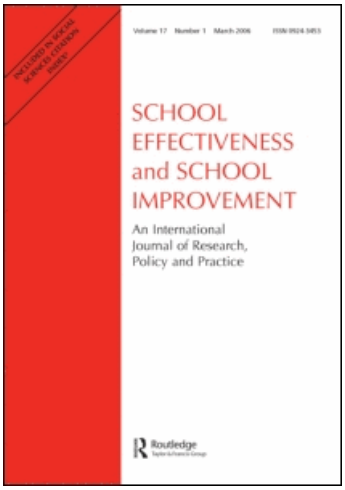

School Effectiveness and School Improvement
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t714592801

Explaining stability and changes in school effectiveness by looking at changes in the functioning of school factors

Bert P. M. Creemers[a]; Leonidas Kyriakides[b]
[a] Faculty of Psychology, Education and Sociology, University of Groningen, Groningen, The Netherlands [b] Department of Education, University of Cyprus, Nicosia, Cyprus

First published on: 29 September 2010

To cite this Article Creemers, Bert P. M. and Kyriakides, Leonidas(2010) 'Explaining stability and changes in school effectiveness by looking at changes in the functioning of school factors', School Effectiveness and School Improvement,, First published on: 29 September 2010 (iFirst)
To link to this Article: DOI: 10.1080/09243453.2010.512795
URL: http://dx.doi.org/10.1080/09243453.2010.512795

School Effectiveness and School Improvement
2010, 1–19, i*First* article

Explaining stability and changes in school effectiveness by looking at changes in the functioning of school factors

Bert P.M. Creemers[a] and Leonidas Kyriakides[b]*

[a]*Faculty of Psychology, Education and Sociology, University of Groningen, Groningen, The Netherlands;* [b]*Department of Education, University of Cyprus, Nicosia, Cyprus*

(*Received 18 November 2009; final version received 29 May 2010*)

This paper investigates the extent to which changes in the effectiveness status of schools can be related to changes in the functioning of school factors included in the dynamic model of educational effectiveness. The methods of a follow-up study were identical to those of a study conducted 4 years ago in order to test the validity of the dynamic model. Since the follow-up study took place in the same schools where the original study took place, changes in the effectiveness status of schools and in the functioning of effectiveness factors were identified. Discriminant function analysis reveals that changes not only in the functioning of some school factors but also in the quality of teaching practice can help us classify the schools into those which improved their effectiveness status, those which remained equally effective, and those which even reduced their effectiveness status. Implications of findings for the development of educational effectiveness research are drawn.

Keywords: dynamic perspectives of educational effectiveness; changes in school effectiveness over time; school policy; school learning environment; testing educational theories

Introduction

Teaching and learning are dynamic processes that are constantly adapting to changing needs and opportunities. Effective schooling should, therefore, be treated as a dynamic, ongoing process. This idea is consistent with the contingency theory (Donaldson, 2001; Mintzberg, 1979) and with the main assumptions upon which the dynamic model of educational effectiveness is based (Creemers & Kyriakides, 2008). Specifically, the dynamic model assumes that schools which are able to identify their weaknesses and take actions to improve their policy on aspects associated with teaching and their school learning environment (SLE) are able to improve their effectiveness status. This element of the dynamic model reveals an essential difference of this model from all the theoretical models of educational effectiveness developed during the last decade (e.g., Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992). Some supportive material for the validity of the dynamic model has been provided (Kyriakides, 2008). A longitudinal study was designed to test the main

*Corresponding author. Email: kyriakid@ucy.ac.cy

ISSN 0924-3453 print/ISSN 1744-5124 online

DOI: 10.1080/09243453.2010.512795
http://www.informaworld.com

assumptions of the model, and empirical support to the importance of the main teacher and school factors of the model has been provided (Creemers & Kyriakides, 2010; Kyriakides & Creemers, 2008). Moreover, a quantitative synthesis of the results of studies exploring the impact of school factors on student achievement has provided some further support to the validity of the model at the school level (see Kyriakides, Creemers, Antoniou, & Demetriou, 2010). However, this meta-analysis revealed that there is no study searching for factors that can explain changes in the effectiveness status of schools over time. Thus, in this paper we look at two important issues that may help us understand better the dynamic nature of educational effectiveness. First, we address the issue of time stability of school effectiveness by investigating the extent to which the effectiveness status of schools can be seen as something more or less stable over time or whether changes in the effectiveness status of schools can be observed. Second, we search for factors that may explain the type of changes in the effectiveness status of schools that may be observed.

In this context, we conducted a follow-up study in the same schools where the original study testing the validity of the dynamic model was conducted 4 years ago. Since the design of the follow-up study is identical to several aspects of the original study, the follow-up study can also be seen as a replication study that aims to test the generalizability of the findings of the original study. Although the follow-up study provided support to the generalizability of the original study and generated evidence supporting the validity of some aspects of the dynamic model (see Kyriakides & Creemers, 2009), this paper investigates the extent to which school-level factors included in the dynamic model can help us explain changes in the effectiveness status of schools. By comparing the data emerged from the two studies, it was possible to identify whether changes/stability in the effectiveness status of schools participating in both studies can be explained by changes/stability in the functioning of school factors. Thus, this paper provides an overview of the dynamic model and presents the main results emerging from this comparison, which helps us to test one of the major assumptions of this model. This comparison will also help us identify factors that explain changes in the effectiveness status of schools and understand better the dynamic nature of educational effectiveness.

The dynamic model of educational effectiveness: an overview

The essential characteristics of the dynamic model

The dynamic model is multilevel in nature and refers to factors associated with learning outcomes that are situated at four different levels: student, classroom, school, and system. The teaching and learning situation is emphasized and the roles of the two main actors (i.e., teacher and student) are analysed. Above these two levels, the dynamic model also refers to school-level factors. It is expected that school-level factors influence the teaching-learning situation by developing and evaluating the school policy on teaching and the policy on creating a learning environment at the school. The final level refers to the influence of the educational system through a more formal way, especially through developing and evaluating the educational policy at the national/regional level. The model also assumes that factors at the school and system level have both direct and indirect effects on student achievement. Finally, the dynamic model is based on the assumption that, although there are different effectiveness factors, each factor can be defined and measured using five dimensions: *frequency, focus, stage, quality, and differentiation*. Frequency is a quantitative way to

measure the functioning of each effectiveness factor. The other four dimensions examine qualitative characteristics of the functioning of the factors and help us describe the complex nature of effective teaching. Creemers and Kyriakides (2006) provide a description of each dimension and draw implications of using these dimensions for establishing comprehensive strategies to improve the functioning of school factors by taking into account both their quantitative and qualitative characteristics.

Classroom factors of the dynamic model

Based on the main findings of teacher effectiveness research (e.g., Brophy & Good, 1986; Muijs & Reynolds, 2001; Rosenshine & Stevens, 1986), the dynamic model refers to factors which describe teachers' instructional role and are associated with student outcomes. These factors refer to observable instructional behaviour of teachers in the classroom rather than to factors that may explain such behaviour (e.g., teacher beliefs and knowledge and interpersonal competences). The eight factors included in the model are as follows: orientation, structuring, questioning, teaching-modelling, applications, management of time, teacher role in making classroom a learning environment, and classroom assessment. These eight factors do not refer only to one approach of teaching such as structured or direct teaching (Joyce, Weil, & Calhoun, 2000) or to approaches associated with constructivism (Schoenfeld, 1998). An integrated approach in defining quality of teaching is adopted.

School factors of the dynamic model

School factors are expected to influence classroom-level factors, especially the teaching practice. Therefore, the dynamic model gives emphasis to the following two main aspects of the school policy which affect learning at both the level of teachers and students: (a) school policy for teaching and (b) school policy for creating a learning environment at school. Guidelines are seen as one of the main indications of school policy, and this is reflected in the way each school-level factor is defined. However, in using the term guidelines we refer to a range of documents, such as staff meeting minutes, announcements, and action plans, which make the policy of the school more concrete to the teachers and other stakeholders. These two factors do not imply that each school should simply develop formal documents to install its policy. The factors concerned with the school policy mainly refer to the actions taken by the school to help teachers and other stakeholders have a clear understanding of what is expected of them. Support offered to teachers and other stakeholders to implement the school policy is also an aspect of these two school factors (Creemers & Kyriakides, 2008).

Based on the assumption that the essence of a successful organization in the modern world is the search for improvement, the dynamic model is also concerned with the processes and the activities which take place in the school in order to improve the teaching practice and its learning environment. For this reason, the processes which are used to evaluate the school policy for teaching and the SLE are investigated. It is expected that evaluation mechanisms will generate data that will help schools to take decisions on how to improve the functioning of school factors. Thus, the following four overarching factors at the school level are included in the model:

(1) school policy for teaching and actions taken for improving teaching practice;
(2) evaluation of school policy for teaching and of actions taken to improve teaching;
(3) policy for creating a SLE and actions taken for improving the SLE; and
(4) evaluation of the SLE.

Methods

Rationale of the study

Since the major aim of conducting the follow-up study was to search for changes in the functioning of school factors and relate them to changes in the effectiveness status of schools, the design of this study is identical to several aspects of the original study, which was conducted in order to test the validity of the dynamic model. First, this study took place in the same schools where the original study was conducted. Since the student body and most teaching staff of these schools have changed, by collecting data from the same 50 schools where the original study was conducted it was possible to measure the changes/stability in the effectiveness status of these schools and try to relate them with changes/stability in the functioning of the school factors irrespective of the fact that the teaching staff has changed or remained the same. This implies that the study is measuring changes or stability in the functioning of teacher and school factors rather than changes in the composition of the staff population. Second, we use the same instruments to measure teacher and school factors. Our decision to use the same instruments was not only based on the fact that the construct validity of each instrument was demonstrated (Kyriakides & Creemers, 2008) but was also taken in order to ensure that differences in results emerged from the two studies are not due to using different instruments to collect data. Finally, the same age group of students is used in order to avoid problems that may arise due to the fact that some factors may have differential effects on achievement of different age groups of students. The next part of the methods section refers to the participants and the methods used to conduct the original and the follow-up study.

Participants

The school sample of the follow-up study was identical with the sample of the original study and consisted of 50 primary schools. In both studies, all the Grade 6 students from each class of the school sample were chosen. The chi-square test did not reveal any statistically significant difference between the sample of each study and the population in terms of students' sex (original study: $X^2 = 0.84$, $df = 1$, $p = 0.36$; follow-up study: $X^2 = 0.69$, $df = 1$, $p = 0.41$). Moreover, the t test did not reveal any statistically significant difference between the research sample of each study and the population in terms of the size of class (original: $t = 1.21$, $df = 107$, $p = 0.22$; follow-up study: $t = 1.62$, $df = 111$, $p = 0.11$). Although these two studies refer to other variables such as the socioeconomic status of students and their achievement levels in different outcomes of schooling, there are no data about these characteristics of the Cypriot students of Grade 6. Therefore, it was not possible to examine whether the sample of each study was nationally representative in terms of any other characteristic than students' sex and size of class. However, it can be

claimed that in each study a nationally representative sample of Cypriot Grade 6 students in terms of these two characteristics was drawn.

The overall teacher and student sample size for each study is shown in Appendix 1, along with a variety of descriptive statistics on the student background variables. For each study, students with missing prior attainment or background data (less than 7% of the original sample of each study) were excluded from the analyses. Appendix 1 also shows that there is no statistically significant difference in any of the background characteristics between the samples of these two studies. It is, however, important to note that in Cyprus, teacher appointments in all public primary schools are the responsibility of the educational service committee and each teacher is appointed at a school for a maximum period of 5 years (Kyriakides, 1999). Therefore, the great majority of teachers and headteachers who were teaching at our school sample during the school year 2004–2005 were not serving at the same school when the follow-up study took place. Although it is acknowledged that this policy creates an unstable context of schooling, the trade union of Cypriot teachers, which has a very strong political power, is in favour of this policy (Kyriakides, Demetriou, & Charalambous, 2006). As a consequence, during the last 20 years, none of the Ministers of Education attempted to change the teacher appointment and transfer system.

Variables

Output measures

Data on achievement in Mathematics and Greek Language were collected by using external forms of assessment. Written tests were administered to our student sample when they were at the beginning of Grade 6 and when they were at the end of Grade 6. The construction of the tests was subject to controls for reliability and validity. For each subject, the Extended Logistic Model of Rasch (Andrich, 1988) was used to analyse the emerging data at the beginning and at the end of the school year separately, and two scales per subject were created. Analysis of data on student achievement revealed that each scale had relatively satisfactory psychometric properties. Specifically, for each scale the indices of cases (i.e., students) and item separation were higher than 0.80, indicating that the separability of each scale was satisfactory (Wright, 1985). Moreover, the infit mean squares and the outfit mean squares of each scale were near one and the values of the infit t scores and the outfit t scores were approximately zero. Furthermore, each analysis revealed that all items had item infit with the range 0.84 to 1.19. Therefore, each analysis revealed that there was a good fit to the model (Keeves & Alagumalai, 1999). Thus, for each student, it was possible to generate two different scores for his/her achievement in each subject at the beginning of Grade 6 and at the end of Grade 6, by calculating the relevant Rasch person estimate. For each study, descriptive statistics for student achievement measures and teacher and school factors are presented in Appendix 2. No statistically significant difference between the samples of these two studies in any measure of student achievement and any measure of teacher and school factor has been identified.

Student background factors

Information was collected on two student background factors: sex (0 = boys, 1 = girls), and socioeconomic status (SES). Five SES variables were available:

father's education level, mother's education level (i.e., graduate of a primary school, graduate of a secondary school, or graduate of a college/university), the social status of father's job, the social status of mother's job, and the financial situation of the family. Following the classification of occupations used by the Ministry of Finance, it was possible to classify parents' occupation into three groups which have relatively similar sizes: occupations held by working class, occupations held by middle class, and occupations held by upper-middle class. Representative parental occupations for the working class are: farmer, truck driver, machine operator in a factory; for the middle class: police officer, teacher, bank officer; and for the upper-middle class: doctor, lawyer, business executive. Relevant information for each child was taken from the school records. Then standardized values of the above five variables were calculated, resulting in the SES indicator.

Quality of teaching

The explanatory variables which refer to the eight factors of the dynamic model dealing with teacher behaviour in the classroom were measured by both independent observers and students. Taking into account the way the five dimensions of each effectiveness factor are defined, one high-inference and two low-inference observation instruments were developed. These observation instruments generate data for all eight factors and their dimensions (Kyriakides & Creemers, 2008). In each study, observations were carried out by four members of the research team who attended a series of seminars on how to use the three observation instruments. The external observers visited each class 6 times and observed three lessons per subject. For each scale of the three observation instruments, the alpha reliability coefficient was higher than 0.83, and the inter-rater reliability coefficient ρ^2 was higher than 0.75.

The eight factors and their dimensions were also measured by administering a questionnaire to students. Specifically, students were asked to indicate the extent to which their teacher behaves in a certain way in their classroom, and a Likert scale was used to collect data. A generalizability study (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson, Webb, & Rowley, 1989) on the use of students' ratings was conducted. It was found that the data emerged from almost all the questionnaire items can be used for measuring the quality of teaching of each teacher in each subject separately (see Kyriakides & Creemers, 2009).

In order to test the construct validity of the instruments and establish measures of teacher factors, data emerged from each study were analysed separately as follows. At the first stage, for each subject, separate Confirmatory Factor Analyses (CFA) for each effectiveness factor were conducted in order to identify the extent to which data emerged from different methods can be used to measure each factor in relation to the five dimensions of the dynamic model. The main results which emerged from using CFA approaches to analyse the multitrait multimethod matrix (MTMM) concerned with each classroom-level factor of the dynamic model in relation to each subject are briefly presented below. First, support to the construct validity of the five measurement dimensions of most effectiveness factors was provided by the two separate analyses of the data emerged from each study. The few exceptions which were identified reveal the difficulty of defining the quality dimension of some factors (see Kyriakides & Creemers, 2009). Second, the comparison of CFA models used to test each factor confirmed convergent and discriminant validity for the five dimensions. Convergent validity for most measures

was demonstrated by the relatively high (i.e., higher than .60) standardized trait loadings, in comparison to the relatively lower (i.e., lower than .38) standardized method loadings. These findings support the use of multimethod techniques for increasing measurement validity, construct validity, and, thus, stronger support for the validity of subsequent results.

Having established the construct validity of the framework used to measure the functioning of the teacher-level factors of the dynamic model, it was then decided to analyse the data emerged from each study separately by using the Rasch model in order to identify the extent to which the five dimensions of these factors are reducible to a common unidimensional scale. The Rasch model does not only test the unidimensionality of the scale but is also able to find out whether the tasks can be ordered according to the degree of their difficulty, and at the same time the people who carry out these tasks can be ordered according to their performance in the construct under investigation. This procedure is justified theoretically and is used in studies on teacher evaluation (e.g., Burry & Shaw, 1988; Wang & Cheng, 2001).

Thus, for each study, the Rasch model was applied on the whole sample of teachers and all measures concerned with their teaching skills together using the computer program Quest (Adams & Khoo, 1996). By analysing the data of the original study, it was found that only two teaching skills (i.e., the focus dimension of the structuring factor and the quality dimension of time management) did not fit the model. The results of the various approaches used to test the fitting of the Rasch model to our data revealed that there was a good fit to the model when teachers' performance in the other teaching skills were taken into account (see Kyriakides, Creemers, & Antoniou, 2009). Similar results emerged by analysing the data from the follow-up study. It was found that there was a good fit to the model when teachers' performance in all teaching skills but the focus dimension of orientation were taken into account. Specifically, by using the Rasch model to analyse teacher performance, it was found that the teaching skills included in the dynamic model were well targeted against the teachers' measures since teachers' scores range from –3.06 to 3.01 logits and the difficulties of teaching skills range from –2.79 to 3.09 logits. Moreover, the indices of persons and of teaching skills separation were found to be higher than 0.92, indicating that the separability of the scale is satisfactory. Therefore, for each teacher participating in each study, it was possible to generate two different scores for his/her teaching abilities in each subject, by calculating the relevant Rasch person estimate.

School-level factors of the dynamic model

The explanatory variables which refer to the four school-level factors of the dynamic model were measured by asking all the teachers of the school sample to complete a questionnaire. The questionnaire was designed in such a way that information about the five dimensions of each school factor could be collected. In each study, a satisfactory response rate was obtained (i.e., original study: 86%; follow-up study: 84%). The chi-square test did not reveal any statistically significant difference between the distribution of the teacher sample which indicates at which school each teacher works and the relevant distribution of the whole population of the teachers of the 50 schools of our sample (original: $X^2 = 57.12$, $df = 49$, $p = .20$; follow-up: $X^2 = 61.21$, $df = 49$, $p = .11$). Therefore, the sample was representative of the whole population in terms of how the teachers are distributed in each of these 50 schools.

Results concerning the internal reliability and the discriminate and construct validity of the questionnaire used to measure teacher views of the school factors are presented below.

Reliability, consistency, and variance at school level. Since it is expected that teachers within a school view the policy of their school and the evaluation mechanisms of their school similarly, but differently from teachers in other schools, a generalizability study was initially conducted. For each study, it was found that, for almost all questionnaire items, the object of measurement was the school. (For more information on the results emerged from the statistical models used to conduct the generalizability study, see Kyriakides & Creemers, 2009.) Since the generalizability of the great majority of the items measuring school policy in relation to the development of positive values towards learning was questionable, it was decided to drop all the items which refer to this factor. Thus, reliability was computed for each of the dimensions of the school factors but the factor concerned with the values towards learning by calculating multilevel λ (Snijders & Bosker, 1999) and the Cronbach alpha for data aggregated at the school level. The value of Cronbach alpha represents consistency across items, whereas multilevel λ represents consistency across groups of teachers. For each scale, the reliability coefficients which emerged were found to be high (i.e., original study: between .82 and .88; follow-up study: between .83 and .90).

The intra-class correlations of the scales were also computed. The intra-class correlations, which indicate what amount of variance of the teacher questionnaire is located at the school level, were between 0.32 and 0.48. These percentages are rather high compared to other instruments that measure perceptions of people or objects in clustered or interdependent situations (Den Brok, Brekelmans, Levy, & Wubbels, 2002). This finding reveals that there is enough homogeneity in the views of teachers of the same school about each scale of the questionnaire concerned with the school factors, and thereby aggregated scores at the level of school were estimated.

Discriminate validity. The mean correlation of one scale with the other scales measuring a multidimensional construct indicates the degree of discriminate validity. The lower the scales correlate amongst each other, the less they measure the same dimension of the construct. Thus, the discriminate validity was calculated for the 45 scales of school factors. For each study, it was found that the scales correlated between 0.10 and 0.45. Moreover, in each study, less than 75 out of 1,035 correlations were statistically significant, and all of them refer to the relationships of indicators of different dimensions of the same school factor. Finally, for each study, the values of the mean correlation of a scale with the other scales were smaller than .30. This implies that the 45 scales of the questionnaire, which refer to indicators of the five dimensions of the school factors, differed sufficiently, although they partly measured the same general construct.

Construct validity. For the identification of the factor structure of the teacher questionnaire, structural equation modelling (SEM) analyses were conducted using EQS (Bentler, 1995). Each model was estimated by using normal theory maximum likelihood methods (ML). More than one fit index was used to evaluate the extent to which the data fit the models tested. More specifically, the scaled chi-square, Bentler's (1990) Comparative Fit Index (CFI), and the Root Mean Square Error of

Approximation (RMSEA) (Brown & Mels, 1990) were examined. Finally, the factor parameter estimates for the models with acceptable fit were examined to help interpret the models. Although for each study separate SEM analyses were conducted, the main results of SEM analyses emerged from the data of the original study (see Creemers & Kyriakides, 2010) were very similar to those emerged by conducting comparable SEM analyses of the data emerged from the follow-up study (see Kyriakides & Creemers, 2009). Both studies reveal that, for each factor, the model which fits the data better is the one that refers to the existence of the five dimensions of the dynamic model. As a consequence, for each study, factor scores for each dimension of each school factor were calculated (see Appendix 2).

Results

In this section, we present the results of our attempt to identify whether changes in the functioning of school factors predict changes in the effectiveness status of schools in each subject. In order to achieve this aim, the following procedure was undertaken. For each study, we conducted separate multilevel modelling analyses to identify the impact of school factors included in the dynamic model upon achievement in each subject (see Creemers & Kyriakides, 2010; Kyriakides & Creemers, 2009). For the purposes of this paper, prior achievement and background factors were only controlled in order to estimate the schools' ''value-added'' contributions. These are typically referred to as the effectiveness scores of schools, but they also reflect other unmeasured factors (outside the control of the school) which were not controlled for in the analysis (Thomas, 2001). Therefore, based on the results of Model 1, which emerged by adding student prior attainment and background factors into the empty model (see Appendix 3), the difference between the expected and the actual score for each school was plotted. The standard error of estimate for each school was also taken into account and is represented by the length of a vertical line. This line can be conceptualised as the range within which we are 95% confident that the ''true'' estimate of the school's residual lies (Goldstein, 2003). Thus, where this vertical line does not cross the horizontal zero line and is also situated below the zero line, the school it represents is considered as one of the least effective schools of our sample. On the other hand, where this line does not cross the horizontal zero line and is situated above the zero line, the school it represents is characterised as one of the most effective schools. All the other schools are characterised as typical.

At the next step, for each subject, it was possible to compare the effectiveness status of each school during the school year 2004–2005 with its effectiveness status during the school year 2008–2009. Table 1 illustrates the distribution of changes in the effectiveness status of our school sample in each subject separately. The following observations arise from this table. First, for each subject, no change in the effectiveness status of more than 50% of our school sample can be observed. It is also important to note that three schools were among the most effective schools in both time periods and managed to achieve this aim for both subjects. Second, in each subject, approximately 12 schools managed to improve their effectiveness status, whereas the effectiveness status of an almost equal number of schools declined. Third, extreme changes in the effectiveness status of the schools are observed in only one school, which dropped down from the most to least effective in language.

Since the figures of Table 1 reveal that changes in the effectiveness status of a relatively large number of schools took place, we conducted a Discriminant Function

Table 1. The distribution of the school sample according to their effectiveness status in mathematics and Greek language during the school year 2004–2005 and during the school year 2008–2009.

Groups of schools	Mathematics	Language
A) Stability		
Remain Typical	14	15
Remain Least Effective	6	6
Remain Most Effective	7	6
B) Improvement		
From Least Effective to Typical	5	5
From Least Effective to Most Effective	0	0
From Typical to Most Effective	6	7
C) Declining		
From Most Effective to Typical	6	6
From Typical to Least Effective	6	4
From Most Effective to Least Effective	0	1

Analysis (DFA) to find out whether changes in the effectiveness status of schools can be explained by taking into account the observed changes in the functioning of the school factors of the dynamic model. DFA is a statistical technique used for classifying observations (Klecka, 1980) and involves the predicting of a categorical dependent variable by one or more continuous or binary independent variables. It is statistically the opposite of Multiple Analysis of Variance (MANOVA), and it is very useful in determining whether a set of variables is effective in predicting category membership. Thus, the main purpose of the DFA employed for the purposes of this study was to predict to which of the following three groups each school of our sample belongs: (a) schools which managed to improve their effectiveness status, (b) schools which managed to keep their status to the same level, or (c) schools which reduced their effectiveness status. In the first part of this section, a classification of the observed changes in the effectiveness status of our schools in each subject has been presented. The next step in this analysis was to create a set of observations where both group membership and the values of the interval variables will be known. For the purposes of this study, changes in the functioning of each school factor included in the dynamic model were assumed to be the interval variables (i.e., the predictors).

For each subject, DFA was applied in order to distinguish our school samples into those which: (a) improved their effectiveness status, (b) did not change their effectiveness status, and (c) reduced their effectiveness status. At the first stage, DFA is used to reveal a function that is able to distinguish among those schools which managed to improve their status from the other two groups of schools (i.e., those which did not improve their status). Then, we identified a function which helps us to distinguish between those schools which did not change their status and those schools where a decline in their status was observed. The eigenvalues which emerged reveal that, in the case of mathematics, the first function accounts for 59% of the variance, whereas the second function accounts for 41%. Similarly, in the case of language, the first function accounts for 57% of the variance, whereas the second function accounts for 43%. The significance of Wilks lambda reveals that for each subject both functions were found to be statistically significant, so both of them can

help us distinguish between the three groups of our schools. These figures also reveal that for both subjects it was relatively easier to distinguish between the schools which managed to improve their effectiveness status and those which did not improve their status rather than to distinguish between those which did not change their status and those which reduced their status.

One of the benefits of DFA is that it produces a classification table showing where the data were categorized and in which groups they were predicted to be (see Table 2). The table also shows the percentage of cases which were correctly classified through the prediction of group membership. Since DFA will classify cases into the largest group, a statistic, tau, can be computed showing the proportional reduction of error (PRE) when using the predicted model. In the analysis of mathematics achievement, Table 2 shows that the percentage of schools which were correctly classified was 74%, whereas the percentage of the largest group was 54%. The value of PRE for the analysis in mathematics shows that placements based on this model increase by 43.5%, which translates into about 22 schools placed more correctly using this model. Similar results emerged from the analysis of achievement in language. Table 2 shows that 76% of schools were correctly classified, and the PRE reveals that placements based on this model increase by 47.8%, which translates into about 24 schools placed more correctly using this model. It is finally important to note that the main weakness of the classification emerged by DFA was concerned with our difficulty to identify more than 35% of the declining schools. More specifically, in the case of mathematics, 5 out of 12 declining schools were expected not to change their effectiveness status. Similarly, using DFA to analyse achievement in language, it was found that 4 out of 11 declining schools were expected to remain equally effective. For each subject, our difficulty was to identify schools which were among the most effective and dropped to typical. In the case of mathematics, 6 schools were among the most effective and dropped to typical, and 4 of them were misclassified as not changing their effectiveness status. In the case of language, all the declining schools which were misclassified as not changing their status ($n = 4$) were among the most effective schools and dropped to typical.

Table 3 shows the standardized weights for the model. All the variables used for this analysis refer to changes in the functioning of school factors and emerged by comparing for each school its scores associated with each school factor emerged from the original study with the one that emerged from the follow-up study. In the case of quality of teaching, an average of the skills of teachers of each school

Table 2. Classification results of changes in the school effectiveness status in each subject.

Groups of schools	Predicted group membership			Total
	Improvement	Stability	Declining	
Changes in the effectiveness status of schools in mathematics				
Improvement	8 (72.7%)	2 (18.2%)	1 (9.1%)	11
Stability	4 (14.8%)	22 (81.5%)	1 (3.7%)	27
Declining	0 (0%)	5 (41.7%)	7 (58.3%)	12
Changes in the effectiveness status of schools in Greek language				
Improvement	10 (83.3%)	2 (16.7%)	0 (0.0%)	12
Stability	4 (14.8%)	21 (77.8%)	2 (7.4%)	27
Declining	0 (0%)	4 (36.4%)	7 (63.6%)	11

Table 3. Standardized canonical discriminant function coefficients based on analysis of school effectiveness in each subject separately.

Variables concerned with changes in the functioning of school factors	Mathematics		Greek Language	
	Function 1	Function 2	Function 1	Function 2
Quality of teaching practice (aggregated at the school level)	0.532	0.416	0.509	0.382
School policy for teaching				
Frequency	0.201	0.191	0.211	0.227
Stage	0.108	0.115	0.128	0.107
Quality	0.136	0.146	0.206	0.171
Evaluation of school policy for teaching				
Frequency	0.218	0.190	0.231	0.214
Quality	0.180	0.109	0.147	0.139
Teacher Collaboration				
Stage	0.125	0.109	0.104	0.134
Quality	0.174	0.169	0.184	0.164
Partnership policy				
Frequency	0.209	0.120	0.269	0.200
Quality	0.239	0.175	0.225	0.207
Differentiation	Not used	Not used	0.125	0.169
Provision of resources (frequency)	0.167	0.170	Not used	Not used
Evaluation of SLE (quality)	0.291	0.251	0.321	0.287

(i.e., person estimate) in each subject was estimated in order to have an estimation of the quality of teaching practice at the school level. Then, for each school, we compared the two estimates of quality of teaching practice and found out whether their teaching practice was improved, remained the same, or even declined. The following observations arise from Table 3.

First, for each subject, it was found that changes that were observed in the actual teaching practice of our school sample was the most relevant variable that helps us distinguish between schools which managed to improve their effectiveness status and those which did not improve their effectiveness. Change in the teaching practice was also the most powerful variable helping us to distinguish between schools which remained equally effective and those which declined. Second, changes of some dimensions of each overarching school factor were found to contribute in helping us predicting changes in the effectiveness status of our school sample. This provides some support to the assumption of the dynamic model that the impact of school factors upon the effectiveness of schools should be measured by investigating the impact of changes in the functioning of these factors upon the improvement of school effectiveness. Third, the importance of using different dimensions to measure the school factors is also confirmed, especially since more than half of the variables included in each analysis were not concerned with the frequency dimension of school factors. Moreover, the quality rather than the frequency dimension of two school factors (i.e., teacher collaboration and evaluation of SLE) was included in each analysis. Fourth, the quality dimension of the evaluation of the SLE had stronger effect than any other school factors. On the other hand, the provision of resources was found to have the weakest contribution, and in the case of language its

contribution was not even statistically significant at the .05 level. Finally, only one aspect of the SLE (i.e., student behaviour outside the classroom) was not found to contribute to each analysis.

Discussion and conclusion

Implications of findings for the development of Educational Effectiveness Research (EER) are discussed, and suggestions for further research are provided. First, by conducting a follow-up study to the same schools where the original study testing the validity of the dynamic model was conducted, one of the most important assumptions of the dynamic model was tested. By comparing the effectiveness status of our school sample during 2 different school years, it was found out that changes in the effectiveness status of a significant number of schools can be observed. Although almost 55% of schools remained equally effective and dramatic changes (i.e., from most effective suddenly to become least effective or vice versa) were not observed, significant improvement in the effectiveness status of almost 25% of our school sample was observed. This finding provides some support to studies investigating school effectiveness over a long period of time which reveal that there are limits to improvement over a long period of time and show that a relatively small proportion of schools have significant improvement patterns (Thomas, Peng, & Gray, 2007). But, although improvement of school effectiveness status appears to be difficult, the limited number of studies focusing on longer term changes in schools seems also to reveal that improvement is not an impossible task (e.g., Gray, Goldstein, & Jesson, 1996; Gray, Goldstein, & Thomas, 2001; Thomas, 2001). Moreover, by taking into account that in all 50 schools the headteacher and most teachers were moved from these schools to some other schools, one could claim that changing the personnel of the school does not explain changes in the effectiveness status of the schools. On the contrary, this paper seems to reveal that changes in the functioning of school factors included in the dynamic model help us predict the type of change that took place in our school sample. The results of DFA revealed that, in each subject, we can predict changes in the effectiveness status of schools by looking first of all at changes in the quality of teaching practice and then at changes in the functioning of most school factors of the dynamic model. By relating changes in the functioning of school factors with changes in their effectiveness status, this paper provides more robust evidence about the validity of the dynamic model than the previous studies which managed to show that there are associations between school factors and student achievement gains during a school year. Based on the findings reported in this paper, we can also provide suggestions to schools on how to improve their effectiveness status and undertake studies investigating the extent to which the dynamic model can be used for improvement purposes.

Second, implications for the development of the methodology of EER can be drawn. This paper shows that by conducting two parallel studies in the same school sample, researchers can consider the possibility to raise causality issues and attempt to demonstrate cause and effect relations among changes in the school factors and changes in school effectiveness. Causality is an important methodological issue, especially since EER is searching for factors explaining student achievement. However, most of the effectiveness studies conducted during the last 3 decades were designed in such a way that only associations between factors and student achievement were demonstrated. Therefore, this paper draws attention to the

importance of conducting follow-up/replication studies which will not only test the generalizability of the findings of the original studies but may also help us study changes in school effectiveness over a long period of time.

But beyond providing suggestions on how the methodology of EER could be improved in order to develop and test theoretical models explaining educational effectiveness, this paper also draws attention to the importance of investigating changes in the effectiveness status of schools. Rather than treating achievement of a single group of students as the dependent variable, implying stability in school effectiveness, our attempt to explain changes in the effectiveness status of schools reveals that the main aim of effectiveness studies should move from understanding variation in the effectiveness status of a number of schools to understand better why changes in the effectiveness status of schools are observed. This study seems to reveal that school factors included in the dynamic model help us understand changes in the effectiveness status of schools. However, further studies are needed to test the generalizability of this finding. It is also acknowledged that by collecting data in more than two periods from the same schools and if possible following them during the whole period of 5 years, we may get a better picture of the complex process of change in the effectiveness status of schools. Mixed research methods could also be employed especially for studying schools where dramatic changes in their effectiveness is observed. For example, in our study, we could have employed a case study approach to find out why one school was dropped from among the most effective to the least effective schools in language (and also dropped from most effective to typical in mathematics).

Finally, an issue that needs further investigation is concerned with our difficulties to identify changes that occur in schools which were among the most effective and dropped to typical. In these schools, no change in the quality of teaching and in the functioning of school factors was observed, and thereby DFA misclassified them as stable schools, expecting them to remain among the most effective. However, by looking at the functioning of school factors in those schools which remained among the most effective, one can observe that significant improvement in the functioning of school factors took place. One could therefore argue that schools cannot remain among the most effective unless improvement in the functioning of school factors is observed. This assumption is also supported by the results of studies investigating changes in the effectiveness status of schools by using a mixed-methods approach (e.g., Teddlie & Stringfield, 1993; Hargreaves & Goodson, 2006). However, research is needed to test this assumption further and help us explain better the dynamic nature of educational effectiveness. Such studies may not only contribute to the establishment of a knowledge base of effective school improvement efforts but may also encourage schools to use the knowledge in the field about ''what works in education and why'' in order to develop strategies to improve their effectiveness.

Notes on contributors

Dr Bert Creemers is an Honorary Professor in Educational Sciences at the University of Groningen, The Netherlands. He is the founding editor of the journals *School Effectiveness and School Improvement* and *Educational Research and Evaluation* and board member of other journals in educational research. His main research interests are educational effectiveness, especially teacher effectiveness, and educational evaluation and improvement.

Dr Leonidas Kyriakides is Associate Professor of Educational Research and Evaluation at the University of Cyprus. His main research interests are modelling educational effectiveness, research methods in educational effectiveness, and educational evaluation and improvement. Currently, his research agenda is concerned with the development of a dynamic model of educational effectiveness and the application of effectiveness research to promote quality and equity in education. He is a member of the editorial board of various international journals with referee systems, the chair of the AERA Special Interest Group on "School Effectiveness and School Improvement", and board member of the International Congress for School Effectiveness and Improvement (ICSEI).

References

Adams, R.J., & Khoo, S.T. (1996). *Quest: The interactive test analysis system*. Camberwell, Victoria, Australia: ACER.

Andrich, D. (1988). A general form of Rasch's Extended Logistic Model for partial credit scoring. *Applied Measurement in Education*, *1*, 363–378.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bentler, P.M. (1995). *EQS: Structural equations program manual*. Encino, CA: Multivariate Software Inc.

Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York, NY: Macmillan.

Brown, M.W., & Mels, G. (1990). *RAMONA PC: User manual*. Pretoria, South Africa: University of South Africa.

Burry, J.A., & Shaw, D. (1988). *Defining teacher effectiveness on a continuum: A Rasch model approach*. Paper presented at the National Council on Measurement in Education conference, New Orleans, LA.

Creemers, B.P.M. (1994). *The effective classroom*. London, UK: Cassell.

Creemers, B.P.M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement*, *17*, 347–366.

Creemers, B.P.M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK: Routledge.

Creemers, B.P.M., & Kyriakides, L. (2010). School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research*, *54*, 263–294.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability scores and profiles*. New York, NY: Wiley.

Den Brok, P., Brekelmans, M., Levy, J., & Wubbels, T. (2002). Diagnosing and improving the quality of teachers' interpersonal behavior. *The International Journal of Educational Management*, *16*, 176–184.

Donaldson, L. (2001). *The contingency theory of organizations: Foundations for organisational science*. Thousand Oaks, CA: Sage.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London, UK: Edward Arnold.

Gray, J., Goldstein, H., & Jesson, D. (1996). Changes and improvements in schools' effectiveness: Trends over five years. *Research Papers in Education*, *11*, 35–51.

Gray, J., Goldstein, H., & Thomas, S. (2001). Predicting the future: The role of past performance in determining trends in institutional effectiveness at A level. *British Educational Research Journal*, *27*, 391–405.

Hargreaves, A., & Goodson, I. (2006). Educational change over time? The sustainability and nonsustainability of three decades of secondary school change and continuity. *Educational Administration Quarterly*, *42*, 3–41.

Joyce, B., Weil, M., & Calhoun, E. (2000). *Models of teaching*. Boston, MA: Allyn & Bacon.

Keeves, J.P., & Alagumalai, S. (1999). New approaches to measurement. In G.N. Masters & J.P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 23–42). Oxford, UK: Pergamon.

Klecka, W.R. (1980). *Discriminant analysis: Quantitative applications in the social sciences*. Newbury Park, CA: Sage.

Kyriakides, L. (1999). The management of curriculum improvement in Cyprus: A critique of a "centre-periphery" model in a centralised system. In T. Townsend, P. Clarke, & M. Ainscow (Eds.), *Third millennium schools: A world of difference in effectiveness and improvement* (pp. 107–124). Lisse, The Netherlands: Swets & Zeitlinger.

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: A step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement*, *19*, 429–446.

Kyriakides, L., & Creemers, B.P.M. (2008). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement*, *19*, 183–205.

Kyriakides, L., & Creemers, B.P.M. (2009, August). *Explaining stability and changes in schools: A follow-up study testing the validity of the dynamic model*. Paper presented at the EARLI 2009 conference, Amsterdam, The Netherlands.

Kyriakides, L., Creemers, B.P.M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education*, *25*, 12–23.

Kyriakides, L., Creemers, B.P.M., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *British Educational Research Journal*, *36*, 807–830.

Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research*, *48*, 1–20.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, NJ: Prentice Hall.

Muijs, D., & Reynolds, D. (2001). *Effective teaching: Evidence and practice*. London, UK: Sage.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York, NY: Macmillan.

Scheerens, J. (1992). *Effective schooling: Research, theory and practice*. London, UK: Cassell.

Schoenfeld, A.H. (1998). Toward a theory of teaching-in-context. *Issues in Education*, *4*(1), 1–94.

Shavelson, R.J., Webb, N.M., & Rowley, G.L. (1989). Generalizability theory. *American Psychologist*, *44*, 922–932.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Stringfield, S.C., & Slavin, R.E. (1992). A hierarchical longitudinal model for elementary school effects. In B.P.M. Creemers & G.J. Reezigt (Eds.), *Evaluation of educational effectiveness* (pp. 35–69). Groningen, The Netherlands: ICO.

Teddlie, C., & Stringfield, S. (1993) *Schools make a difference: Lessons learned from a ten-year study of school effects*. New York, NY: Teachers College Press.

Thomas, S. (2001). Dimensions of secondary school effectiveness: Comparative analyses across regions. *School Effectiveness and School Improvement*, *12*, 285–322.

Thomas, S., Peng, W.J., & Gray, J. (2007). Modelling patterns of improvement over time: Value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education*, *33*, 261–295.

Wang, W.C., & Cheng, Y.Y. (2001). Measurement issues in screening outstanding teachers. *Journal of Applied Measurement*, *2*, 171–186.

Wright, B.D. (1985). Additivity in psychological measurement. In E.E. Roskam (Ed.), *Measurement and personality assessment* (pp. 101–112). Amsterdam, The Netherlands: Elsevier Science Publishers.

Appendix 1. Table for descriptive statistics for the dataset of each study and statistical figures of tests used to compare the background characteristics of the two sample

Characteristics of sample	Original study	Follow-up study	Statistical figures emerged from comparing the two samples
Original Sample			
Number of pupils	2,503	2,716	Non Applicable
Number of teachers of Grade 6	108	112	Non Applicable
Number of teachers	364	387	Non Applicable
Sample used in the analysis			
Number of pupils	2,369	2,682	Non Applicable
Percentage of girls	1,239 (52.3%)	1,432 (53.4%)	Chi-square test: ($X^2 = 0.60$, $df = 1$, $p = 0.44$)
Educational background of father			
Graduate of a primary school	829 (35%)	912 (34%)	Kolmogorov-Smirnov two-sample test (K-S Z = 0.364, $p = 0.999$)
Graduate of secondary school	900 (38%)	1,073 (40%)	
Graduate of a college/ university	640 (27%)	697 (26%)	
Educational background of mother			
Graduate of a primary school	805 (34%)	912 (34%)	Kolmogorov-Smirnov two-sample test (K-S Z = 0.341, $p = 0.999$)
Graduate of secondary school	995 (42%)	1,100 (41%)	
Graduate of a college/ university	569 (24%)	670 (25%)	
Father occupation			
Occupations held by working class	782 (33%)	939 (35%)	Kolmogorov-Smirnov two-sample test (K-S Z = 0.710, $p = 0.695$)
Occupations held by middle class	876 (37%)	965 (36%)	
Occupations held by upper-middle class	711 (30%)	778 (29%)	
Mother occupation			
Occupations held by working class	875 (37%)	938 (35%)	Kolmogorov-Smirnov two-sample test (K-S Z = 0.682, $p = 0.740$)
Occupations held by middle class	877 (37%)	1,073 (40%)	
Occupations held by upper-middle class	617 (26%)	671 (25%)	
Financial situation of the family	$M = 2.02$ $SD = 1.12$	$M = 1.98$ $SD = 1.09$	t test for independent samples ($t = 1.28$, $df = 5049$, $p = 0.20$)

Appendix 2. Table for descriptive statistics for student achievement, quality of teaching, and school factors emerged from the data of each study separately

	Original study		Follow-up study	
Variables	*M*	*SD*	*M*	*SD*
Student achievement in mathematics (initial)	−0.22	0.98	−0.21	1.01
Student achievement in mathematics (final)	0.17	0.96	0.20	0.98
Student achievement in Greek language (initial)	−0.07	0.99	−0.09	1.00
Student achievement in Greek language (final)	0.38	0.96	0.37	0.97
Quality of teaching in mathematics	−0.06	0.98	0.02	0.95
Quality of teaching in Greek language	−0.12	1.02	−0.10	0.98
School Policy of Teaching				
A) Frequency	3.46	1.02	3.51	1.06
B) Stage	2.97	1.12	3.02	1.08
C) Focus	3.00	0.99	2.97	0.96
D) Quality	2.85	1.05	2.92	1.01
E) Differentiation	2.54	0.92	2.61	0.90
Evaluation of school policy of teaching				
A) Frequency	3.24	1.02	3.14	1.13
B) Stage	3.05	1.06	3.02	1.04
C) Focus	2.87	0.94	2.94	0.95
D) Quality	2.78	0.99	2.99	1.02
E) Differentiation	2.20	0.92	2.24	0.90
Policy on student behaviour outside the classroom				
A) Frequency	2.87	0.94	2.99	0.98
B) Stage	2.80	0.96	2.90	0.97
C) Focus	2.45	0.91	2.68	0.94
D) Quality	2.62	0.93	2.65	0.97
E) Differentiation	2.02	0.90	2.12	0.93
Policy on teacher collaboration				
A) Frequency	3.42	0.99	3.45	1.02
B) Stage	3.24	1.02	3.26	1.00
C) Focus	2.88	1.09	2.90	1.03
D) Quality	3.01	1.02	3.05	0.99
E) Differentiation	2.45	0.90	2.52	0.93
Policy on parental involvement				
A) Frequency	3.61	1.09	3.51	1.05
B) Stage	3.02	1.08	3.09	1.04
C) Focus	2.99	1.01	3.02	1.05
D) Quality	2.89	0.96	2.82	1.02
E) Differentiation	2.04	0.92	2.01	0.97
Policy on Educational Resources				
A) Frequency	3.20	1.09	3.32	1.12
B) Stage	3.02	1.01	3.09	1.05
C) Focus	2.89	1.02	2.97	1.07
D) Quality	2.66	0.99	2.76	1.02
E) Differentiation	2.02	0.93	2.13	1.00
Evaluation of SLE				
A) Frequency	3.29	1.02	3.59	1.03
B) Stage	3.12	1.03	3.16	1.00
C) Focus	3.00	1.03	3.06	1.03
D) Quality	2.89	1.04	2.92	1.00
E) Differentiation	2.07	0.95	2.12	0.90

Notes: The extended logistic model of Rasch was used to analyse data on *student achievement* and *quality of teaching*, and person estimates are given in logits. (For model identification, the mean of the difficulties of all items of each scale was fixed at zero.) A 5-point Likert scale was used to measure each dimension of each *school factor*.

Appendix 3. Parameter estimates and (standard errors) for the analyses of student achievement in Greek language and mathematics of each study

	Follow-Up Study				Original Study			
	Greek Language		Mathematics		Greek Language		Mathematics	
Factors	Model 0	Model 1	Model 0	Model 1	Model 0	Model 1	Model 0	Model 1
Fixed part/intercept	.42 (.06)	.31 (.05)	−.41 (.09)	−.33 (.08)	−.31 (.08)	−.22 (.08)	.35 (.05)	.28 (.05)
Student Level								
Prior knowledge		.43 (.11)		.38 (.05)		.39 (.05)		.45 (.10)
Sex (boys = 0, girls = 1)		.12 (03)		N.S.S.*		.19 (.08)		−.14 (.06)
SES		.19 (.10)		.19 (.07)		.30 (.06)		.30 (.12)
Classroom Level								
Context								
Average knowledge		.10 (.04)		.10 (.04)		.12 (.05)		.28 (.10)
Average SES		N.S.S.		N.S.S.		.08 (.03)		.12 (.05)
Percentage of girls		N.S.S.		N.S.S.		N.S.S.		−.05 (.02)
School level								
Context								
Average SES		N.S.S.		N.S.S.		N.S.S.		N.S.S.
Average knowledge		.07 (.03)		.06 (.03)		.09 (.04)		.11 (.05)
Percentage of girls		N.S.S.		N.S.S.		N.S.S.		N.S.S.
Variance								
School	9.5%	8.8%	9.1%	8.7%	9.0%	8.2%	11.2%	9.8%
Class	18.0%	14.9%	16.8%	14.1%	14.7%	10.3%	14.8%	10.0%
Student	72.5%	29.0%	74.1%	29.5%	76.3%	31.3%	74.0%	30.2%
Explained		47.3%		47.7%		50.2%		50.0%
Significance test								
X^2	924.3	605.2	915.6	499.6	815.6	507.2	1144.9	795.5
Reduction		319.1		416.0		308.4		349.4
Degrees of freedom		5		4		6		7
p value		.001		.001		.001		.001

Note: Each model was estimated without the variables that did not have a statistically significant effect at the .05 level.
N.S.S. = No statistically significant effect at the .05 level.