all three cycles. For two of these 32 schools, only one sample was drawn since they organized the first cycle together when the questionnaire was administered. From now on, the word ‘school’ will be used in the meaning of a ‘cycle group’. Our final sample comprised 1550 teachers from 113 ‘cycle groups’. This sample comprised approximately 44% teachers from the first cycle and 56% teachers from the highest two cycles. The teacher sample ranged from 1 to 18 within each of the cycle groups. Both sexes were (unintentionally) nearly equally represented, and the teachers’ age (with 69 missing) ranged from 23 until 63, with a mean age of approximately 42 ($SD = 7.80$).

2.2 Data: School Characteristics Questionnaire for Teachers

To develop measures for school process variables, we used data from the ‘School Characteristics Questionnaire for Teachers’. This questionnaire encompassed two different parts consisting of more than 300 items and addressed issues concerning the teachers’ background, teaching practice, educational and pedagogical framework, and daily school life (Van Damme et al. 2004). For the construction of this questionnaire we refer to Opdenakker (2003).

Due to the complexity of MCFA (cf., many convergence problems) when analysing a large amount of items, we selected a subsample of 66 items from the second part of the questionnaire: ‘the daily school life of teachers’. A group of 41 items was judged on a 4-point scale, whereas 25 items were judged on a 5-point scale (see overview in Appendix 1). We acknowledge that this decision inevitably affected the factor structure.

During the selection process, we tried to take account of the empirical support concerning the association with several student outcomes in the research literature. Four different aspects of the daily school life of teachers were included in the present study:

(1) Cooperation between teachers, both between teachers of the same class and between teachers instructing the same subject. For more information on the association with student achievement and school well-being, see for instance Anderson (1982), Mortimore (1997), Scheerens (1990), Scheerens and Bosker (1997), Vandenberghe et al. (1994) and Verhoeven et al. (1992).
(2) Decision making, in particular regarding instruction. For more information on the relationship with student achievement, see Sweetland and Hoy (2000).
(3) Social system, which relates to the relationships among teachers and between teachers and the principal. For more information on the direct link with teachers’ participatory decision making concerning class and instruction, and thus on the indirect connection with student achievement, see Sweetland and Hoy (2000).
(4) Establishment of rules, as one aspect of an orderly school atmosphere (Scheerens 1990; Scheerens and Bosker 1997; Stockard and Mayberry 1992).

2.3 Analysis

First, our data were screened and necessary assumptions were assessed using SAS 9.1 (SAS Institute Inc. 2003) and SPSS 15.0 (SPSS 2006).

Next, the design effect was calculated for each of the observed items to determine whether all of our selected items were able to differentiate between the schools included in our study. The design effect is a function of both the intraclass correlation coefficient (ICC), indicating the proportion of variance observed between the schools, and the average school size:

$$\text{Design effect} = 1 + (\text{average group size} - 1)\,^{*}\,\text{ICC}$$

A design effect with a value larger than 2 requires multilevel analysis (Muthén and Satorra 1995). We decided to delete the items that insufficiently discriminated between schools. This way, we were able to illustrate more clearly the consequences of ignoring the dependency present in the data. Furthermore, our goal was to construct process characteristics that vary adequately between schools.

We used Robust Maximum Likelihood (MLR) estimation to deal with missing data and non-normality of our observed data. This estimator yields both a chi-square test statistic and standard errors that are robust to non-normality (Muthén and Muthén 1998–2006). Basically, robust standard errors for missing data require data to be *missing completely at random*, although some simulation results have suggested these standard errors to be relatively accurate under the assumption of *missing at random* as well (see Enders 2006). *Missing at random* means the missingness of a certain variable is related to one or more other observed variables in the model, while *missing completely at random* indicates that the missingness of the variable is unrelated to the values of other observed variables as well as to the values of the variable with missings itself. In Sect. 3, it will be shown that *missing at random* is a reasonable assumption in this case. Since MLR is computationally very demanding, especially in combination with two levels, we were required to split up our data set of 66 items. Based on the review of Scheerens and Bosker (1997), we constructed one group of items concerning school climate, including an orderly atmosphere (i.e., establishment of rules) and internal relationships (i.e., social system). This way, a second group of items concerned school organizational characteristics, including the decision making process and cooperation between teachers. Again, we acknowledge that this decision undoubtedly affected the factor structure.

Due to frequent convergence problems with MCFA-applications, Muthén (1994) has proposed a five-step procedure to conduct MCFAs. However, another interesting alternative is the use of the software program STREAMS (i.e., STRuctural Equation Modeling made Simple; Gustafsson and Stahl 2005). STREAMS provides pre- and postprocessors to software programs for covariance structure analysis, which make it relatively easy to set up, estimate, and interpret two-level CFAs. An interesting feature offered by STREAMS concerns the use of the previously estimated model to provide starting values for the currently estimated model. We used STREAMS 3.0.4 in combination with M*plus* 4.0 to estimate and analyse the necessary covariance matrices (Muthén and Muthén 1998–2006). This way, in case of non-convergence, we were able to estimate more constrained models to obtain starting values for the less constrained models we were actually interested in.

As advised in the STREAMS-manual (Gustafsson and Stahl 2005), we applied following analysis procedure to each group of items. First, a conventional CFA was performed on the teacher-level data. To test whether significant differences existed between cycle-1 versus cycle-2 and -3, a multiple-group CFA was conducted. Next, MCFAs were estimated. The factor structure at the teacher level was first imposed based on the results of the conventional CFA, without estimating a particular factor structure at the school level (i.e., a saturated group level). After obtaining a satisfactory structure at the teacher level, the school-level structure was estimated as well.

Based on our expectations, different alternative models were tested. In addition, some modifications were also made according to modification indices and other model results such as correlations between latent variables. Modification indices demonstrate how much the model fit would improve in case a parameter would be added or freed, or when an equality restriction would be relaxed (Sörbom 1989). When evaluating the largest modification indices, considerable attention was given to the interpretation of the proposed modifications. Thus, model testing was partly confirmatory, partly exploratory.

To evaluate model fit, different fit indices were used. First, the chi-square index (i.e., Yuan-Bentler T_2* test statistic), which yields an indication of the difference between the observed and the predicted association matrix, was used. However, since this goodness-of-fit index has the undesirable property of being affected by sample size, it has been advised to look at several other categories of fit indices as well (e.g., Heck and Thomas 2000). Relative chi-square, chi-square divided by degrees of freedom, is one attempt to make it less dependent on sample size. This ratio should be between 2 and 3 to be indicative of good model fit (Schermelleh-Engel et al. 2003). A ratio between 3 and 4 indicates an acceptable model fit (see e.g., Yang 2003). Other model fit indices that take into account the difference between the observed and the model-implied association matrix are Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). RMSEA is relatively independent of sample size and favours parsimonious models. RMSEA-values smaller than 0.05 are indicative of good fit, while values between 0.05 and 0.08 are considered acceptable (e.g., Schermelleh-Engel et al. 2003). As for SRMR, Hu and Bentler (1999) have proposed a cut-off value of 0.08: Values lower than 0.08 are considered good. Two other model fit indices that have been advised by, among others, Schermelleh-Engel et al. concern the Tucker-Lewis Index (TLI) and the Comparative Fit Index (CFI). Both are descriptive measures that compare the fit of the estimated model with the fit of the independence model, which assumes that the observed indicators are measured without error. Analogous to RMSEA, both indices reward model parsimony and are not as dependent on sample size as the chi-square value. Values should be larger than 0.95 to be indicative of good model fit (Hu and Bentler 1999). It should be stressed that the combination of these fit indices was used in an indicative way, and not as an absolute criterion. Alternative nested models were compared by means of a chi-square difference test. Since MLR-estimation was used, the difference between two scaled chi-square values for nested models is not distributed as a chi-square. Therefore, chi-square values were corrected prior to performing a difference test (Muthén and Muthén 1998–2006).

In addition, with regard to the finally selected models, the percentage variance extracted and the internal consistency were computed for each factor at the teacher level and at the school level. The percentage variance extracted reflects the overall amount of variance in the items accounted for by the latent construct. Higher percentages indicate a better representation of the latent construct by the items. Guidelines suggest values above 0.50. The internal consistency was measured by means of standardized Cronbach's alpha, which is a function of both the number of items and the inter-item correlations (Hair et al. 1998). Confidence intervals for Cronbach's alpha were determined using ScoreRel CI (Barnette 2005).

3 Results

3.1 Data screening and assumptions

Prior data screening was performed on the total set of 66 items. Missing data for each item ranged from 1.00% until 10.00%. Little's MCAR test in SPSS 15.0 indicated that these missings were not *missing completely at random* ($\chi^2_{17665} = 18264.18$, $p < 0.001$), meaning that missingness was related either to observed variables or to unobserved variables. However, MLR-estimation minimally requires *missing at random* (see Enders 2006). Collins et al. (2001) have suggested including variables that are correlates of the missingness and/or of the variables of interest to help reduce bias. These variables can be part of the analysis. Sufficiently high correlations between our items were found to justify the use of MLR.

Since investigations of both the histogram and the skewness and kurtosis values revealed a departure from normality for the majority of our items, MLR-estimation was chosen.

Next, the design effect was calculated. ICCs for each separate item ranged from 0.03 to 0.33 (see Appendix 2). An average cluster size of 13.61 needs an ICC larger than 0.08 to justify a multilevel analysis. Thus, 25 items with an ICC smaller than 0.08 were excluded, while retaining 41 of the 66 original items.

Finally, the amount of multicollinearity between the observed items was evaluated to test the appropriateness of a factor analysis. This was investigated using Kaiser's Measure of Sampling Adequacy (MSA) (Hair et al. 1998). Regarding each of the correlation matrices, all 41 items had an MSA-value larger than 0.50 and the overall MSA amounted to very good values above 0.85.

Thus, after performing data screening and splitting up our data, we obtained two data sets of 21 and 20 items respectively. Only teachers with missings on all items were deleted from subsequent analyses. A total number of 1538 teachers were included in the first data set on school climate, while 1540 teachers were retained in the second data set on school organizational characteristics.

3.2 Conventional confirmatory factor analysis

3.2.1 CFA 1 on school climate

Regarding school climate, several models were tested based on our expectations. For each of the analyses reported in Sect. 3, the scale of the latent variables was defined by means of the highest loading item whenever possible (i.e., unstandardized loading fixed to 1.0). In addition, correlations between the latent variables were estimated and evaluated afterwards.

First, we postulated a model with two latent variables, consistent with the aspects of the teachers' daily school life we outlined in Sect. 2: 'social relationships among school staff' and 'establishment of rules'. Since the social relationships among school staff concerned both relationships among teachers, and between teachers and the principal, we also tested a model with three latent variables: 'social relationships among teachers', 'social relationships between teachers and the principal', and 'establishment of rules'. Next, we also subdivided the items concerning establishment of rules to obtain a model with four latent variables: 'social relationships among teachers', 'social relationships between teachers and the principal', 'establishment of rules concerning tests and homework', and 'establishment of rules concerning students' behaviour'. Every addition of a latent variable resulted in a significantly better chi-square test statistic (see Table 1).

Table 1 Alternative CFA-models concerning school climate: model fit

Model	χ^2	df	p	$\Delta\chi^{2a}$	Δdf	p
Model comparison						
2 LVs	3349.636	188	<0.0001	–	–	–
3 LVs	1843.888	186	<0.0001	1490.840	2	<0.001
4 LVs	1493.874	183	<0.0001	285.493	3	<0.001
5 LVs	1043.509	179	<0.0001	328.614	4	<0.001
5 LVs + adjustments	562.189	176	<0.0001	491.645	3	<0.001
5 LVs + adjustments + equality constraints	614.473	185	<0.0001	42.097	9	< 0.001

Model	χ^2/df	TLI	CFI	RMSEA	SRMR
Other model fit indices					
2 LVs	17.817	0.621	0.661	0.105	0.083
3 LVs	9.913	0.799	0.822	0.076	0.067
4 LVs	8.163	0.838	0.859	0.068	0.059
5 LVs	5.830	0.891	0.907	0.056	0.053
5 LVs + adjustments	3.194	0.951	0.959	0.038	0.038
5 LVs + adjustments + equality constraints	3.322	0.948	0.954	0.039	0.044

LVs latent variables

[a] Chi-square difference test: in comparison with previous model

However, model fit could still be improved substantially. Therefore, we consulted the modification indices of the model with four latent variables. Both the size of these indices and the theoretical interpretation of the suggested changes were taken into account. The modification indices suggested the subdivision of ‘social relationships between teachers and the principal’ into ‘general social relationships between teachers and the principal’ and ‘social relationships between teachers and the principal concerning educational aspects’. Again, a significantly better fit was obtained. Further adjustments were made to this five-factor model: Three residual covariances were added. These residual covariances indicated that between three pairs of items, some common element was present that was not taken into account by the latent variable. They were only added when yielding a plausible interpretation. Obviously, this concerned a subjective aspect of the analysis. Residual covariances were added for the items 118 and 119 (StdYX $= 0.54$; $t = 19.55$), the items 70 and 79 (StdYX $= 0.20$; $t = 8.41$), and the items 71 and 73 (StdYX $= 0.12$; $t = 6.20$) (see explanation below for StdYX). This model yielded a good global fit (5 LVs + adjustments; see Table 1).

Since the estimation of the MCFA required additional equality constraints on the unstandardized factor loadings at the teacher level to assure model identification, these were tested in the conventional CFA as well. For instance, the unstandardized factor loadings of the items 75, 81 and 86 on ‘social relationships between teachers and the principal concerning educational aspects’ were constrained to be equal. Although this resulted in a slightly worse model fit, the absolute model fit values did not change considerably and were still indicative of a good model fit (5 LVs + adjustments + equality constraints; see Table 1). Thus, these additional equality constraints were retained.

A multi-group analysis showed no significant differences between cycle-1 teachers and cycle-2 and -3 teachers, justifying the use of one global CFA across all three cycles.

The results of the final model with equality constraints are shown in Table 2. Pattern coefficients are factor loadings that represent the unique contribution of each indicator to the factor, thus accounting for the inter-factor correlations (Hair et al. 1998). The completely standardized solution is shown, which refers to a standardization based on the variances of both the latent factors and the observed indicators (i.e., StdYX; Muthén and Muthén 1998–2006).

3.2.2 CFA 2 on school organizational characteristics

Based on our expectations, several alternative models were tested. First, a simple model with two latent variables was evaluated, again consistent with the aspects of the teachers' daily school life we outlined in Sect. 2: 'cooperation between teachers' and 'decision making process'. Next, 'cooperation between teachers' was split up into 'cooperation between teachers of the same subject' and 'cooperation between teachers of the same class'. This yielded a significantly better chi-square value (see Table 3). Subsequently, 'decision making process' was also subdivided into 'participatory decision making concerning the organization of education' and 'autonomy in decision making concerning tests and homework'. Again, a significantly better model fit was obtained.

Since there was still room for further improvement, modification indices were evaluated. These suggested separating 'cooperation between teachers of the same class concerning class aspects' and 'cooperation between teachers of the same class concerning student aspects', which yielded a model with five latent variables. Although chi-square improved significantly, further adjustments were necessary. Again, modification indices indicated splitting up a latent variable. Thus, 'cooperation between teachers of the same subject' was divided into 'cooperation between teachers of the same subject concerning didactical aspects' and 'cooperation between teachers of the same subject concerning the subject content'. Two items were allowed to load on both factors, items 10 and 18, which resulted in a significantly better chi-square value. Several further adjustments were made to this model with six latent variables: Three plausible residual covariances were added, another two items were allowed to load on two factors (i.e., items 53 and 56), and one item was allowed to load on a third factor (i.e., item 18). The residual covariances referred to the items 13 and 67 (StdYX $= -0.20$; $t = -9.30$), the items 66 and 68 (StdYX $= 0.27$; $t = 9.46$), and the items 54 and 55 (StdYX $= 0.16$; $t = 7.13$). This model yielded an acceptable model fit (see Table 3).

Analogous to the MCFA on school climate, the estimation of the MCFA on the school organizational characteristics required additional equality constraints at the teacher level to assure model identification. Regarding the second MCFA, not only several unstandardized factor loadings were constrained to be equal, but also residual variances of the observed items. Thus, each of these constraints was also tested in the conventional CFA. For instance, the unstandardized factor loadings of items 47, 53 and 54 on 'participatory decision making concerning the organization of education', and the residual variances of the items 27 and 28 were constrained to be equal. Since this resulted in a non-significant factor loading of the item 10 on 'cooperation between teachers of the same subject concerning didactical aspects' (i.e., $t < 2$), this relationship was deleted from the final model. Altogether, these constraints resulted in a slightly worse model fit, but the absolute model fit values did not change considerably and were still indicative of a good model fit (6 LVs + adjustments + equality constraints; see Table 3). Therefore, the additional constraints were retained.

Table 2 Conventional CFA on school climate: results of the selected model

Factors	I	II	III	IV	V	
% Variance extracted	53.03	60.70	42.23	23.06	31.80	
Standardized Cronbach's alpha (95% confidence interval)	0.85 (0.84–0.86)	0.86 (0.85–0.87)	0.74 (0.72–0.76)	0.60 (0.57–0.63)	0.57 (0.53–0.60)	
Number of items	5	4	4	5	3	
Item no.	**Concerning**	**Pattern coefficients[a,b] Completely standardized solution**				h^{2c}
82	Social, among teachers	0.80 (–)				0.64
71	Social, among teachers	0.77 (–)				0.59
77	Social, among teachers	0.76 (–)				0.58
73	Social, among teachers	0.74 (33.11)				0.55
80	Social, among teachers	0.54 (19.25)				0.29
74	Social, principal, education		0.85 (–)			0.73
75	Social, principal, education		0.76 (42.53)			0.58
86	Social, principal, education		0.75 (42.53)			0.57
81	Social, principal, education		0.74 (42.53)			0.55
83	Social, principal, general			0.74 (–)		0.55
72	Social, principal, general			0.73 (–)		0.54
79	Social, principal, general			0.68 (–)		0.47
70	Social, principal, general			0.37 (12.03)		0.14
112	Rules, students' behaviour				0.71 (–)	0.51
118	Rules, students' behaviour				0.45 (14.05)	0.21

Table 2 continued

Factors	I	II	III	IV	V		
% Variance extracted	53.03	60.70	42.23	23.06	31.80		
Standardized Cronbach's alpha (95% confidence interval)	0.85 (0.84–0.86)	0.86 (0.85–0.87)	0.74 (0.72–0.76)	0.60 (0.57–0.63)	0.57 (0.53–0.60)		
Number of items	5	4	4	5	3		
Item no.	Concerning	Pattern coefficients[a,b] Completely standardized solution					h^{2c}
109	Rules, students' behaviour				0.42 (12.54)		0.18
111	Rules, students' behaviour				0.40 (14.05)		0.16
119	Rules, students' behaviour				0.32 (12.54)		0.10
115	Rules, tests and homework					0.64 (–)	0.41
113	Rules, tests and homework					0.60 (12.15)	0.36
114	Rules, tests and homework					0.43 (12.15)	0.18
	Inter-factor correlations[a]						
	I	II	III	IV	V		
I	–						
II	0.59 (15.24)	–					
III	0.53 (13.35)	0.73 (17.42)	–				
IV	0.15 (4.37)	0.26 (6.88)	0.20 (5.13)	–			
V	0.16 (4.13)	0.15 (3.96)	0.08 (1.92)	0.33 (8.39)	–		

[a] t-Values are in brackets; (–) means the t-value was not estimated, since the respective item(s) was (were) used to define the scale of the factor
[b] Blanks refer to pattern coefficients constrained to be zero
[c] h^2 = communality coefficient

Table 3 Alternative CFA-models concerning school organizational characteristics: model fit

Model	χ^2	df	p	$\Delta\chi^{2a}$	Δdf	p
Model comparison						
2 LVs	2868.278	169	<0.0001	–	–	–
3 LVs	2360.083	167	<0.0001	306.234	2	<0.001
4 LVs	1756.952	164	<0.0001	361.373	3	<0.001
5 LVs	1273.752	160	<0.0001	457.143	4	<0.001
6 LVs	914.945	153	<0.0001	286.130	7	<0.001
6 LVs + adjustments	507.620	147	<0.0001	371.308	6	<0.001
6 LVs + adjustments + equality constraints	550.813	161	<0.0001	39.975	14	<0.001
Model	**χ^2/df**	**TLI**	**CFI**	**RMSEA**	**SRMR**	
Other model fit indices						
2 LVs	16.972	0.606	0.649	0.102	0.104	
3 LVs	14.132	0.676	0.715	0.092	0.097	
4 LVs	10.713	0.760	0.793	0.079	0.069	
5 LVs	7.961	0.828	0.855	0.067	0.053	
6 LVs	5.980	0.877	0.901	0.057	0.046	
6 LVs + adjustments	3.453	0.939	0.953	0.040	0.033	
6 LVs + adjustments + equality constraints	3.421	0.940	0.949	0.040	0.036	

LVs latent variables

[a] Chi-square difference test: in comparison with previous model

A multi-group analysis revealed evidence for some differences at the 0.01-level between cycle-1 teachers and cycle-2 and -3 teachers regarding factor loadings and residual variances of the observed items. However, inspection of the results only showed rather small differences: The largest completely standardized difference in factor loadings amounted to 0.098, and the largest completely standardized difference in residual variance amounted to 0.111. Furthermore, all factor loadings were significant and the relative importance of the items to the factor was similar in both groups. Therefore, we decided to use one global CFA for both groups of teachers. The results of the final model are summarized in Table 4.

3.3 Two-level confirmatory factor analysis

3.3.1 MCFA 1 on school climate

Regarding the model at the teacher level, the same factor structure as in the conventional CFA was confirmed (SRMR-within = 0.034–0.036; see Table 6). All estimated factor loadings remained significant, and modification indices did not indicate any further improvements. In addition, the estimates of the residual covariances were still significant: for the items 118 and 119 (StdYX = 0.50; t = 16.50), for the items 70 and 79 (StdYX = 0.17; t = 6.96), and for the items 71 and 73 (StdYX = 0.13; t = 7.01). The only necessary adjustment at the teacher level concerned constraining one insignificant inter-factor correlation to zero to ensure model identification. The results of this final model at the teacher level are presented in Table 5.

Table 4 Conventional CFA on school organizational characteristics: results of the selected model

Factors	I	II	III	IV	V	VI	
% Variance extracted	36.12	46.55	52.26	56.29	39.10	22.99	
Standardized Cronbach's alpha (95% confidence interval)	0.72 (0.70–0.74)	0.70 (0.68–0.73)	0.69 (–)[a]	0.72 (–)	0.79 (0.78–0.81)	0.57 (0.54–0.61)	
Number of items	6	3	2	2	7	4	
Item no.	Concerning	Pattern coefficients[b,c] Completely standardized solution					h^{2d}
9	Meetings, subj., cont.	0.81 (32.16)					0.66
13	Meetings, subj., cont.	0.73 (–)					0.54
10	Meetings, subj., cont.	0.68 (32.16)					0.46
18	Meetings, subj.	0.31 (6.36)	0.19 (3.83)			0.14 (5.20)	0.25
12	Meetings, subj., did.		0.85 (28.64)				0.73
11	Meetings, subj., did.		0.73 (–)				0.54
28	Meetings, class, stud.			0.72 (–)			0.52
27	Meetings, class, stud.			0.72 (24.37)			0.52
34	Meetings, class, class				0.78 (–)		0.61
35	Meetings, class, class				0.72 (23.40)		0.52
47	Partic. dec. mak.					0.77 (–)	0.59
53	Partic. dec. mak.	0.15 (7.99)				0.75 (–)	0.61
54	Partic. dec. mak.					0.73 (–)	0.53
55	Partic. dec. mak.					0.62 (26.27)	0.38
56	Partic. dec. mak.	0.15 (7.99)				0.54 (26.27)	0.33
57	Partic. dec. mak.					0.49 (18.96)	0.24

Table 4 continued

Factors	I	II	III	IV	V	VI		
% Variance extracted	36.12	46.55	52.26	56.29	39.10	22.99		
Standardized Cronbach's alpha (95% confidence interval)	0.72 (0.70–0.74)	0.70 (0.68–0.73)	0.69 (–)[a]	0.72 (–)	0.79 (0.78–0.81)	0.57 (0.54–0.61)		
Number of items	6	3	2	2	7	4		
Item no.	Concerning	Pattern coefficients[b,c] Completely standardized solution						h^{2d}
67	Auton. dec. mak.						0.69 (14.02)	0.47
69	Auton. dec. mak.						0.50 (–)	0.25
68	Auton. dec. mak.						0.34 (14.02)	0.11
66	Auton. dec. mak.						0.29 (7.83)	0.09
	Inter-factor correlations[b]							
	I	II	III	IV	V	VI		
I	–							
II	0.70 (18.33)	–						
III	0.20 (5.81)	0.14 (3.93)	–					
IV	0.14 (3.86)	0.22 (5.88)	0.06 (1.50)	–				
V	0.08 (2.25)	0.17 (4.98)	0.22 (6.38)	0.31 (8.30)	–			
VI	−0.62 (−10.63)	−0.37 (−8.28)	−0.08 (−1.99)	−0.05 (−1.10)	−0.07 (−1.79)	–		

[a] (–) Means the 95% confidence interval could not be estimated due to the small number of items
[b] t-Values are in brackets; (–) here means that the t-value was not estimated, since the respective item(s) was (were) used to define the scale of the factor
[c] Blanks refer to pattern coefficients constrained to be zero
[d] h^2 = communality coefficient

Table 5 MCFA on school climate: results of the selected model at the teacher level

Factors	I	II	III	IV	V	
% Variance extracted	46.82	56.85	40.21	18.74	31.35	
Standardized Cronbach's alpha (95% confidence interval)	0.82 (0.80–0.83)	0.84 (0.83–0.85)	0.73 (0.71–0.75)	0.57 (0.53–0.60)	0.55 (0.51–0.59)	
Number of items	5	4	4	5	3	
Item no.	Concerning	Pattern coefficients[a,b] Completely standardized solution				h^{2c}
82	Social, among teachers	0.75 (–)				0.57
71	Social, among teachers	0.73 (–)				0.53
77	Social, among teachers	0.72 (–)				0.52
73	Social, among teachers	0.70 (26.35)				0.48
80	Social, among teachers	0.49 (15.48)				0.24
74	Social, principal, education		0.82 (–)			0.67
75	Social, principal, education		0.75 (35.45)			0.56
86	Social, principal, education		0.74 (35.45)			0.54
81	Social, principal, education		0.71 (35.45)			0.51
72	Social, principal, general			0.72 (–)		0.52
83	Social, principal, general			0.71 (–)		0.50
79	Social, principal, general			0.66 (–)		0.43
70	Social, principal, general			0.40 (12.82)		0.16
112	Rules, students' behaviour				0.63 (–)	0.40
111	Rules, students' behaviour				0.40 (10.81)	0.16
118	Rules, students' behaviour				0.40 (10.81)	0.16
109	Rules, students' behaviour				0.38 (8.99)	0.14

Table 5 continued

Factors	I	II	III	IV	V		
% Variance extracted	46.82	56.85	40.21	18.74	31.35		
Standardized Cronbach's alpha (95% confidence interval)	0.82 (0.80–0.83)	0.84 (0.83–0.85)	0.73 (0.71–0.75)	0.57 (0.53–0.60)	0.55 (0.51–0.59)		
Number of items	5	4	4	5	3		
Item no.	Concerning	Pattern coefficients[a,b] Completely standardized solution					h^{2c}
119	Rules, students' behaviour				0.28 (8.99)		0.08
115	Rules, tests and homework					0.67 (–)	0.44
113	Rules, tests and homework					0.56 (11.52)	0.32
114	Rules, tests and homework					0.43 (11.52)	0.18
	Inter-factor correlations[a]						
	I	II	III	IV	V		
I	–						
II	0.56 (13.27)	–					
III	0.52 (12.14)	0.73 (14.86)	–				
IV	0.17 (5.14)	0.32 (6.90)	0.25 (5.55)	–			
V	0.12 (2.97)	0.14 (4.36)	0.00 (0.00)	0.30 (6.57)	–		

[a] t-Values are in brackets; (–) means the t-value was not estimated, since the respective item(s) was (were) used to define the scale of the factor
[b] Blanks refer to pattern coefficients constrained to be zero
[c] h^2 = communality coefficient

Table 6 Alternative MCFA-models concerning school climate: model fit

Model	χ^2	df	χ^2/df	$\Delta\chi^{2a}$	Δdf	p
Model comparison						
W 5 LVs; B 1 LV	1413.921	381	3.711	–	–	–
W 5 LVs; B 2 LVs	1214.331	380	3.196	20.364	1	<0.001
W 5 LVs; B 3 LVs	1199.319	378	3.173	14.216	2	<0.001
W 5 LVs; B 4 LVs	1039.016	375	2.771	115.826	3	<0.001
W 5 LVs; B 5 LVs	1001.041	371	2.698	54.328	4	<0.001
Model	TLI	CFI	RMSEA (total)	RMSEA (between)	SRMR (within)	SRMR (between)
Other model fit indices						
W 5 LVs; B 1 LV	0.892	0.902	0.042	0.204	0.036	0.287
W 5 LVs; B 2 LVs	0.912	0.921	0.038	0.170	0.035	0.191
W 5 LVs; B 3 LVs	0.913	0.922	0.038	0.168	0.035	0.181
W 5 LVs; B 4 LVs	0.929	0.937	0.034	0.143	0.034	0.179
W 5 LVs; B 5 LVs	0.932	0.940	0.033	0.140	0.034	0.148

W within-level or teacher level, *LV(s)* latent variable(s), *B* between-level or school level

[a] Chi-square difference test: in comparison with previous model

Concerning the model at the school level, several models were tested. These alternative models were equivalent to those tested with the conventional CFA, albeit in a slightly different order. Apart from the equality constraints at the teacher level, several constraints had to be imposed at the school level as well. First, different sets of factor loadings were constrained to be equal, such as the unstandardized loadings of the items 81 and 86. This was necessary to yield model identification, since we only had a limited number of groups (i.e., $n = 113$). Next, some residual variances of the observed items were constrained to be zero in case these values were very small and negative. This promoted model convergence, since negative variances are statistically impossible. These constraints did not produce a significantly worse model fit. Importantly, these constraints were equal across the different rival models allowing chi-square difference testing on nested models.

Based on the few MCFA-studies conducted so far, we expected a model with fewer factors than at the teacher level. Therefore, we started with a model including one general factor: 'school climate'. However, this model yielded a bad fit (see Table 6). Thus, this general factor was split up into 'social relationships among school staff' and 'establishment of rules'. Further model fit improvement appeared necessary. In a third and fourth step, both factors were split up, with each step yielding a significant model fit improvement. Thus thirdly, 'establishment of rules' was divided into 'establishment of rules concerning tests and homework' and 'establishment of rules concerning students' behaviour'. Fourth, 'social relationships among school staff' was separated into 'social relationships among teachers' and 'social relationships between teachers and the principal'. Nevertheless, model fit could still be improved substantially at the school level. An inspection of the same five latent variables than at the teacher level again resulted in a significantly smaller chi-square value. This way, 'social relationships between teachers and the principal' was further divided into 'general social relationships between teachers and the principal' and 'social relationships between teachers and the principal concerning educational aspects'.

In spite of these additional factors, model fit improvement was still required at the school level, while the model fit indices regarding the global model (i.e., within + between) were much better (see Table 6). Thus, modification indices were inspected, but these did not indicate any further plausible improvements. Therefore, residuals were consulted. These represent the difference between the observed and the estimated covariance matrix. Surprisingly, no severe misspecifications were found at the school level. In fact, residuals at the school level were even as small as those at the teacher level. Possibly, the much smaller sample size at the group level caused the model fit indices to be less good. This will be further considered in Sect. 4. Owing to the small residuals at the school level, the results of this final model are reported in Table 7.

3.3.2 MCFA 2 on school organizational characteristics

With regard to the model at the teacher level, identically the same factor model as in the conventional CFA was found (SRMR-within = 0.042–0.043; see Table 9). The same factors appeared to be relevant: All factor loadings remained significant, and no further plausible improvements were indicated by the modification indices. Additionally, the residual covariances were still important: for the items 54 and 55 (StdYX = 0.14; $t = 6.54$), for the items 13 and 67 (StdYX = −0.21; $t = -8.93$), and for the items 66 and 68 (StdYX = 0.23; $t = 6.94$). To promote model identification, all inter-factor correlations were only estimated regarding the finally selected model. As regards the other models, inter-factor correlations that were not significant within conventional CFA were constrained to be zero at the teacher level within MCFA. Therefore, model fit improved considerably in the final MCFA-model (SRMR-within = 0.031; see Table 9). The results of this final model at the teacher level are depicted in Table 8.

Regarding the model at the school level, again different rival models were tested. The models with two and three latent variables tested here differed to some extent regarding their factor structure from the models hypothesized in the conventional CFA. Equivalent to the first MCFA, several constraints had to be imposed on the school level to promote both model identification and model convergence, on top of the constraints at the teacher level. First, several factor loadings were constrained to be equal, for instance between the items 34 and 35. Second, some constraints were put on the residual variances. Since some of the residual variance estimates had a small negative value, these were constrained to be zero. None of these constraints resulted in a significantly worse model fit at the school level. To test the rival models by means of chi-square difference testing, these constraints were held equal across models even when this appeared unnecessary. Concerning the finally chosen model, as much constraints as possible were deleted.

Similar to the MCFA on school climate, we assumed a smaller number of factors at the school level in comparison with the teacher level. Therefore, we started by postulating one general factor: ‘organizational school climate’. This model yielded a bad fit (see Table 9), and it proved a significant improvement to divide this general factor into ‘consultation between teachers’ and ‘participatory decision making concerning the organization of education’. The first factor, which was guided by the modification indices, incorporated both items concerning cooperation between teachers and items about teachers’ autonomy in decision making concerning tests and homework. Nonetheless, model fit could still be improved substantially. Thus, a model with three factors was evaluated. Here, ‘consultation between teachers’ was split up into ‘cooperation between teachers of the same class’ and ‘cooperation and shared decision making concerning didactics’. This last factor included both items about cooperation between teachers of the same subject and items related to teachers’ autonomy in decision

Table 7 MCFA on school climate: results of the selected model at the school level

Factors	I	II	III	IV	V	
% Variance extracted	94.22	83.45	58.09	62.85	37.53	
Standardized Cronbach's alpha (95% confidence interval)	0.98 (0.97–0.99)	0.94 (0.93–0.96)	0.83 (0.77–0.88)	0.77 (0.69–0.83)	0.66 (0.53–0.75)	
Number of items	5	4	4	5	3	
Item no.	Concerning	Pattern coefficients[a,b] Completely standardized solution				h^{2c}
71	Social, among teachers	1.00 (–)				0.999
73	Social, among teachers	0.99 (22.89)				0.98
82	Social, among teachers	0.99 (22.89)				0.97
80	Social, among teachers	0.94 (9.62)				0.89
77	Social, among teachers	0.94 (11.00)				0.87
74	Social, principal, education		1.00 (–)			0.999
81	Social, principal, education		0.92 (13.80)			0.85
75	Social, principal, education		0.89 (13.41)			0.80
86	Social, principal, education		0.83 (13.80)			0.69
83	Social, principal, general			0.98 (6.41)		0.96
72	Social, principal, general			0.87 (–)		0.76
79	Social, principal, general			0.75 (5.52)		0.57
70	Social, principal, general			0.21 (0.93)		0.05
112	Rules, students' behaviour				0.999 (–)	0.999
118	Rules, students' behaviour				0.998 (2.08)	0.995
109	Rules, students' behaviour				0.85 (7.88)	0.72
111	Rules, students' behaviour				0.61 (3.99)	0.37

Table 7 continued

Factors	I	II	III	IV	V		
% Variance extracted	94.22	83.45	58.09	62.85	37.53		
Standardized Cronbach's alpha (95% confidence interval)	0.98 (0.97–0.99)	0.94 (0.93–0.96)	0.83 (0.77–0.88)	0.77 (0.69–0.83)	0.66 (0.53–0.75)		
Number of items	5	4	4	5	3		
Item no.	Concerning	Pattern coefficients[a,b] Completely standardized solution					h^{2c}
119	Rules, students' behaviour				0.25 (0.62)		0.06
114	Rules, tests and homework					0.71 (–)	0.50 [a]
113	Rules, tests and homework					0.61 (1.06)	0.37
115	Rules, tests and homework					0.51 (1.41)	0.26
	Inter-factor correlations[a]						
	I	II	III	IV	V		
I	–						
II	0.64 (5.13)	–					
III	0.55 (3.83)	0.69 (4.52)	–				
IV	0.11 (0.90)	0.14 (1.09)	0.03 (0.28)	–			
V	0.40 (0.97)	0.29 (1.11)	0.41 (2.27)	0.51 (2.30)	–		

t-Values are in brackets; (–) means the t-value was not estimated, since the respective item(s) was (were) used to define the scale of the factor

[b] Blanks refer to pattern coefficients constrained to be zero

[c] h^2 = communality coefficient

Table 8 MCFA on school organizational characteristics: results of the selected model at the teacher level

Factors	I	II	III	IV	V	VI	
% Variance extracted	29.95	45.90	50.57	54.11	37.40	17.91	
Standardized Cronbach's alpha (95% confidence interval)	0.67 (0.64–0.70)	0.67 (0.64–0.70)	0.67 (–)[a]	0.70 (–)	0.78 (0.77–0.80)	0.50 (0.46–0.54)	
Number of items	6	3	2	2	7	4	
Item no.	**Concerning**	**Pattern coefficients[b,c] Completely standardized solution**					**h^{2d}**
9	Meetings, subj., cont.	0.76 (–)					0.57
13	Meetings, subj., cont.	0.63 (20.57)					0.40
10	Meetings, subj., cont.	0.62 (–)					0.38
18	Meetings, subj.	0.36 (5.80)	0.13 (2.06)			0.13 (4.32)	0.24
12	Meetings, subj., did.		0.84 (24.08)				0.70
11	Meetings, subj., did.		0.75 (–)				0.56
28	Meetings, class, stud.			0.73 (–)			0.53
27	Meetings, class, stud.			0.69 (16.69)			0.48
34	Meetings, class, class				0.77 (–)		0.59
35	Meetings, class, class				0.70 (20.56)		0.49
47	Partic. dec. mak.					0.74 (–)	0.55
53	Partic. dec. mak.	0.08 (4.09)				0.73 (–)	0.56
54	Partic. dec. mak.					0.72 (–)	0.51
55	Partic. dec. mak.					0.61 (23.07)	0.37
56	Partic. dec. mak.	0.09 (4.09)				0.57 (23.07)	0.35
57	Partic. dec. mak.					0.45 (14.09)	0.20

Table 8 continued

Factors	I	II	III	IV	V	VI		
% Variance extracted	29.95	45.90	50.57	54.11	37.40	17.91		
Standardized Cronbach's alpha (95% confidence interval)	0.67 (0.64–0.70)	0.67 (0.64–0.70)	0.67 (–)[a]	0.70 (–)	0.78 (0.77–0.80)	0.50 (0.46–0.54)		
Number of items	6	3	2	2	7	4		
Item no.	Concerning	Pattern coefficients[b,c] Completely standardized solution						h^{2d}
67	Auton. dec. mak.						0.61 (7.32)	0.38
69	Auton. dec. mak.						0.43 (–)	0.18
68	Auton. dec. mak.						0.28 (7.32)	0.08
66	Auton. dec. mak.						0.28 (5.20)	0.08
	Inter-factor correlations[b]							
	I	II	III	IV	V	VI		
I	–							
II	0.72 (14.10)	–						
III	0.24 (6.08)	0.12 (3.37)	–					
IV	0.27 (6.01)	0.30 (7.90)	0.20 (4.05)	–				
V	0.16 (3.74)	0.20 (5.25)	0.25 (6.77)	0.36 (8.57)	–			
VI	−0.42 (−6.21)	−0.27 (−5.15)	−0.05 (−1.09)	−0.14 (−2.75)	−0.15 (−3.24)	–		

[a] (–) Means the 95% confidence interval could not be estimated due to the small number of items
[b] t-Values are in brackets; (–) here means that the t-value was not estimated, since the respective item(s) was (were) used to define the scale of the factor
[c] Blanks refer to pattern coefficients constrained to be zero
[d] h^2 = communality coefficient

Table 9 Alternative MCFA-models concerning school organizational characteristics: model fit

Model	χ^2	df	χ^2/df	$\Delta\chi^{2a}$	Δdf	p
Model comparison						
W 6 LVs; B 1 LV	1186.685	348	3.410	–	–	–
W 6 LVs; B 2 LVs	1085.719	347	3.129	615.646	1	<0.001
W 6 LVs; B 3 LVs	998.763	345	2.895	70.985	2	<0.001
W 6 LVs; B 4 LVs	997.661	342	2.917	1.111	3	>0.05
W 6 LVs; B 6 LVs	972.172	332	2.928	23.684	10	<0.01
W 6 LVs; B 3 LVs + adjustments	901.860	331	2.725	92.863*	14	<0.001
Model	**TLI**	**CFI**	**RMSEA (total)**	**RMSEA (between)**	**SRMR (within)**	**SRMR (between)**
Other model fit indices						
W 6 LVs; B 1 LV	0.893	0.902	0.040	0.171	0.043	0.260
W 6 LVs; B 2 LVs	0.905	0.913	0.037	0.149	0.042	0.210
W 6 LVs; B 3 LVs	0.915	0.923	0.035	0.133	0.042	0.188
W 6 LVs; B 4 LVs	0.914	0.923	0.035	0.136	0.042	0.184
W 6 LVs; B 6 LVs	0.914	0.925	0.035	0.138	0.042	0.165
W 6 LVs; B 3 LVs + adjustments	0.923	0.933	0.033	0.122	0.031	0.176

W within-level or teacher level, *LV(s)* latent variable(s), *B* between-level or school level
* Chi-square difference test in comparison with B 3 LVs
[a] Chi-square difference test: in comparison with previous model, except *

making concerning tests and homework. Again, a significantly better model fit was obtained. In a next step, a model with four latent variables was presupposed. Now, 'cooperation and shared decision making concerning didactics' was broken into 'cooperation between teachers of the same subject' and 'autonomy in decision making concerning tests and homework'. However, this did not result in a significantly better fitting model. In a final step, the model with six factors that was selected at the teacher level was also evaluated at the school level. In comparison with the 4-factor model, this resulted in a slightly better fit. On the other hand, both RMSEA and TLI, which favour parsimony, indicated the model with three latent variables. Therefore, the model with three latent variables was preferred and further improved.

Guided by the modification indices, two extra parameters were added to the selected model: a residual covariance between the items 66 and 68, and an extra factor loading of item 56 on 'cooperation and shared decision making concerning didactics'. Both parameters indeed appeared to be significant. The completely standardized estimate of the residual covariance between items 66 and 68 amounted to 0.51 ($t = 3.90$). Other adjustments concerned the removal of several constraints on factor loadings and residual variances, and, as already mentioned, the estimation of all inter-factor correlations at the teacher level. Again, this yielded a substantial improvement in model fit. However, equivalent to the results of the first MCFA, the model fit at the school level was not yet acceptable. Therefore, residuals were also evaluated. Unexpectedly, these were even much better for the school level than for the teacher level. The residuals at the school level did not indicate any severe misspecification. Possibly, the lower sample size at the group level again caused the model fit indices to indicate a rather unacceptable model fit. Nonetheless, the goodness-of-fit indices concerning the entire model were quite good. Consequently, the results of this 3-factor model are reported in Table 10.

Table 10 MCFA on school organizational characteristics: results of the selected model at the school level

Factors	I	II	III
% Variance extracted	69.84	73.32	63.98
Standardized Cronbach's alpha (95% confidence interval)	0.94 (0.93–0.96)	0.86 (0.81–0.90)	0.89 (0.86–0.92)
Number of items	11	4	6

Item no.	Concerning	Pattern coefficients[a,b] Completely standardized solution			h^{2c}
11	Meetings, subject	0.999 (5.97)			0.998
12	Meetings, subject	0.998 (5.12)			0.996
9	Meetings, subject	0.99 (13.19)			0.98
69	Auton. dec. mak.	−0.97 (−10.83)			0.95
13	Meetings, subject	0.96 (–)			0.93
67	Auton. dec. mak.	−0.92 (−13.95)			0.85
10	Meetings, subject	0.84 (7.25)			0.71
68	Auton. dec. mak.	−0.67 (−5.66)			0.46
18	Meetings, subject	0.61 (3.16)			0.37
66	Auton. dec. mak.	−0.37 (−3.36)			0.14
35	Meetings, class		−0.93 (−2.74)		0.86
28	Meetings, class		0.88 (6.56)		0.77
34	Meetings, class		−0.83 (−2.74)		0.70
27	Meetings, class		0.78 (–)		0.61
47	Partic. dec. mak.			0.97 (–)	0.94
53	Partic. dec. mak.			0.94 (2.95)	0.88
57	Partic. dec. mak.			0.84 (2.49)	0.71
54	Partic. dec. mak.			0.83 (12.05)	0.68
55	Partic. dec. mak.			0.57 (2.49)	0.33
56	Partic. dec. mak.	0.46 (4.88)		0.45 (2.49)	0.36

	Inter-factor correlations[a]		
	I	II	III
I	–		
II	0.26 (2.26)	–	
III	−0.13 (−0.54)	0.11 (0.51)	–

[a] t-Values are in brackets; (–) means the t-value was not estimated, since the respective item(s) was (were) used to define the scale of the factor
[b] Blanks refer to pattern coefficients constrained to be zero
[c] h^2 = communality coefficient

4 Discussion and conclusion

To assess the potential of MCFA to construct measures for school process variables in educational effectiveness research, several aspects must be compared accurately between the different types of factor analyses we conducted. First, we will discuss the differences we found concerning our data set between conventional CFA and MCFA at the teacher level. Next, we will focus on the differences between the teacher level and the school level within MCFA.

So first, when we want to construct underlying dimensions of individual teacher perceptions, does it matter whether we control for the nesting of teachers within schools? Concerning the factor structure, we found exactly the same factors for conventional CFA and MCFA at the teacher level. For both groups of items, regarding school climate and school organizational characteristics, the factors were composed of exactly the same items. Moreover, most items even appeared in the same order when comparing CFA with MCFA at the teacher level. Thus, items that were most crucial for an underlying dimension within CFA, also were most crucial for that same underlying dimension within teacher-level MCFA. On the contrary, a clear distinction related to the size of the factor loadings. Regarding MCFA at the teacher level, most items had a slightly lower factor loading. Analogously, the communalities of the indicators, the percentage variance extracted per factor, and the internal consistency of the factors were all slightly lower when taking into account the nesting of teachers, albeit all differences were rather small. This is due to a downward-correction of the covariance matrix when controlling for the dependency of the teachers' answers (Julian 2001; Muthén and Satorra 1995). The same phenomenon was observed when using exploratory factor analyses, where analyses were done across all 41 items (D'Haenens et al. 2008). Interestingly, as regards the inter-factor correlations, different conclusions could be drawn for each group of items. Concerning school climate, most correlations were slightly lower for teacher-level MCFA, while most inter-factor correlations about school organizational characteristics were slightly higher for teacher-level MCFA. Nevertheless, most differences were again rather small.

Next, results are compared between teacher-level MCFA and school-level MCFA. Since differences between conventional CFA and teacher-level MCFA were quite trivial, we decided to make the comparison of the results at the school level with the unbiased results of the teacher-level MCFA. Remarkably, considerable differences were found between both groups of items we analysed. First, the school-level MCFA of the items about school climate revealed exactly the same factors composed of the same items as the teacher-level MCFA. The only major difference in the factor structure concerned the order of the items within each factor, which was different for most items between both analyses. Thus, depending on the level of analysis the items had a different significance regarding their underlying dimension. Furthermore, the solution at the school level did not entail any error covariances, so this solution had a more simple structure than at the teacher level (i.e., with three error covariances) which certainly is to be preferred. Inspection of the size of the factor loadings revealed considerably larger loadings at the school level, although standard errors were also noticeably larger. This way, a few items turned insignificant at the school level ($t < 2$). Therefore, depending on the decision of individual researchers, a slightly different item composition could result at the school level, in case these insignificant items would be deleted. Surprisingly, two of these items did have a considerably large factor loading. The larger standard errors indicate a less stable solution at the school level, probably due to a much smaller sample size in comparison with the teacher level (Hair et al. 1998). Analogous to the size of the factor loadings, the communalities of the individual items, and both the percentage variance extracted and the internal consistency of the factors, were considerably higher at the school level. Finally, the inter-factor correlations differed markedly between each level of analysis. For example, the correlation between 'establishment of rules concerning students' behaviour' and 'establishment of rules concerning tests and homework' appeared much larger at the school level than at the teacher level. Again, larger standard errors were observed at the group level.

On the contrary, school-level analysis of the items about school organizational characteristics revealed a quite different picture. While six factors were identified at the teacher level, only three factors were confirmed at the school level. Two factors at the group level were clear combinations of several factors at the teacher level, whereas one factor was identical to the factor at the teacher level, albeit with a largely different order of the items. Equivalent to the first item set, a more clear factor structure was found at the group level, with only one error covariance and only one item with a cross-loading compared to three error covariances and three items with at least one cross-loading at the teacher level. As regards the size of the factor loadings, again we noticed much higher loadings at the school level, although in combination with larger standard errors. Here, all factor loadings were significant. Parallel to the item set about school climate, higher communalities, higher percentages variance extracted, and higher internal consistencies were observed at the school level, even regarding the factor that was identical between both levels. Although hard to compare straightforwardly, the more global school-level factors correlated rather low, while among the larger group of inter-factor correlations at the teacher level, at least some correlations were rather high. When using these measures in multilevel regression analysis to investigate for example the link with student achievement, low correlations are to be preferred (Shieh and Fouladi 2003).

In sum, almost negligible differences were found between the solutions of the conventional CFA and the solutions of the teacher-level MCFA. We did found smaller estimates concerning factor loadings, communalities, etc., due to the corrected inter-item covariances. These results are largely equivalent to the findings of our previous study (D'Haenens et al. 2008). The implications of these specific findings depend to a certain extent on the decision of how the scores on the underlying factors are computed. Does one use a simple sum score, or does one apply a more complex factor score based on the factor loadings (e.g., Grice 2001)? Does one first delete all the items with a factor loading smaller than a particular cut-off value, for instance |0.30|, or does one include all the items? These decisions are left over to the individual researcher, but they will definitely influence the implications of the results.

On the other hand, the results of the comparison between teacher-level MCFA and school-level MCFA were mixed to some extent. While the items about school climate yielded quite the same factors across levels, the items concerning school organizational characteristics yielded more global factors at the group level in comparison with the teacher level. Only this last finding is in line with our previous study (D'Haenens et al. 2008), as well as with some other studies applying MCFA within educational research (e.g., Holfve-Sabel and Gustafsson 2005). Identical conclusions concerning both groups of items were: solutions with a more simple structure, higher factor loadings, larger communalities, higher percentages variance extracted, and larger internal consistencies at the group level. Higher percentages of variance extracted indicate a better representation of the latent constructs at the school level. Obviously, implications are largely determined by the question whether the selected latent constructs matter for the learning of students. This will be addressed in a future study, linking the constructed measures for school process variables of both CFA and MCFA to student achievement and psycho-social outcomes.

Several limitations need to be taken into account. First, by splitting up the total set of 41 items into two separate groups of items, we precluded the possibility of covariances between items about school climate on the one hand and items about school organization on the other hand. Obviously, the selection of the items also partly determined the results we obtained. Second, due to a number of adaptations we made to the a priori tested models, our confirmatory analyses became to a certain extent exploratory. Therefore, a replication of our selected models within a new sample would be necessary. Third, although residuals at the school level did not reveal any severe misspecification, the model fit indices at the group level indicated a rather bad fit. Possibly, this is related to the much smaller sample size at the group level. Several simulation studies have already shown that less stringent cut-off values should be used in case of a smaller sample size (e.g., Sivo et al. 2006). By way of test, we conducted exploratory factor analyses at the school level for each of the groups of items. Even when specifying the resulting factor solutions in MCFA, we obtained RMSEA-values above 0.119. Therefore, we computed RMSEA at the group level, hypothesizing a higher sample with exactly the same chi-square value and degrees of freedom. For MCFA 1 concerning school climate, a sample of 500 schools would result in an acceptable RMSEA-value of 0.0668, and with a sample of 1000 groups we would obtain a good RMSEA-value of 0.0472. For MCFA 2 concerning school organizational characteristics, a sample of 500 schools would yield an acceptable RMSEA-value of 0.0578, and a good RMSEA-value of 0.0409 would be obtained with a sample of 1000 groups. Although this does not prove that the poor model fit at the group level was caused by the small sample size, we think that these issues show that sample size definitely needs to be taken into account when considering our results. Fourth, smaller sample sizes also make a factor analysis less stable. Therefore, replications using a larger sample size at the group level are in fact required. Finally, this study merely provides an illustration of different results that *might* be obtained depending on whether you take into account nesting of teachers within schools. Therefore, generalizations are impossible to make at this stage.

Acknowledgements The study reported in this manuscript was made possible by Research Grant 3H051166 of the Research Foundation—Flanders. The authors are also grateful to their colleagues of the Centre for Educational Effectiveness and Evaluation for their helpful comments.

Appendices

Appendix 1. Selected items of the 'School Characteristics Questionnaire for Teachers'

Cooperation between teachers
Meetings among teachers of the same subject
Response categories: (1) never, (2) sometimes, (3) mostly, (4) always
Number of items: 10
Items
How frequently do meetings take place among teachers of the same subject, concerning…
9. …the choice and planning of the study contents?
10. …the choice of the textbook?
11. …the choice of the didactical material?
12. …working with didactical methods and techniques?
13. …the construction of tests and exams?
14. …problems during teaching?
15. …the approach towards students with learning difficulties?
16. …training and educational initiatives organized by the school?

Appendix 1. continued

17. …training and educational initiatives organized by an authority other than the school?
18. …agenda items proposed or imposed by the principal?

Meetings among teachers of the same class

Response categories: (1) never, (2) sometimes, (3) mostly, (4) always

Number of items: 14

Items

How frequently do meetings take place among teachers of the same class, concerning…

27. …the study results obtained by the student?
28. …the student's personality characteristics (intelligence, character…)?
29. …the student's family background characteristics (socio-economic status, relationship between the parents and the students…)?
30. …a student's disturbing behaviour?
31. …the formulation of advices with regard to the study choices for all the students of the class?
32. …other initiatives with regard to the study and/or vocational choices of the students?
33. …the construction of class regulations?
34. …organizing tests?
35. …changes of the lesson schedule for the class?
36. …the approach towards problem students?
37. …the purchase of pedagogical material?
38. …working with and coordination of teaching methods and techniques?
39. …the integration of repeaters in the class group?
40. …agenda items proposed or imposed by the principal?

Decision making

Participatory decision making

Response categories: (1) never, (2) sometimes, (3) mostly, (4) always

Number of items: 6

Items

To what extent the principal involves an average teacher into decision making, concerning…

47. …the study programmes of the school?
53. …educational reform?
54. …the construction of the lesson package?
55. …the composition of the classes?
56. …the construction of the school report (the components)?
57. …extracurricular activities?

Autonomy in decision making

Response categories:

(1) The decision is taken entirely by the individual teacher himself.

(2) Idem as in (1), but after elaborate consultation with the (working) group of teachers instructing the same subject.

(3) The decision is taken primarily by the (working) group of teachers instructing the same subject.

(4) Idem as in (3), but after elaborate consultation with people outside this group of teachers (e.g., principal, cycle coordinators, other working groups).

Appendix 1. continued

(5) The decision is taken primarily by people outside the (working) group of teachers instructing the same subject.

Number of items: 8

Items

62. Concerning decisions about the study content.*
63. Concerning decisions about teaching and learning material.*
64. Concerning the teaching method.*
65. Concerning the way of differentiating (adapting the instruction to the differences between the students in the class).*
66. Concerning decisions about the frequency of homework.*
67. Concerning decisions about the content of a test.*
68. Concerning decisions about the frequency of tests.*
69. Concerning decisions about the criteria for tests (what is an unsatisfactory mark, etc.).*

Social system

Response categories: (1) totally agree, (2) agree, (3) neutral, (4) disagree, (5) totally disagree

Number of items: 17

Items

70. The principal exerts a lot of pressure on me as a teacher.
71. A good team spirit is present among the colleagues.*
72. Teachers are able to contact the principal easily.*
73. In general, good relationships are present among the colleagues at our school.*
74. The principal often stimulates the teachers to reflect together on what we aim for in our school.*
75. The principal stimulates the teachers into educational reform and innovation.*
76. I have the impression that the principal is rather unfriendly towards me.
77. Among the colleagues, there is a considerable willingness to help each other concerning our work.*
78. Meetings among the teachers where decision making is necessary, in general work out rather inefficiently.
79. The principal treats me in a distant and authoritarian way.
80. The atmosphere among the colleagues strongly fosters the quality of my work.*
81. I am stimulated by the principal to perform well as a teacher.*
82. Among the colleagues at our school, there is a sense of joint commitment to a common duty.*
83. The principal is always open to the personal situation of the teachers.*
84. During meetings among teachers, one does not listen well to one another.
85. I have difficulties to express my opinion towards the principal.
86. The principal stimulates the teachers to take their own initiatives concerning instruction at our school.*

Rules

Response categories:

(1) Rules are put in writing. These rules are always applied, without any distinction.

(2) Rules are put in writing. Deviation from this rule is possible after consultation with the principal (or another authority).

(3) Rules are <u>not</u> put in writing. The teacher acts after consultation with the principal (or another authority).

(4) Rules are <u>not</u> put in writing. The teacher acts based on his own evaluation.

Number of items: 11

Appendix 1. continued

Items
109. Concerning students' late arrival.*
110. Concerning students' disturbing behaviour during class.*
111. Concerning cheating.*
112. Concerning students' school absenteeism.*
113. Concerning the announcement of tests by teachers.*
114. Concerning giving homework or not.*
115. Concerning the way of evaluating tests.*
116. Concerning the requirements students need to meet before choosing a subject or a discipline.*
117. Concerning holding a class discussion of tests.*
118. Concerning students' behaviour outside the class.*
119. Concerning students' behaviour outside the school.*

Note: * Refers to a reversal of the item scores prior to the analyses

Appendix 2. Descriptive statistics and intraclass correlation coefficient (ICC) of the total set of 66 items

Item no.	Concerning	*M*	SD	ICC
9	Meetings subject teachers	2.46	0.95	0.24
10	Meetings subject teachers	2.91	1.11	0.18
11	Meetings subject teachers	2.34	0.99	0.08
12	Meetings subject teachers	2.01	0.78	0.08
13	Meetings subject teachers	2.23	1.10	0.33
14	Meetings subject teachers	2.15	0.73	0.05
15	Meetings subject teachers	2.18	0.74	0.03
16	Meetings subject teachers	2.01	0.96	0.05
17	Meetings subject teachers	2.21	0.87	0.06
18	Meetings subject teachers	2.51	0.97	0.09
27	Meetings class teachers	3.44	0.71	0.17
28	Meetings class teachers	3.03	0.71	0.08
29	Meetings class teachers	2.51	0.69	0.04
30	Meetings class teachers	3.13	0.76	0.07
31	Meetings class teachers	2.99	0.86	0.06
32	Meetings class teachers	2.54	0.79	0.03
33	Meetings class teachers	1.72	0.85	0.05
34	Meetings class teachers	1.84	0.92	0.10
35	Meetings class teachers	1.52	0.83	0.09
36	Meetings class teachers	2.92	0.84	0.06
37	Meetings class teachers	1.80	0.85	0.06
38	Meetings class teachers	1.74	0.76	0.04
39	Meetings class teachers	2.30	0.80	0.04
40	Meetings class teachers	2.65	0.97	0.06
47	Participatory decision making	2.13	0.96	0.18
53	Participatory decision making	2.24	0.95	0.10

Appendix 2. continued

Item no.	Concerning	*M*	SD	ICC
54	Participatory decision making	1.95	0.99	0.14
55	Participatory decision making	1.77	0.88	0.19
56	Participatory decision making	2.10	1.00	0.22
57	Participatory decision making	2.75	0.89	0.09
62	Autonomy in decision making	3.83	1.28	0.06
63	Autonomy in decision making	4.07	1.07	0.04
64	Autonomy in decision making	4.51	0.86	0.04
65	Autonomy in decision making	4.55	0.89	0.03
66	Autonomy in decision making	3.81	1.59	0.22
67	Autonomy in decision making	4.41	0.84	0.20
68	Autonomy in decision making	3.17	1.70	0.10
69	Autonomy in decision making	4.12	1.26	0.10
70	Social system	3.69	0.94	0.09
71	Social system	3.52	0.94	0.18
72	Social system	3.81	0.96	0.18
73	Social system	3.80	0.82	0.17
74	Social system	3.49	0.97	0.18
75	Social system	3.43	0.93	0.17
76	Social system	4.09	0.87	0.06
77	Social system	3.43	0.90	0.12
78	Social system	3.12	0.94	0.04
79	Social system	3.90	0.97	0.10
80	Social system	3.42	0.98	0.09
81	Social system	3.44	0.92	0.11
82	Social system	3.30	0.89	0.16
83	Social system	3.62	0.95	0.12
84	Social system	3.36	0.88	0.07
85	Social system	3.65	0.96	0.06
86	Social system	3.49	0.90	0.14
109	Rules	3.32	0.77	0.15
110	Rules	1.86	1.04	0.04
111	Rules	2.68	1.21	0.29
112	Rules	3.32	0.83	0.13
113	Rules	1.38	0.85	0.09
114	Rules	1.88	1.20	0.22
115	Rules	1.56	0.99	0.10
116	Rules	2.40	0.91	0.08
117	Rules	1.49	0.92	0.06
118	Rules	2.44	1.04	0.09
119	Rules	1.98	0.97	0.12

References

Anderson, C.S.: The search for school climate: a review of the research. Rev. Educ. Res. **52**, 368–420 (1982)

Barnette, J.J.: ScoreRel CI: an Excel program for computing confidence intervals for commonly used score reliability coefficients. Educ. Psychol. Meas. **65**, 980–983 (2005)

Collins, L.M., Schafer, J.L., Kam, C.-M.: A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol. Methods **6**, 330–351 (2001)

den Brok, P., Bergen, T., Stahl, R.J., Brekelmans, M.: Students' perceptions of teacher control behaviours. Learn. Instr. **14**, 425–443 (2004)

D'Haenens, E., Van Damme, J., Onghena, P.: Multilevel exploratory factor analysis: evaluating its surplus value in educational effectiveness research. In: Paper Presented at the Annual Meeting of the American Educational Research Association (AERA), New York, 24–28 March 2008

Dyer, N.G., Hanges, P.J., Hall, R.J.: Applying multilevel confirmatory factor analysis techniques to the study of leadership. Leadersh. Q. **16**, 149–167 (2005)

Enders, C.K.: Analyzing structural equation models with missing data. In: Hancock, G.R., Mueller, R.O. (eds.) Structural Equation Modeling: A Second Course, pp. 313–342. Information Age Publishing, Greenwich (2006)

Fraser, B.J., Walberg, H.J. (eds.): Educational Environments: Evaluation, Antecedents and Consequences. Pergamon, Oxford (1991)

Freiberg, H.J. (ed.): School Climate: Measuring, Improving, and Sustaining Healthy Learning Environments. Falmer, London (1999)

Grice, J.W.: Computing and evaluating factor scores. Psychol. Methods **6**, 430–450 (2001)

Gustafsson, J.-E., Stahl, P.A.: STREAMS 3.0 User's Guide. MultivariateWare, Mölndal (2005)

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C.: Multivariate Data Analysis. Prentice Hall, Upper Saddle River (1998)

Halpin, A.W., Croft, D.B.: The Organizational Climate of Schools. Department of Health, Education & Welfare, National Institute of Education, Washington DC (1962)

Heck, R.H., Thomas, S.L.: An Introduction to Multilevel Modeling Techniques. Lawrence Erlbaum Associates, Publishers, Mahwah (2000)

Holfve-Sabel, M.-A., Gustafsson, J.-E.: Attitudes towards school, teacher, and classmates at classroom and individual levels: an application of two-level confirmatory factor analysis. Scand. J. Educ. Res. **49**, 187–202 (2005)

Hoy, W.K., Tarter, C.J., Kottkamp, R.B.: Open Schools/Healthy Schools: Measuring Organizational Climate. Sage, Newbury Park (1991)

Hu, L., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Modeling **6**, 1–55 (1999)

Jöreskog, K.G., Sörbom, D.: LISREL 8: User's Reference Guide. Scientific Software International, Chicago (1996)

Julian, M.W.: The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. Struct. Equ. Modeling **8**, 325–352 (2001)

Mortimore, P.: The positive effects of schooling. In: Rutter, M. (ed.) Psychosocial Disturbances in Young People. Challenges for Prevention, pp. 333–363. Cambridge University Press, Cambridge (1997)

Muthén, B.O.: Latent variable modeling in heterogeneous populations. Psychometrika **54**, 557–585 (1989)

Muthén, B.O.: Mean and Covariance Structure Analysis of Hierarchical Data. University of California, Los Angeles (1990)

Muthén, B.O.: Multilevel covariance structure analysis. Sociol. Methods Res. **22**, 376–398 (1994)

Muthén, L.K., Muthén, B.O.: Mplus User's Guide, 4th edn. Muthén & Muthén, Los Angeles (1998–2006)

Muthén, B.O., Satorra, A.: Complex sample data in structural equation modeling. Sociol. Methodol. **25**, 267–316 (1995)

Opdenakker, M.-C.: Leerling in wonderland? Een onderzoek naar het effect van leerling-, lesgroep-, leerkracht- en schoolkenmerken op prestaties voor wiskunde in het secundair onderwijs (Student in wonderland? A study into the effect of student, class group, teacher and school characteristics on math achievement in secondary education). Dissertation, K.U.Leuven, Belgium (2003)

Rowan, B., Raudenbush, S.W., Kang, S.J.: School climate in secondary schools. In: Raudenbush, S.W., Willms, J.D. (eds.) Schools, Classrooms and Pupils: International Studies of Schooling from a Multilevel Perspective, pp. 203–223. Academic Press, NY (1991)

SAS Institute Inc.: SAS User's Guide: Statistics, 9th edn. SAS Institute Inc., Cary (2003)

Scheerens, J.: School effectiveness research and the development of process indicators of school functioning. Sch. Eff. Sch. Improv. **1**, 61–80 (1990)

Scheerens, J., Bosker, R.J.: The Foundations of Educational Effectiveness. Elsevier Science, Oxford (1997)

Schermelleh-Engel, K., Moosbrugger, H., Müller, H.: Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods. Psychol. Res. **8**, 23–74 (2003). Online http://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf. Accessed 15 Sept 2008

Schumacker, R.E., Lomax, R.G.: A Beginner's Guide to Structural Equation Modeling. Lawrence Erlbaum Associates, Publishers, Mahwah (1996)

Shieh, Y.-Y., Fouladi, R.T.: The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. Educ. Psychol. Meas. **63**, 951–985 (2003)

Sivo, S.A., Fan, X., Witta, E.L., Willse, J.T.: The search for "optimal" cutoff properties: fit index criteria in structural equation modeling. J. Exp. Educ. **74**, 267–288 (2006)

Sörbom, D.: Model modification. Psychometrika **54**, 371–384 (1989)

SPSS: SPSS Base 15.0 User's Guide. SPSS Inc., Chicago (2006)

Stapleton, L.M.: Using multilevel structural equation modeling techniques with complex sample data. In: Hancock, G.R., Mueller, R.O. (eds.) Structural Equation Modeling: A Second Course, pp. 345–383. Information Age Publishing, Greenwich (2006)

Stockard, J., Mayberry, M.: Effective Educational Environments. Corwin Press, Newbury Park (1992)

Sweetland, S.R., Hoy, W.K.: School characteristics and educational outcomes: toward an organizational model of student achievement in middle schools. Educ. Adm. Q. **36**, 703–729 (2000)

Van Damme, J., De Fraine, B., Van Landeghem, G., Opdenakker, M.-C., Onghena, P.: A new study on educational effectiveness in secondary education in Flanders: an introduction. Sch. Eff. Sch. Improv. **13**, 383–397 (2002)

Van Damme, J., Van Landeghem, G., De Fraine, B., Opdenakker, M.-C., Onghena, P.: Maakt de school het verschil? Effectiviteit van scholen, leraren en klassen in de eerste graad van het middelbaar onderwijs (Does the school make the difference? Effectiveness of schools, teachers and classes in the first cycle of secondary education). Acco, Leuven (2004)

Vandenberghe, R., Bohets, A., Claus, G., Vernelen, A., Viaene, M.: De determinanten van het professionele handelen van leerkrachten secundair onderwijs en de invloed op de onderwijskwaliteit (The determinants of secondary teachers' professional capacities and the influence on the quality of education). K.U.Leuven, Departement Pedagogische Wetenschappen, Centrum voor Onderwijsbeleid en –vernieuwing, Belgium (1994)

Verhoeven, J.C., Vandenberghe, R., Van Damme, J., Clement, M., Maetens, D., Vergauwen, G.: Schoolmanagement en kwaliteitsverbetering van het onderwijs. Een empirisch onderzoek in secundaire scholen (School management and quality improvement in education. An empirical research in secondary schools). K.U.Leuven, Departement Pedagogische Wetenschappen, Studiegroep voor Onderwijsbeleid en –begeleiding, Belgium (1992)

Westling Allodi, M.: A two-level analysis of classroom climate in relation to social context, group composition, and organization of special support. Learn. Environ. Res. **5**, 253–274 (2002)

Yang, Y.: Measuring socioeconomic status and its effects at individual and collective levels: a cross-country comparison. Dissertation, University Göteborg, Sweden (2003)