Routledge
Taylor & Francis Group

Scandinavian Journal of Educational Research

ISSN: 0031-3831 (Print) 1470-1170 (Online) Journal homepage: http://www.tandfonline.com/loi/csje20

School Factors Explaining Achievement on Cognitive and Affective Outcomes: Establishing a Dynamic Model of Educational Effectiveness

Bert Creemers & Leonidas Kyriakides

To cite this article: Bert Creemers & Leonidas Kyriakides (2010) School Factors Explaining Achievement on Cognitive and Affective Outcomes: Establishing a Dynamic Model of Educational Effectiveness, Scandinavian Journal of Educational Research, 54:3, 263-294, DOI: 10.1080/00313831003764529

To link to this article: http://dx.doi.org/10.1080/00313831003764529

Published online: 21 Jul 2010.

Scandinavian Journal of Educational Research
Vol. 54, No. 3, June 2010, 263–294

Routledge
Taylor & Francis Group

School Factors Explaining Achievement on Cognitive and Affective Outcomes: Establishing a Dynamic Model of Educational Effectiveness

Bert Creemers
University of Groningen

Leonidas Kyriakides
University of Cyprus

The dynamic model of educational effectiveness defines school level factors associated with student outcomes. Emphasis is given to the two main aspects of policy, evaluation, and improvement in schools which affect quality of teaching and learning at both the level of teachers and students: a) teaching and b) school learning environment. Five measurement dimensions are used to define each factor: frequency, stage, focus, quality and differentiation. This paper reports the results of a longitudinal study testing the validity of the dynamic model at the school level. The multidimensional approach to measure the school level factors was supported and most of the factors and their dimensions were found to be associated with student achievement in different learning outcomes. Implications for the development of the dynamic model and for educational practice are drawn.

Keywords: school effectiveness, school policy, school learning environment, multilevel modeling

Educational Effectiveness Research (EER) addresses the questions on what works in education and why. Over the last two decades EER has been improved considerably by the criticism on research design, the sampling, and statistical techniques. Methodological advances, particularly the availability of particular software for the analysis of multilevel data, have enabled more efficient estimates of teacher and school differences in student achievement to be obtained (Goldstein, 2003). There is also substantial agreement as to appropriate methods of estimating school differences/effects and the kinds of data required for valid comparisons to be made (Hopkins, Reynolds, & Gray, 1999). As far as the theoretical component of the field is concerned, progress was made by a more precise definition of the concepts used and the relations between the concepts (e.g. Creemers, 1994; Levin & Lezotte, 1990; Scheerens & Bosker, 1997). One of the most influential theoretical models of the field was developed in the 1990s and attempted to provide a comprehensive view of the education by relating factors operating at different levels to outcomes of schooling (Creemers, 1994). During the last decade six studies, conducted in two different countries, (de Jong, Westerhof, & Kruiter, 2004; Driessen & Sleegers, 2000; Kyriakides, 2005; Kyriakides, Campbell, & Gagatsis, 2000; Kyriakides & Tsangaridou, 2008; Reezigt,

Bert Creemers, Faculty of Behavioural and Social Sciences, University of Groningen; Leonidas Kyriakides, Department of Education, University of Cyprus.

Correspondence concerning this article should be addressed to Leonidas Kyriakides, Department Of Education, University Of Cyprus, Nicosia 1678, Cyprus. E-mail: kyriakid@ucy.ac.cy

ISSN 0031-3831 print/ISSN 1470-1170 online

DOI: 10.1080/00313831003764529
http://www.informaworld.com

Guldemond, & Creemers, 1999) provided some support to the validity of the comprehensive model. A synthesis of these studies has revealed suggestions for further development of the model especially by taking into account the dynamic nature of educational effectiveness (Kyriakides, 2008). In this context, Creemers and Kyriakides (2008) developed a dynamic model of educational effectiveness that attempts to define the dynamic relations between the multiple factors found to be associated with effectiveness. A longitudinal study testing the validity of the dynamic model has been conducted and provided support for the validity of the model at the classroom level. In this paper, the results of the study testing the model at the school level are presented and implications for the development of the model and for educational practice are drawn.

The Dynamic Model of Educational Effectiveness: An Overview

The Essential Characteristics of the Dynamic Model

The dynamic model takes into account the fact that effectiveness studies conducted in several countries reveal that the influences on student achievement are multilevel (Teddlie & Reynolds, 2000). Therefore, the dynamic model is multilevel in nature and refers to 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 analyzed. 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. It is also taken into account that the teaching and learning situation is influenced by the wider educational context in which students, teachers, and schools are expected to operate. Factors such as the values of the society for learning and the importance attached to education play an important role both in shaping teacher and student expectations.

The interrelations between the components of the model are also illustrated. In this way, the model assumes that factors at the school and context level have both direct and indirect effects on student achievement since they are able not only to influence student achievement directly but also to influence the teaching and learning situations. Therefore, teaching is emphasized and the description of the classroom level refers mainly to the behavior of the teacher in the classroom and especially to his/her contribution in promoting learning at the classroom level. Moreover, defining factors at the classroom level is seen as a prerequisite for defining the school and the system level. 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 educational effectiveness. A brief description of these four dimensions is given below. Specifically, two aspects of the *focus* dimension are taken into account. The first one refers to the specificity of the activities associated with the functioning of the factor, whereas the second one with the number of purposes for which an activity takes place.

The *stage* at which tasks associated with a factor take place is also examined. It is expected that the factors need to take place over a long period of time to ensure that they have a continuous direct or indirect effect on student learning. The *quality* refers to properties of the specific factor itself, as these are discussed in the literature. Finally, *differentiation* refers to the extent to which activities associated with a factor are implemented in the same way for all the subjects involved with it (e.g. all the students, teachers, schools). It is expected that adaptation to specific needs of each subject or group of subjects will increase the successful implementation of a factor and ultimately maximize its effect on student learning outcomes.

School Factors in the Dynamic Model

The definition of the school level is based on the assumption that factors at the school level are expected to have not only direct effects on student achievement but also mainly indirect effects. School factors are expected to influence classroom-level factors, especially the teaching practice. This assumption is based on the fact that EER has shown that the classroom level is more significant than the school level (e.g. Kyriakides et al., 2000; Teddlie & Reynolds, 2000). Moreover, defining factors at the classroom level is seen as a prerequisite for defining the school level (Creemers, 1994). Therefore, the dynamic model refers to factors at the school level that are related to the same key concepts of quantity of teaching, provision of learning opportunities, and quality of teaching that are used to define the classroom-level factors of the dynamic model. Specifically, emphasis is given to the following two main aspects of the school policy, which affect learning at both the teacher and student level: (1) school policy for teaching, and (2) 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 (see Creemers & Kyriakides, 2008). 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. This factor does not imply that each school should simply develop formal documents to install the 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 from them. Support offered to teachers and other stakeholders to implement the school policy is also an aspect of these two factors.

Based on the assumption that the essence of a successful organization in the modern world is the search for improvement (Hopkins, 2001), we also examine the processes and the activities that take place in the school in order to improve the teaching practice and the School Learning Environment (SLE). For this reason, the processes that are used to evaluate the school policy for teaching and the SLE are investigated. Thus, the following four factors at the school level are included in the model:

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

Figure 1 illustrates the interrelations among the school factors, which are briefly described below (for more information see Creemers and Kyriakides, 2008). It is, finally, important to note that the inclusion of these factors is also based on the results of a synthesis of 123 studies on school effectiveness conducted in different countries since 1986 (see Kyriakides, Creemers, Antoniou, & Demetriou, in press). This meta-analysis has provided support to the importance of the factors included in the model and also revealed that the effect sizes of other factors not taken into account by the dynamic model are extremely low. For example, the average effect size of leadership in this meta-analysis was 0.07 and this finding is in line with the results of two earlier meta-analyses, which were also conducted by using multilevel modeling approaches (see Scheerens, Seidel, Witziers, Hendriks, & Doornekamp, 2005; Witziers, Bosker, & Kruger, 2003). Similar results were obtained from studies that were conducted in order to measure indirect effects of leadership on student achievement (Leithwood & Jantzi, 2006). Therefore, the model is not

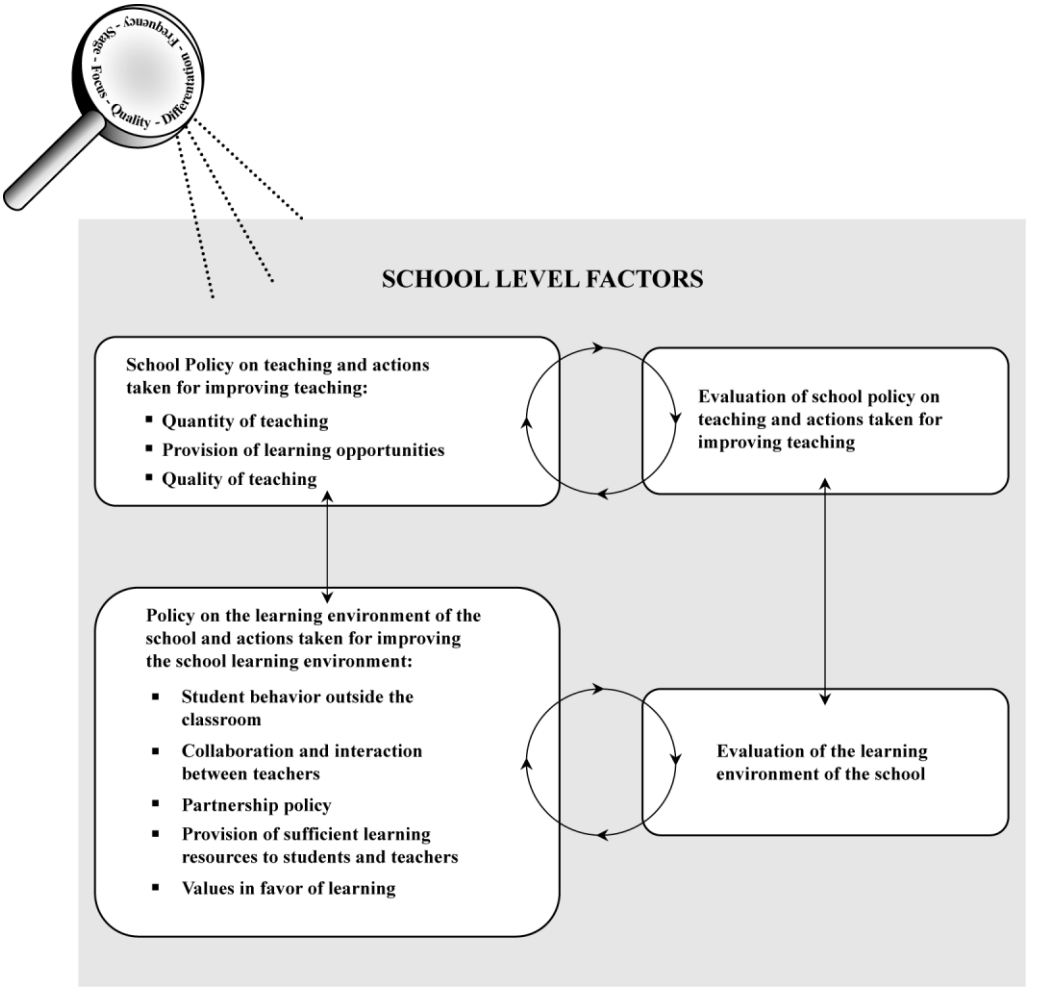

Figure 1. Factors of the dynamic model operating at the school level.

concerned with who is in charge of designing and/or implementing the school policy, but with the content of the school policy and the type of activities that take place in school. This reveals one of the major assumptions of the model, which is not focused on individuals as such, but on the effects of the actions that take place at classroom/school/context levels.

School Policy for Teaching and Actions Taken for Improving Teaching

Since the definition of the dynamic model at the classroom level (see Creemers & Kyriakides, 2006) refers to factors related to the key concepts of quality, time on task, and opportunity to learn, the model attempts to investigate aspects of school policy for teaching associated with quantity of teaching, provision of learning opportunities, and quality of teaching. Actions taken for improving the above three aspects of teaching practice, such as the provision of support to teachers for improving their teaching skills, are also taken into account. More specifically, the following aspects of school policy on quantity of teaching are taken into account:

- school policy on the management of teaching time (e.g. lessons start on time and finish on time; there are no interruptions of lessons for staff meetings and/or for preparation of school festivals and other events);
- policy on student and teacher absenteeism;
- policy on homework; and
- policy on lesson schedule and timetable.

School policy on provision of learning opportunities is measured by looking at the extent to which the school has a mission concerning the provision of learning opportunities, which is reflected in its policy on curriculum. We also examine school policy on long-term and short-term planning and school policy on providing support to students with special needs. Furthermore, the extent to which the school attempts to make good use of school trips and other extra-curricular activities for teaching/learning purposes is investigated. Finally, school policy on the quality of teaching is seen as closely related to the classroom-level factors of the dynamic model, which refer to the instructional role of teachers (Creemers & Kyriakides, 2006).

Therefore, the way school policy for teaching is examined reveals that effective schools are expected to make decisions on maximizing the use of teaching time and the learning opportunities offered to their students. In addition, effective schools are expected to support their teachers in their attempt to help students learn by using effective teaching practices. In this context, the definition of this factor implies that we should measure the extent to which: (1) the school makes sure that teaching time is offered to students, (2) learning opportunities beyond those offered by the official curricula are offered to the students, and (3) the school attempts to improve the quality of teaching practice.

School Policy for Creating a SLE and Actions Taken for Improving the SLE

School climate factors have been incorporated in effectiveness models in different ways. Stringfield (1994) defines the school climate very broadly as the total environment of the school. This makes it difficult to study specific factors of the school climate and examine their impact on student achievement. The dynamic model refers to the extent to which a

learning environment has been created in the school. This element of school climate is seen as the most important predictor of school effectiveness since learning is the key function of a school (Linnakyla, Malin, & Taube, 2004). Moreover, EER has shown that effective schools are able to respond to the learning needs of both teachers and students and to be involved in systematic changes of the school's internal processes in order to achieve educational goals more effectively in conditions of uncertainty (Harris, 2001). In this context, the following five aspects, which define the SLE, are taken into account:

(1) student behavior outside the classroom,
(2) collaboration and interaction between teachers,
(3) partnership policy (i.e., relations of school with community, parents, and advisors),
(4) provision of sufficient learning resources to students and teachers, and
(5) values in favor of learning.

The first three aspects refer to the rules that the school has developed for establishing a learning environment inside and outside the classrooms. Here the term learning does not refer exclusively to student learning. For example, collaboration and interaction between teachers may contribute in their professional development (i.e., learning of teachers) but may also have an effect on teaching practice and thereby may improve student learning. The fourth aspect refers to the policy on providing resources for learning. The availability of learning resources in schools may not have only an effect on student learning but may also encourage the learning of teachers. For example, the availability of computers and software for teaching geometry may contribute to teacher professional development since it encourages teachers to find ways to make good use of the software in their teaching practice and thereby to become more effective. The last aspect of this factor is concerned with the strategies that the school has developed in order to encourage teachers and students to develop positive attitudes towards learning.

Following a similar approach as the one concerned with school policy on teaching, the dynamic model attempts to measure the school policy for creating a SLE. Actions taken for improving the SLE beyond the establishment of policy guidelines are also taken into account. Specifically, actions taken for improving the SLE can be directed at: (1) changing the rules in relation to the first three aspects of the SLE factor mentioned above, (2) providing educational resources (e.g. teaching aids, educational assistance, new posts), and/or (3) helping students/teachers develop positive attitudes towards learning. For example, a school may have a policy for promoting teacher professional development, but this might not be enough, especially if some teachers do not consider professional development as an important issue. In this case, actions should be taken to help teachers develop positive attitudes towards learning, which may help them become more effective.

The last two overarching school factors of the dynamic model refer to the mechanisms used to evaluate the functioning of the first two overarching factors. Creemers (1994) claims that control is one of the major principles operating in generating educational effectiveness. This implies that goal attainment and the school climate should be evaluated (Grosin, 1993; Torres & Preskill, 2001). It was therefore considered important to treat evaluation of policy for teaching and of other actions taken to improve teaching practice as well as evaluation of the SLE as overarching factors operating at school level. Data emerging from these evaluation mechanisms are expected to help schools develop their policies and

improve the teaching practice at the classroom level as well as their SLE (see Creemers & Kyriakides, 2008).

Research Aims

A criticism that may arise from the theoretical background and the outline of the dynamic model concerns the complexity of the model and the difficulties of testing it empirically. As a consequence, we conducted a longitudinal study on teacher and school effectiveness in Cyprus in order to investigate the validity of the dynamic model. This study does not only attempt to investigate educational effectiveness in mathematics and language, but also measures concerned with both cognitive and affective aims of religious education are taken into account. In this way we can find out whether each factor and its dimensions are associated with achievement in different subjects and in both cognitive and affective outcomes. Thus, we can investigate the extent to which the dynamic model could be considered as a generic model (Scheerens & Bosker, 1997).

The results of the first phase of this study, which was concerned with the validity of the model at the classroom level, not only reveal that the dynamic model is a theoretical model that can be put into testing, but also provided support for the construct validity of the five measurement dimensions of most effectiveness factors at the classroom level (Kyriakides & Creemers, 2008). Furthermore, this study revealed the added value of using the five dimensions to measure the classroom-level factors for explaining variation of student achievement in different outcomes. Testing the validity of the model at the classroom level can be seen as the starting point for the development and the testing of the dynamic model at the school and the system level. Thus, the second phase of this longitudinal study, which is presented in this paper, attempts to test the validity of the dynamic model at the school level. Specifically, the second phase of this study investigates:

(1) the extent to which each school-level factor can be defined by reference to the five dimensions of the model, and
(2) the type(s) of relations that each school factor and its dimensions have with student learning outcomes in mathematics, language, and religious education.

Methods

Participants

Stratified sampling (Cohen, Manion, & Morrison, 2000) was used to select 52 out of 191 Cypriot primary schools, but only 50 schools participated in the study. All the grade 5 students (n = 2,503) from each class (n = 108) of the school sample were chosen. The chi-square test did not reveal any statistically significant difference between the research sample and the population in terms of students' sex ($X^2 = 0.84$, $df = 1$, $p = 0.42$). Moreover, the t-test did not reveal any statistically significant difference between the research sample and the population in terms of the size of class ($t = 1.21$, $df = 507$, $p = 0.22$). Although this study refers to other variables such as the socio-economic status (SES) of students and their achievement levels in different outcomes of schooling, there is no national data about these characteristics of the Greek Cypriot students. Therefore, it was not possible to examine whether the sample was nationally representative in terms of any other characteristic except from students' sex and the size of the class. However, it can be claimed that a nationally

representative sample of Greek Cypriot grade 5 students in terms of these two characteristics was drawn.

Dependent Variables: Student Achievement in Mathematics, Greek Language and Religious Education at the end of Grade 6

Data on student achievement in mathematics, Greek language, and religious education were collected by using external forms of assessment designed to assess knowledge and skills in mathematics, Greek language, and religious education, which are identified in the Cyprus Curriculum for grade 6 students (Ministry of Education, 1994). Student achievement in relation to the affective aims included in the Cyprus curriculum for religious education was also measured. Criterion-reference tests are more appropriate than norm-referenced tests for relating achievement to what a student should know and for testing competence rather than general ability. Thus, criterion-reference tests were constructed and students were asked to answer at least two different tasks related to each objective in the teaching programs of mathematics, Greek language, and religious education for grade 6 students. Scoring rubrics, used to differentiate among four levels of task proficiency (0–3) on each task were also constructed. Thus, ordinal data about the extent to which each student had acquired each skill included in the grade 6 curriculum of mathematics, Greek language, and religious education were collected. The construction of the tests was subject to controls for reliability and validity. Specifically, the Extended Logistic Model of Rasch (Andrich, 1988) was used to analyze the emerging data in each subject separately. Four scales, which refer to student knowledge in mathematics, Greek language, and religious education and to student attitudes towards religious education, were created and analyzed for reliability, fit to the model, meaning, and validity. Analysis of the data revealed that each scale had relatively satisfactory psychometric properties (see Creemers & Kyriakides, 2008). Thus, for each student four different scores for his/her achievement at the end of grade 6 were generated by calculating the relevant Rasch person estimate in each scale. The written tests are available upon request from the second author. It is also important to note that none of the respondents gained a full score in any of these tests. Moreover, less than 5% of the students achieved over 80% of the maximum score, and less than 12% of the students achieved over 70% of the maximum score in each test. Therefore, the ceiling effect was less probable. The floor effect was also not real in the data, because no student showed full zero-performance in any test.

Explanatory Variables at Student Level

Aptitude. Aptitude refers to the degree to which a student is able to perform the next learning task (Gustafsson & Balke, 1993). For the purpose of this study, it consists of prior knowledge of each subject (i.e. mathematics, Greek language, and religious education) and prior attitudes towards religious education emerged from student responses to the external forms of assessment administered to students when they were at the end of grade 5. Thus, external forms of assessment were also used to measure the achievement of our sample when they were at the end of grade 5. The Extended Logistic Model of Rasch was used to analyze the emerging data in each subject separately, and four scales, which refer to student knowledge in mathematics, Greek language, religious education, and to student attitudes towards religious education at the end of grade 5, were created. The psychometric properties

of these scales were satisfactory (see Creemers & Kyriakides, 2008). Thus, for each student four different scores for his/her achievement at the end of grade 5 were generated, by calculating the relevant Rasch person estimate in each scale.

Student background factors. Information was collected on two student background factors: sex (0 = boys, 1 = girls), and SES. Five SES variables were available: father's and mother's education level (i.e., graduate of a primary school, graduate of secondary school, or graduate of a college/university), the social status of father's job, the social status of mother's job, and the economic 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 that have relatively similar sizes: occupations held by working class (33%), occupations held by middle class (37%), and occupations held by upper-middle class (30%). 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.

Explanatory Variables at School Level

The explanatory variables that 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 during the last term of the school year. The questionnaire was designed in such a way that information about the five dimensions of the four school-level factors of the dynamic model could be collected. A Likert scale was used to collect data on teachers' perceptions of the school level factors. We also attempted to generate data on school factors by collecting documents about policy and actions at school level and by conducting a content analysis. However, we did not succeed in collecting the documents in a sufficient way mainly because some headteachers were not willing to provide the documents in order to protect privacy of their students and teachers. Thus, data on school factors are only based on teacher questionnaires and limitations of using perceptual methods to measure school factors should be acknowledged. Nevertheless, the quality, and especially the generalizability, of the data were tested systematically, as is explained below. In addition, perceptual measures were found to produce valid data in other areas within education such as measures of teacher interpersonal behavior and/or quality of teaching through student questionnaires (e.g. den Brok, Brekelmans, Levy, & Wubbels, 2002; Marsh & Roche, 1997).

Of the 364 teachers approached, 313 responded, a response rate of 86%. The chi-square test did not reveal any statistically significant difference between the distribution of the teacher sample that 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 ($X^2 = 57.12$, $df = 49$, $p = .38$). It can be claimed that our sample is representative to the whole population in terms of how the teachers are distributed in each of these 50 schools. Moreover, the missing responses to each questionnaire item were very small (less than 5%).

Results

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 in the first part of the results section. This section enables us to identify the extent to which the

proposed measurement dimensions can be used to define the functioning of the school factors of the model. The second part of this section is an attempt to identify the extent to which the school factors of the dynamic model and their dimensions show the expected effects upon each dependent variable (i.e., student achievement in each outcome of schooling).

The Questionnaire Measuring Teacher Views About the School Factors

Reliability, consistency, and variance at class 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. It was found that for 132 out of the 140 questionnaire items, the object of measurement was the school. It is important to note that six out of the eight items for which the generalizability of the data at the level of the school was questionable had very small variance and referred to the school policy in relation to the development of positive values towards learning. Since only eight items were used to collect data on teacher views about this factor, it was decided to drop all the items that refered to this factor. We also dropped the data that emerged from the other two items that were found not to be generalizable at the level of school. These two items were concerned with the focus dimension of two other factors (i.e., school policy for teaching, and evaluation of the SLE).

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 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. The results are presented in Table 1. We can observe that reliability coefficients were very high (around .90). Moreover, the reliability of the focus dimension of the factors concerned with the school policy on teaching and the focus dimension of the evaluation of the SLE were somewhat lower, while the reliability of the frequency dimension of the factor concerned with the evaluation of school policy for teaching was the highest.

Using the Mplus (Muthén & Muthén, 2001) the intra-class correlations of the scales were computed. The intra-class correlations, which indicate what amount of variance of the teacher questionnaire is located at the between-level, are also illustrated in Table 1. We can observe that the percentages of variance at the between-level (school-level) were between 37 and 48. These percentages are rather high compared to other instruments that measure perceptions of people or objects in clustered or interdependent situations (den Brok et al., 2002).

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 teacher-scales. It was found that the scales correlated between 0.10 and 0.35. Moreover, only 71 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, the values of the mean correlation of a scale with the other scales were smaller than .22. This implies that the 45 scales of the questionnaire, which refer to indicators of the five dimensions of the school factors, differed sufficiently.

Table 1
Cronbach Alpha (Reliability), Multilevel Lambda (Consistency), and Intra-Class Correlations (ICC) of Scales Emerging from Teacher Questionnaire Concerned with Each Dimension of Each School Factor at the School Level

School factors	Cronbach alpha					Multilevel Lambda (consistency)					Intra-class correlations				
	Freq	Focus	Stage	Quality	Diff	Freq	Focus	Stage	Quality	Diff	Freq	Focus	Stage	Quality	Diff
School policy for teaching															
Quantity of teaching	.90	.82	.93	.95	.92	.90	.80	.92	.91	.90	.41	.42	.46	.42	.45
Provision of learning opportunities	.91	.82	.87	.90	.88	.88	.81	.88	.87	.89	.39	.37	.45	.45	.41
Quality of teaching	.89	.83	.85	.87	.83	.85	.82	.83	.82	.80	.44	.40	.44	.43	.40
Policy on the school as a learning environment															
Student behavior outside the classroom	.88	.85	.89	.88	.86	.87	.86	.88	.90	.89	.38	.36	.36	.39	.43
Collaboration and interaction between teachers	.87	.84	.88	.87	.84	.85	.83	.84	.85	.87	.37	.36	.39	.38	.41
Partnership policy	.86	.87	.84	.88	.86	.89	.82	.84	.88	.86	.39	.37	.37	.41	.36
Provision of resources	.84	.83	.84	.89	.85	.87	.83	.84	.89	.85	.42	.38	.43	.40	.37
Evaluation of school policy for teaching	.94	.87	.90	.91	.88	.93	.85	.86	.90	.88	.46	.39	.39	.38	.38
Evaluation of SLE	.91	.82	.88	.90	.89	.88	.80	.84	.87	.89	.41	.35	.40	.40	.40

Note. The five dimensions for each school factor are as follows: frequency (Freq), focus, stage, quality and differentiation (Diff).

Construct validity. Using a unified approach to test validation (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Messick, 1989), this study provides construct-related evidence of the questionnaire measuring teacher views of the school factors and their dimensions. For the identification of the factor structure of the questionnaire, Structural Equation Modeling (SEM) analyses were conducted using the structural equations program, EQS (Bentler, 1995). Each model was estimated by using normal theory maximum likelihood methods (ML). The ML estimation procedure was chosen because it does not require an excessively large sample size. 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. The main results of SEM analysis for each factor are presented below.

School Policy for Teaching

A first-order Confirmatory Factor Analysis model designed to test the multidimensionality of research instruments was used to examine the construct validity of the first part of the questionnaire measuring school policy for teaching (Byrne, 1998). Specifically, the model hypothesized that: (1) the 15 variables (i.e., scale scores measuring each dimension of each of the three aspects of this factor) could be explained by five factors concerning the five measurement dimensions of this school factor; (2) each variable would have a nonzero loading on the factor that it was designed to measure, and zero loadings on all other factors; (3) the five factors would be correlated; and (4) measurement errors would be uncorrelated.

The findings of the first order factor SEM analysis generally affirmed the theory upon which the questionnaire was developed. Although the scaled chi-square for the five-factor structure ($X^2 = 123.2$, $df = 80$, $p < .001$) as expected was statistically significant, the values of RMSEA (0.029) and CFI (0.981) met the criteria for acceptable level of fit. Kline (1998) argues that: "even when the theory is precise about the number of factors of a first-order model, the researcher should determine whether the fit of a simpler, one-factor model is comparable" (p. 212). Criteria fit for a one-factor model ($X^2 = 1249.4$, $df = 90$, $p < .001$; RMSEA = 0.141 and CFI = 0.469) provided values that fell outside generally accepted guidelines for model fit. Thus, a decision was made to consider the five-factor structure as reasonable and thereby the analysis proceeded and the parameter estimates were calculated. Figure 2 depicts the five-factor model and presents the factor parameter estimates. All parameter estimates were statistically significant ($p < .001$).

The following observations arise from Figure 2. First, the standardized factor loadings were all positive and moderately high. Their standardized values ranged from 0.63 to 0.81 and the great majority of them were higher than 0.65. Second, the correlations among the five factors were positive and ranged between 0.08 and 0.17. Moreover, the majority of factor inter-correlations were smaller than 0.13. The relatively small values of the factor intercorrelations provided support for arguing the separation of the five measurement dimensions of the school factor concerned with school policy for teaching. In order to test this assumption further, we also tested the fitting of a higher order model that could explain the correlations among the five first-order factors in each analysis. Specifically, this model hypothesized that: (1) responses to the teacher questionnaire could be explained by five

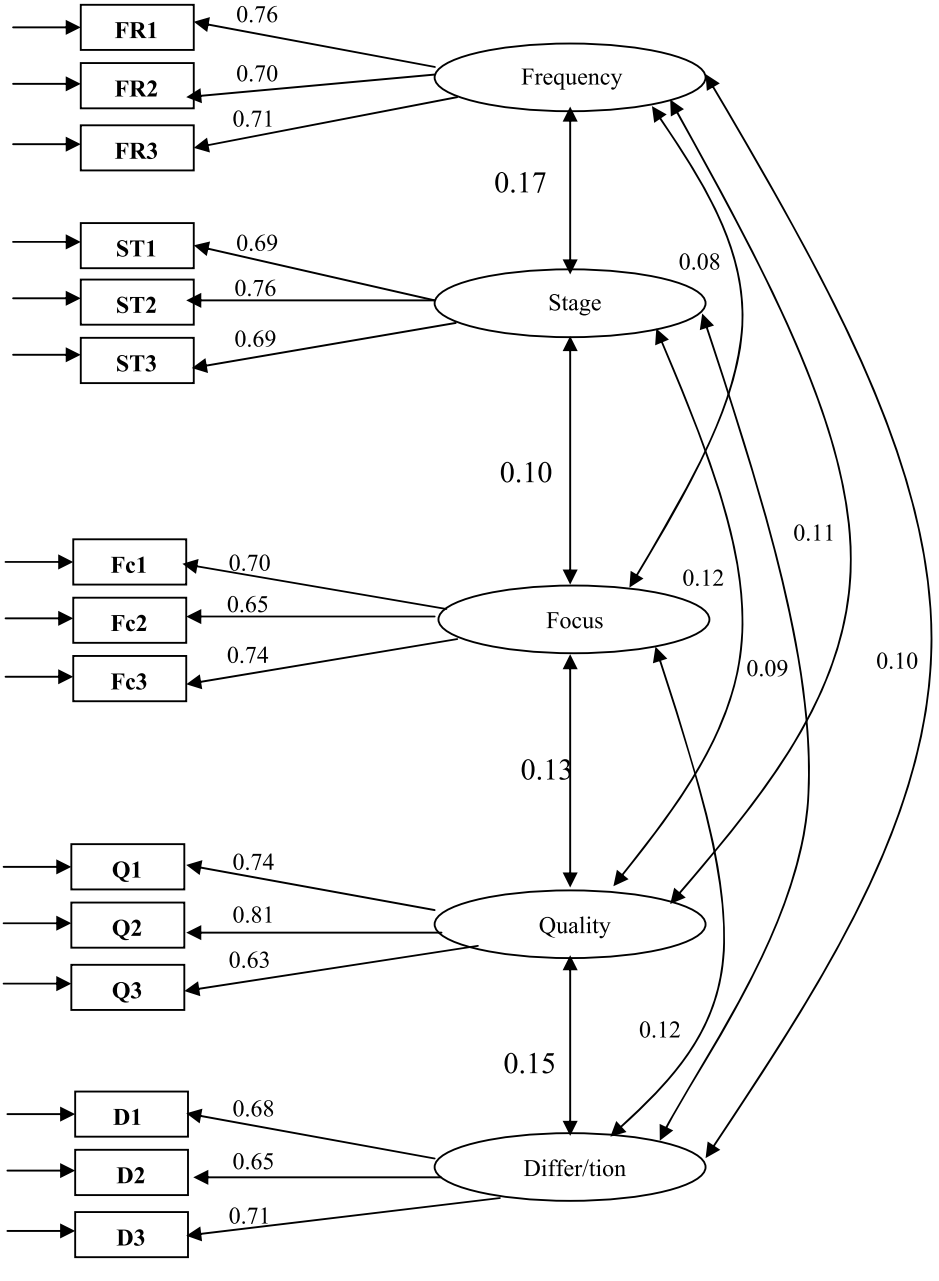

Figure 2. First-order factor model of school policy for teaching with factor parameter estimates.

first-order factors and one second-order factor (i.e., school policy for teaching in general); (2) each item (i.e., sub-scale score) would have a nonzero loading on the factor it was designed to measure, and zero loadings on all other factors; (3) error terms associated with each item would be uncorrelated; and (4) covariation among the five first-order factors

would be explained by their regression on the second order factor. However, the fit statistics of this model (X^2 = 350.4, df = 85, $p < .001$; RMSEA = 0.152 and CFI = 0.782) provided values that fell outside generally accepted guidelines for model fit. Thus, for each school, five scores of the factor concerned with school policy of teaching were generated by aggregating at the school level the factor scores that emerged from teacher responses to the questionnaire.

Evaluation of School Policy on Teaching

A similar procedure to the one used to test the construct validity of the part of the questionnaire measuring the school policy for teaching was used to test the factor concerned with the evaluation of school policy on teaching. The first-order factor structure of the 15 items concerned with the evaluation of the school policy for teaching was investigated in order to determine whether the five proposed measurement dimensions of the dynamic model explain the variability in the items that are logically tied to each other, or whether there is a single latent factor that can better explain the variability in the 15 items. The findings of the first-order factor SEM analysis generally affirmed the assumption of the dynamic model that this factor could be measured in relation to each of the five measurement dimensions. Although the scaled chi-square for the five-factor structure (X^2 = 164.4, df = 80, $p < .05$) was statistically significant, the RMSEA was 0.032 and the CFI was 0.968 and both of them met the criteria for acceptable level of fit. Therefore, validation of the five-order factor structure of this part of the questionnaire provided support to the use of item scores for making inferences about five different measurement dimensions of this factor rather than treating it as a unidimensional construct. Thus, for each school, five scores of its evaluation of school policy for teaching were generated by aggregating at the school-level the factor scores that emerged from teacher responses to the relevant questionnaire items.

School Policy on the Learning Environment of the School

As it has been explained above, five aspects of the SLE are taken into account in defining the factor investigating policy on the learning environment of the school. However, it was possible to generate data about only four of these aspects (see Table 1). Therefore, for each of these four aspects of the SLE, a first-order Confirmatory Factor Analysis model was used in order to find out whether the 15 variables (i.e., subscale scores measuring each dimension of the relevant aspect of SLE) could be explained by five factors concerning the five measurement dimensions of the relevant aspect of SLE. The findings of the first order factor SEM analysis generally affirmed the assumption of the dynamic model that each aspect of SLE could be measured in relation to each of the five measurement dimensions since they provided fit statistic values that were acceptable (i.e., student behavior outside the classroom [X^2 = 116.8, df = 80, $p < .001$; RMSEA = 0.029 and CFI = 0.971]; collaboration between teachers [X^2 = 102.4, df = 80, $p < .001$; RMSEA = 0.019 and CFI = 0.982]; partnership policy [X^2 = 99.7, df = 80, $p < .001$; RMSEA = 0.015 and CFI = 0.984]; provision of learning resources [X^2 = 112.5, df = 80, $p < .001$; RMSEA = 0.026 and CFI = 0.972]), whereas the criteria fit for a one-factor model for each of these four aspects of the SLE provided values that fell outside generally accepted guidelines for model fit. Thus, based on the results of the CFA analysis, for each school, five scores of each aspect of the SLE were generated by aggregating at the school level the factor scores that emerged from teacher responses to the questionnaire.

Evaluation of the Learning Environment of the School

The first-order factor structure of the 14 items concerned with the evaluation of the SLE was investigated in order to determine whether the five proposed measurement dimensions of the dynamic model explain the variability in the items that are logically tied to each other (i.e., refer to the same measurement dimension), or whether there is a single latent factor that can explain better the variability in these items. The null model and the four CFA nested models are presented in Table 2. The null model (Model 1) represents the most restrictive model, with 14 uncorrelated variables measuring the perceptions of teachers about the evaluation of the SLE. Models 2 through 4 are first-order models, and comparisons between the chi-squares of these models helped us evaluate the construct validity of the part of the teacher questionnaire concerned with this school-level factor. Model 5 was a higher-order model and is compared with the lower-order model found to fit better than any other first-order factor model.

The following observations arise from Table 2. First, comparing the null model with Model 2, we can observe that although the overall fit of Model 2 was not acceptable, it was a significant improvement in chi-square compared to the null model. This result can be seen as an indication of the importance of searching for the factor structure of the data emerging from the teacher questionnaire. Second, Model 2 can be compared with Models 3 and 4 to determine the best trait structure of evaluation of SLE that is able to explain better the variability in the 14 questionnaire items. Model 3 represents the five-factor model, which investigates whether each of the 14 items has a nonzero loading on the factor (i.e., measurement dimension) it was designed to measure, and zero loadings on all other factors. The five factors are also correlated but the measurement errors of these items are uncorrelated. The chi-square difference between Models 2 and 3 showed a significant decrease in chi-square and a significant improvement over the one factor only model. Clearly, the use of different dimensions to measure this factor is supported since their treatment as separate factors helps us increase the amount of covariation explained. On the other hand, Model 4 was found to fit reasonably well and was a significant improvement over both Models 2 and 3. This Model hypothesized a structure of four factors, which refer to all but the focus dimension of the evaluation of SLE (see Figure 3) since the two items concerned with the measurement of the focus dimension were found to belong to two other dimensions (i.e., one item is correlated with the factor representing the frequency dimension whereas the other is associated with the quality dimension). Moreover, one of the three items expected to measure the stage dimension was found to be correlated with both the stage and the quality dimensions.

Table 2
Goodness-of-Fit-Indices for Structural Equation Models Used to Test the Validity of the Proposed Framework for Measuring the Evaluation of the SLE

Structural equation models	X^2	*df*	CFI	RMSEA	X^2/df
1. Null model	2131.5	105	–	–	20.3
2. 1 first order factor	298.7	76	.878	.13	3.93
3. 5 correlated factors	142.1	67	.901	.09	2.12
4. 4 correlated factors (see Figure 2)	122.5	70	.947	.03	1.75
5. 1 second order general, 4 correlated factors	286.1	71	.921	.08	4.03

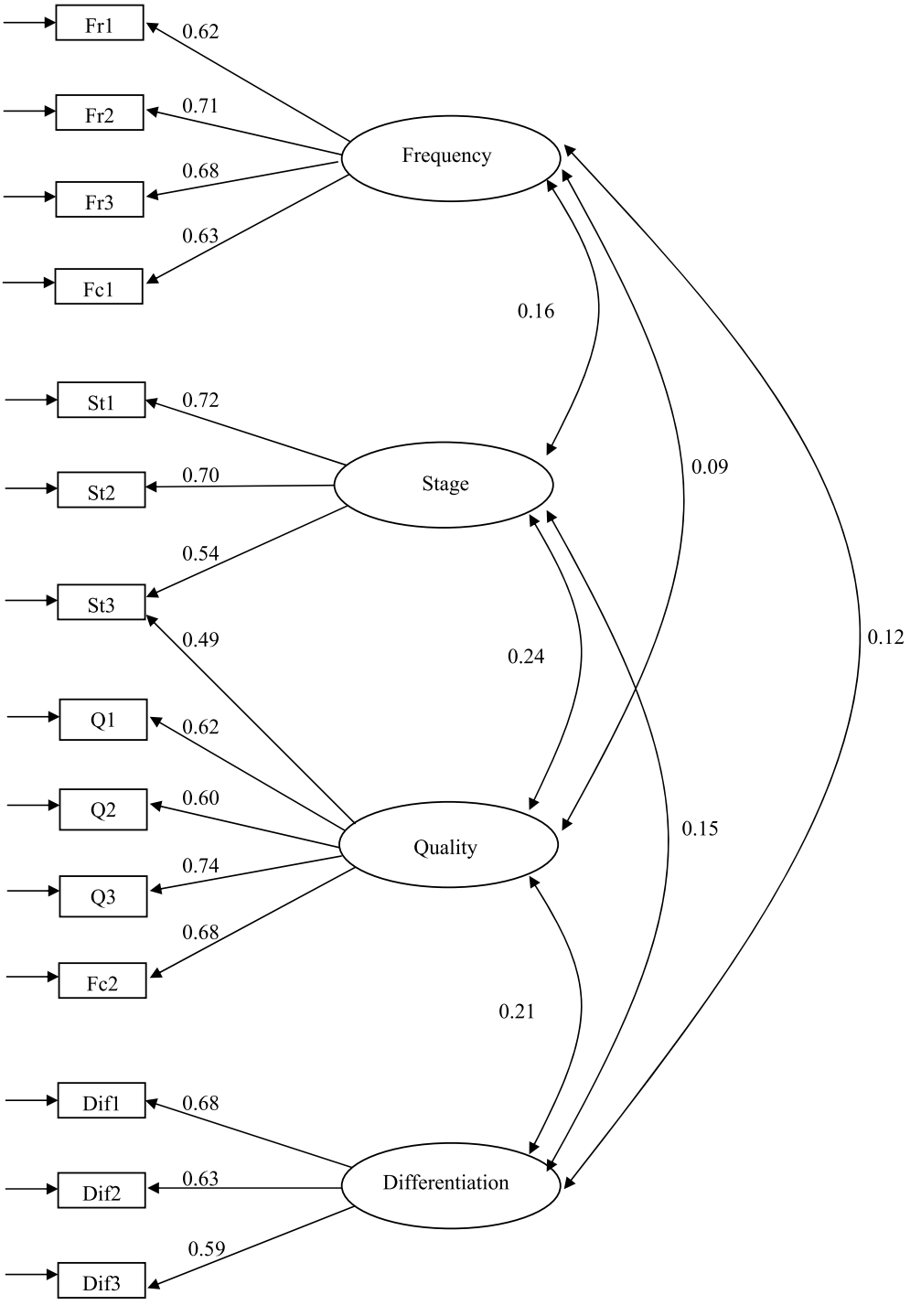

Figure 3. First-order four factors model of the questionnaire measuring the evaluation of the learning environment of the school with factor parameter estimates.

Third, Model 5 was examined to determine if a second-order structure would explain the lower-order trait factors, as these are described in Model 4, more parsimoniously. Specifically, Model 5 hypothesized that the scores which emerged from the 14 items could be explained by the four first-order factors (as these appear in Model 4) and one second-order factor (i.e., evaluation of SLE in general). In this study, for each subject the fit values of Model 5 do not meet the criteria for acceptable level of fit. We also tested three additional second-order models with varying factor structure but none of them was found to meet the criteria for acceptable level of fit. This finding provides support for arguing the importance of measuring each of the four dimensions of the evaluation of SLE factor separately rather than treating this school factor as unidimensional. Thus, for each school, four factor scores based on the results of Model 4 were estimated.

The Effect of School-Level Factors on Achievement in Four Outcomes of Schooling

Having established the construct validity of the framework used to measure the dimensions of the school factors of the dynamic model, it was decided to examine the extent to which the first-order factors, which were established through the SEM analyses, show the expected effects upon each of the four dependent variables, and thereby the analyses were performed separately for each variable. Specifically, the dynamic model was tested using MLwiN (Goldstein et al., 1998) because the observations are interdependent and because of multi-stage sampling since students are nested within classes, and classes within schools. The dependency has an important consequence. If students' achievement within a class or a school has a small range, institutional factors at class or school level may have contributed to it (Snijders & Bosker, 1999). Thus, the first step in the analysis was to determine the variance at individual, class, and school level without explanatory variables (empty model). In subsequent steps, explanatory variables at different levels were added. Explanatory variables, except from grouping variables, were entered as *Z*-scores with a mean of 0 and a standard deviation of 1. This is a way of centering around the grand mean (Bryk & Raudenbush, 1992) and yields effects that are comparable. Thus, each effect expresses how much the dependent variable increases (or decreases in case of a negative sign) by each additional deviation on the independent variable (Snijders & Bosker, 1999). Grouping variables were entered as dummies with one of the groups as baseline (e.g. boys = 0). The models presented in Tables 3 and 4 were estimated without the variables that did not have a statistically significant effect at .05 level.

A comparison of the empty models of the four outcome measures reveals that the effect of the school and classroom was more pronounced on achievement in mathematics and Greek language rather than in religious education. Moreover, the school and the teacher (classroom) effects were found to be higher on achievement of cognitive rather than affective aims of religious education. Furthermore, in each analysis the variance at each level reaches statistical significance ($p < .05$) and this implies that MLwiN can be used to identify the explanatory variables that are associated with achievement in each outcome of schooling (Goldstein, 2003).

In Model 1, the context variables at student, classroom and school levels were added to the empty model. The following observations arise from the figures of the four columns illustrating the results of Model 1 for each analysis. First, Model 1 explains approximately 50% of the total variance of student achievement in each outcome and most of the explained variance is at the student level. However, more than 30% of the total variance remained

Table 3
Parameter Estimates and (Standard Errors) for the Analyses of Greek Language and of Mathematics Achievement

Factors	Greek language							Mathematics						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Fixed part (Intercept)	−0.31(.08)	−0.22(.08)	−0.19(.08)	−0.20(.08)	−0.19(.08)	−0.22(.08)	−0.21(.08)	0.35 (.05)	0.28(.05)	0.23(.03)	0.24(.03)	0.26(.04)	0.20(.03)	0.24(.03)
Student level														
Prior knowledge		0.39(.05)	0.37(.05)	0.36(.05)	0.35(.05)	0.38(.05)	0.37(.05)		0.45(.10)	0.40(.10)	0.42(.11)	0.42(.10)	0.40(.09)	0.38(.09)
Sex (boys = 0, girls = 1)		0.19(.08)	0.18(.08)	0.20(.09)	0.22(.09)	0.19(.08)	0.20(.08)		−0.14(.06)	−0.13(.05)	−0.12(.05)	−0.13(.06)	−0.12(.05)	−0.13(.06)
SES		0.30(.06)	0.28(.05)	0.27(.05)	0.23(.05)	0.29(.05)	0.27(.05)		0.30(.12)	0.25(.09)	0.25(.09)	0.21(.08)	0.23(.09)	0.22(.10)
Classroom level														
Context														
Average prior knowledge		0.12(.05)	0.10(.04)	0.09(.04)	0.11(.05)	0.09(.04)	0.10(.04)		0.28(.10)	0.26(.09)	0.25(.10)	0.24(.10)	0.23(.09)	0.22(.09)
Average SES		0.08(.03)	0.07(.03)	0.08(.03)	0.08(.04)	0.07(.03)	0.06(.03)		0.12(.05)	0.13(.05)	0.10(.04)	0.09(.04)	0.11(.05)	0.10(.04)
Percentage of girls		NSS*	NSS	NSS	NSS	NSS	NSS		−0.05(.02)	−0.05(.02)	−0.04(.02)	−0.04(.02)	−0.05(.02)	−0.05(.02)
School level														
Context														
Average SES		NSS	NSS	NSS	NSS	NSS	NSS		NSS	NSS	NSS	NSS	NSS	NSS
Average prior knowledge		0.09(.04)	0.11(.05)	0.10(.05)	0.13(.06)	0.11(.05)	0.10(.05)		0.11(.05)	0.09(.04)	0.08(.04)	0.09(.04)	0.08(.05)	0.08(.04)
Percentage of girls		NSS	NSS	NSS	NSS	NSS	NSS		NSS	NSS	NSS	NSS	NSS	NSS
Frequency														
Policy for teaching			0.08(.03)							0.12(.03)				
Evaluation policy for teaching			0.10(.04)							0.09(.02)				
Student behavior outside the classroom			NSS							NSS				
Teacher collaboration			NSS							0.04(.01)				
Partnership policy			0.09(.03)							0.08(.03)				
Provision of resources			0.06(.02)							NSS				

Table 3
(*Continued*)

Factors	Greek language							Mathematics						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Evaluation of the SLE			0.03(.01)							NSS				
Stage														
Policy for teaching				0.03(.01)							0.04(.02)			
Evaluation policy for teaching				0.12(.02)							0.11(.03)			
Student behavior outside the classroom				NSS							NSS			
Teacher collaboration				NSS							0.06(.03)			
Partnership policy				0.11(.03)							0.10(.03)			
Provision of resources				0.05(.02)							NSS			
Evaluation of the SLE				0.13(.03)							0.12(.02)			
Focus														
Policy for teaching					NSS							0.04(.02)		
Evaluation policy for teaching					0.06(.02)							NSS		
Student behavior outside the classroom					NSS							NSS		
Teacher collaboration					NSS							0.06(.03)		
Partnership policy					0.05(.02)							NSS		
Provision of resources					NSS							NSS		
Quality														
Policy for teaching						0.07(.02)							0.06(.02)	
Evaluation policy for teaching						0.05(.02)							0.05(.02)	

Table 3
(*Continued*)

Factors	Greek language							Mathematics						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Student behavior outside the classroom						0.08(.02)							0.10(.02)	
Teacher Collaboration						NSS							0.05(.02)	
Partnership policy						0.10(.03)							NSS	
Provision of resources						NSS							0.06(.02)	
Evaluation of the SLE						0.09(.02)							0.08(.02)	
Differentiation														
Policy for teaching							0.11(.03)							0.12(.02)
Evaluation policy for teaching							NSS							NSS
Student behavior outside the classroom							0.07(.02)							NSS
Teacher Collaboration							0.06(.02)							NSS
Partnership policy							0.08(.02)							0.06(.02)
Provision of resources							NSS							0.05(.02)
Evaluation of SLE							NSS							0.08(.02)
Variance components (%)														
School	9.0	8.2	4.2	4.9	6.5	3.6	3.8	11.2	9.8	4.0	3.7	5.9	3.5	4.0
Class	14.7	10.3	9.8	9.2	10.2	9.4	9.7	14.8	10.0	9.3	9.4	9.9	8.5	9.2
Student	76.3	31.3	29.2	29.6	30.8	28.7	29.5	74.0	30.2	29.6	30.0	30.0	29.5	30.0
Explained		50.2	56.8	56.3	52.5	58.3	57.0		50.0	57.1	56.9	54.2	58.5	56.8

Table 3
(*Continued*)

Factors	Greek language							Mathematics						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Significance test														
X^2	815.6	507.2	279.3**	312.3	470.3	236.9	294.9	1144.9	795.5	650.7	656.3	774.2	504.1	619.8
Reduction		308.4	227.9	194.9	36.9	270.3	212.3		349.4	144.8	139.2	21.3	291.4	175.7
Degrees of freedom		6	5	5	2	5	4		7	4	5	2	6	4
p-value		.001	.001	.001	.001	.001	.001		.001	.001	.001	.001	.001	.001

Notes. *NSS = No statistically significant effect at .05 level; **for each alternative model 2 (i.e., Models 2a up to 2e) the reduction is estimated in relation to the deviance of Model 1. SES = socio-economic status.

Table 4
Parameter Estimates and (Standard Errors) for the Analyses of Achievement in Religious Education (Cognitive and Affective Outcomes)

Factors	Religious education (Cognitive aims)							Religious education (Affective aims)						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Fixed part (Intercept)	−0.59(.11)	−0.43(.09)	−0.41(.08)	−0.40(.08)	−0.43(.09)	−0.34(.08)	−0.40(.08)	0.41 (.08)	0.40 (.07)	0.30 (.07)	0.31 (.07)	0.40 (.07)	0.30 (.07)	0.34 (.07)
Student level														
Prior knowledge		0.41 (.05)	0.39 (.05)	0.38 (.05)	0.41 (.05)	0.42 (.05)	0.40 (.05)		0.36 (.10)	0.35 (.10)	0.34 (.10)	0.36 (.10)	0.35 (.10)	0.38 (.10)
Sex (boys = 0, girls = 1)		0.13 (.06)	0.12 (.05)	0.10 (.04)	0.13 (.06)	0.11 (.04)	0.10 (.05)		0.16 (.06)	0.15 (.06)	0.15 (.06)	0.16 (.06)	0.17 (.06)	0.15 (.06)
SES		0.12 (.05)	0.10 (.05)	0.09 (.04)	0.12 (.05)	0.10 (.05)	0.08 (.04)		NSS	NSS	NSS	NSS	NSS	NSS
Classroom level														
Context														
Average prior knowledge		0.15 (.06)	0.14 (.06)	0.13 (.06)	0.15 (.06)	0.12 (.05)	0.13 (.06)		0.19 (.08)	0.17 (.07)	0.16 (.07)	0.19 (.08)	0.18 (.07)	0.19 (.18)
Average SES		0.09 (.04)	0.08 (.04)	0.09 (.04)	0.09 (.04)	0.07 (.03)	0.06 (.03)		NSS	NSS	NSS	NSS	NSS	NSS
Percentage of girls		NSS	NSS	NSS	NSS	NSS	NSS		0.05 (.02)	0.04 (.02)	0.04 (.02)	0.05 (.02)	0.04 (.02)	0.03 (.01)
School level														
Context														
Average SES		NSS	NSS	NSS	NSS	NSS	NSS		NSS	NSS	NSS	NSS	NSS	NSS
Average prior knowledge		0.13 (.05)	0.13 (.05)	0.12 (.05)	0.13 (.05)	0.12 (.05)	0.13 (.05)		0.07 (.02)	0.06 (.02)	0.06 (.02)	0.07 (.02)	0.07 (.02)	0.06 (.02)
Percentage of girls		NSS	NSS	NSS	NSS	NSS	NSS		NSS	NSS	NSS	NSS	NSS	NSS
Frequency														
Policy for teaching			0.10 (.04)							NSS				
Evaluation policy for teaching			0.09 (.03)							0.11 (.02)				
Student behavior outside the classroom			NSS							NSS				
Collaboration/ interaction among teachers			NSS							0.07 (.02)				

Table 4
(*Continued*)

Factors	Religious education (Cognitive aims)							Religious education (Affective aims)						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Partnership policy			0.07 (.02)							0.06 (.03)				
Provision of resources			0.05 (.02)							NSS				
Evaluation of the SLE			0.05 (.02)							NSS				
Stage														
Policy for teaching				NSS							NSS			
Evaluation policy for teaching				0.11 (.03)							0.10 (.03)			
Student behavior outside the classroom				NSS							NSS			
Teacher collaboration				0.06 (.02)							NSS			
Partnership policy				0.09 (.02)							0.08 (.02)			
Provision of resources				NSS							NSS			
Evaluation of the SLE				0.10 (.02)							0.10 (.03)			
Focus														
Policy for teaching					NSS							NSS		
Evaluation policy for teaching					NSS							NSS		
Student behavior outside the classroom					NSS							NSS		
Teacher collaboration					NSS							NSS		
Partnership policy					0.05 (.02)							NSS		
Provision of resources					NSS							0.06 (.02)		

Table 4
(*Continued*)

Factors	Religious education (Cognitive aims)							Religious education (Affective aims)						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Quality														
Policy for teaching						0.07 (.02)							0.06 (.02)	
Evaluation policy for teaching						0.08 (.02)							0.09 (.02)	
Student behavior outside the classroom						0.07 (.02)							0.12 (.02)	
Teacher collaboration						0.06 (.02)							NSS	
Partnership policy						0.07 (.02)							NSS	
Provision of resources						NSS							NSS	
Evaluation of the SLE						0.09 (.02)							0.09 (.02)	
Differentiation														
Policy for teaching							0.10 (.02)							0.11 (.03)
Evaluation policy for teaching							NSS							NSS
Student behavior outside the classroom							0.05 (.02)							NSS
Teacher collaboration							0.06 (.02)							NSS
Partnership policy							NSS							0.07 (.03)
Provision of resources							0.07 (.02)							NSS
Evaluation of the SLE							NSS							0.07 (.02)

Table 4
(*Continued*)

Factors	Religious education (Cognitive aims)							Religious education (Affective aims)						
	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Variance components (%)														
School	8.5	7.2	4.9	5.0	6.5	4.5	4.6	7.0	6.9	4.7	4.6	6.0	4.2	4.6
Class	13.7	12.9	11.2	11.5	12.0	11.3	11.7	10.2	9.4	8.8	8.9	9.0	8.8	8.5
Student	77.8	31.2	30.3	29.5	31.0	29.2	30.3	82.7	32.7	31.9	32.3	32.6	31.6	32.0
Explained		48.7	53.6	54.0	50.5	55.0	53.4		51.0	54.6	54.2	52.4	55.4	54.9
Significance test														
X^2	985.6	676.7	485.8**	477.3	650.0	427.4	490.5	1024.3	684.9	488.9	495.7	660.4	451.4	471.4
Reduction		308.9	190.9	199.4	26.7	249.3	186.2		339.4	196.0	189.2	24.5	233.5	213.5
Degrees of freedom		6	5	4	1	6	4		5	3	3	1	4	3
p-value		.001	.001	.001	.001	.001	.001		.001	.001	.001		.001	.001

Notes. *NSS = No statistically significant effect at .05 level; **for each alternative model 2 (i.e., Models 2a up to 2e) the reduction is estimated in relation to the deviance of Model 1.

unexplained at the student level. Second, the likelihood statistic (X^2) shows a significant change between the empty model and model 1 ($p < .001$), which justifies the selection of Model 1. Second, the effects of all contextual factors at student level (i.e., SES, prior knowledge, sex) are significant, but the SES was not found to be associated with achievement of affective aims of religious education. Moreover, gender was not found to be consistently associated with student achievement in each outcome. Girls were found to have better results than boys in each outcome but mathematics. The results concerning gender differences in Greek language and mathematics are in line with findings of effectiveness studies conducted in Cyprus (Kyriakides et al., 2000; Kyriakides, 2005). Third, prior knowledge (i.e., aptitude) has the strongest effect in predicting student achievement at the end of grade 6. Moreover, aptitude is the only contextual variable that had a consistent effect on student achievement when it was aggregated either at the classroom or the school level.

At the next step of the analysis, for each dependent variable, five different versions of Model 2 were established. In each version of Model 2, the scores of the school-level factors that refer to the same measurement dimension and emerged through our attempt to test the construct validity of the teacher questionnaire were added to Model 1. The fitting of these five models was tested against Model 1. The likelihood statistic (X^2) reveals a significant change ($p < .001$) between Model 1 and each version of Model 2. This implies that variables measuring each of the five dimensions of the school effectiveness factors have significant effects on student achievement in all four outcomes of schooling taken into account by this study. This approach was deliberately chosen since the dimensions of the same factor are interrelated. Therefore, adding all dimensions into a single model causes difficulties of identifying which variables have effects on student achievement. Specifically, some variables may correlate with achievement when they are studied in isolation, but because of multicolinearity their effects may disappear when they are studied together. It was, therefore, considered appropriate to study the effect of each dimension of the school-level factors in isolation.

The following observations arise from the figures of Model 2a, which refer to the impact of the frequency dimension of the school factors on each of the four dependent variables. First, the only factor that did not have any statistically significant effect is concerned with student behavior outside the classroom. On the other hand, the evaluation of school policy for teaching and the schools' relations with parents were found to be associated with student achievement in each of the four dependent variables. Second, the figures of Models 2 reveal that the stage dimension of the two overarching factors concerned with school evaluation are associated with each outcome measure, whereas the stage dimension of only one factor (i.e., student behavior outside the classroom) does not have any statistically significant effect at the .05 level on student achievement. Moreover, the effects of the stage dimension of the two evaluation factors were found to be stronger than the effect of any other factor. Third, according to the results of Model 2c, the focus dimension of most school-level factors has no impact on achievement in any of the four dependent variables of this study. Moreover, the focus dimension of only one factor was found to be related with achievement in religious education. Fourth, the figures of Model 2d, which refer to the impact of the quality dimension of each effectiveness factor upon student achievement, reveal that there is no quality measure of a school-level factor that does not have any statistically significant effect upon at least one learning outcome measure. Moreover, for each outcome measure, Model 2d explains more variance than any other alternative Model 2, and this reveals the importance of using this dimension to measure the impact of effectiveness factors on

Table 5
Percentage of Explained Variance of Student Achievement in Each Outcome Provided by Each Alternative Model Testing the Effect of the Frequency Dimension of the School-Level Factors and the Effect of Combinations of Frequency Dimensions With Each of the Other Dimensions

Alternative models	Greek language (%)	Mathematics (%)	Cognitive religious education (%)	Affective religious education (%)
Model 2a (frequency dimension of school level factors)	56.8	57.1	53.6	54.6
Model 2f (frequency and stage dimensions)	58.5	58.2	56.7	57.2
Model 2g (frequency and focus dimensions)	57.4	57.7	54.2	55.2
Model 2h (frequency and quality dimensions)	59.3	59.9	57.1	58.6
Model 2i (frequency and differentiation dimensions)	58.7	58.8	56.8	57.4
Model 3 (all five dimensions of school level factors)	60.7	61.5	58.5	59.4

student achievement. Finally, the figures of the four Models of 2e reveal that the differentiation dimension of the overarching factor concerned with the school policy for teaching is not only consistently related with student achievement, but its effect size is stronger than the effect of the differentiation dimension of any other school-level factor. On the other hand, the differentiation dimension of the evaluation of school policy for teaching is not associated with student achievement in any outcome measure.

At the next stage of the analysis, we attempted to identify the amount of variance that can be explained when researchers take into account the effects of the frequency dimension of the school-level factors and the effects of at least another dimension. For this reason, four alternative models were created that took into account a combination of frequency dimension with another dimension of the school-level factors. Each model was compared with Model 2a, which takes into account only the frequency dimension. The likelihood statistics for each model justifies the inclusion of more than one dimension of factors in explaining variation of student achievement in each outcome of schooling. Table 5 illustrates the total explained variance of Model 2a and of five alternative models taking into account combinations of frequency with other dimensions of measurement. We can observe that for each outcome each alternative model explains more than the variance explained by considering only the frequency dimension. Moreover, the model with a combination of frequency with the quality dimension of the school-level factors explains more total variance than any other combination of the frequency with each of the three other dimensions. Finally, Model 3, combining all five dimensions, explains most of the variance. This model was found to fit better than any other alternative model. It is important to note that this model is able to explain more than 85% of the variance at the school-level of student achievement in each outcome. This implies that all five dimensions should be taken into account in order to explain as much variance as possible at the school level. However, none of these models explains more than about 62% of the total variance. Nevertheless, this can be attributed to the fact that no classroom, and only three student factors, were taken into account.

Discussion

In the first part of this section, implications of findings for the development of the dynamic model are drawn. First, this study reveals that the dynamic model is a theoretical

model that can be put into testing. This argument is based on the fact that a generalizability study revealed that almost all items (i.e., 102 out of 110) were generalizable at the level of school and the scales for each factor and its dimension had relatively high internal reliability. The only exception was concerned with the measurement of only one aspect of the SLE factor. Specifically, it was not possible to get any reliable data from the teacher questionnaire on the establishment of positive attitudes towards learning. This might be attributed either to the quality of these six questionnaire items and/or to the difficulties of measuring this aspect of policy by looking at its perceived impact on teachers who are not likely to have similar ideas on what positive attitudes towards learning are. This difficulty might even be attributed to the lack of any national policy at primary schools in Cyprus in respect to this aspect of the SLE (see Kyriakides, 1999).

Second, the results of SEM analyses provided support to the proposed framework that was used to measure the functioning of school factors. Specifically, for each factor, the model that fits the data better is the one that refers to the existence of the five dimensions proposed in the dynamic model. This implies that factor scores for each dimension of each factor should be calculated and their impact on achievement on both cognitive and affective outcomes can be investigated. A comparison of each proposed model with alternative models implying that all five dimensions should be treated as a single factor provided further support to the argument that each school factor should be measured in relation to the five dimensions of the dynamic model. The only exception was concerned with the focus dimension of the factor concerned with the evaluation of SLE.

Third, it was possible not only to demonstrate the construct validity of the measurement framework but also to reveal the added value of using five dimensions to measure the school-level factors for explaining variation of student achievement in different outcomes. Specifically, it has been shown that the five alternative models used to examine the impact of each of the five measurement dimensions fit the data better than Model 1, which was concerned with the impact of contextual factors on student achievement (see Tables 3 and 4). This implies that all five dimensions can be used to identify factors associated with student achievement in both cognitive and affective aspects of education. Moreover, taking into account the combination of frequency dimension with other dimensions of school-level factors increases the explained variance on student achievement. In this particular study, the frequency dimension explains less than 65% of the school-level variance (i.e., Model 2a explains 7.4% out of 11.2% of the total variance in mathematics achievement that is situated at the school level) whereas using all five dimensions increases the explained variance to at least 85% of the school variance in each learning outcome.

Fourth, this study reveals that there are factors that were found to have no statistically significant effect on student achievement by measuring the impact of their frequency dimension but that did have a significant impact on student achievement when other dimensions were taken into account. This implies that previous studies concerned only with the frequency dimension might draw wrong conclusions about the impact of a factor and might fail to explain as much variance as possible at the school level. For example, in this study, the frequency dimension of school policy on student behavior outside the classroom was not associated with student achievement in any outcome, but the quality dimension of this factor had an impact on achievement in each subject and a relatively high impact on affective outcomes. Therefore, the findings of this study reveal that emphasis should be given to other dimensions of effectiveness factors and not only to frequency, which has been used predominantly in all effectiveness studies in the past (Teddlie & Reynolds, 2000).

Fifth, looking at the impact that each proposed factor has on student achievement we can claim that the importance of the four overarching school-level factors is confirmed. At least one dimension of each of these factors was found to be associated with student achievement in different outcomes of schooling that refer to both the core subjects of the curriculum and the subsidiary subjects, and to both cognitive and affective aims. Therefore, these four overarching factors could be considered as generic in nature (Campbell, Kyriakides, Muijs, & Robinson, 2004).

Sixth, the three aspects of the overarching factor concerned with policy on teaching were found to be strongly related to each other. When this is confirmed in other studies, a question that can arise is why these three aspects of teaching should be treated separately when we study the teacher-level factors but not when we look at the school-level factors. A possible explanation is that school policy has to be developed in a more comprehensive way in order to have an effect on student outcomes. Establishing an effective policy for teaching seems to imply that all three aspects of teaching are integrated into the school guidelines and the support activities that are offered to teachers. Therefore, we should consider the possibility of establishing a more parsimonious model, which may refer to a single factor on school policy on teaching but should be defined in a way that investigates whether and how all three aspects of policy for teaching are covered.

Seventh, the effect sizes of the school factors reported in this study were relatively small compared to the results of two meta-analyses of the effect of school factors on student achievement that have been carried out recently (Kyriakides et al., in press; Scheerens et al., 2005). However, we were able to explain a large proportion of school-level variance (at least 85%), indicating that the theoretical model stands and the way we measure the school-level factors is appropriate. Moreover, the large number of variables used to measure the five dimensions of these factors can be seen as a possible reason for finding relatively smaller effect sizes than those reported in other studies. Obviously, replication studies should be conducted, especially since the small effect size of the school-level factors reported here could also be attributed to the context of the educational system of Cyprus, which is highly centralized and does not allow schools to be autonomous on issues concerned with teaching and learning environment (Kyriakides, 1999).

Finally, some more specific implications of this study for policy and practice at school level can be drawn. For example, this study reveals that establishing policy and taking actions to improve the relations between schools and parents has a significant impact on student outcomes (both cognitive and affective). More specifically, the study points to the fact that schools should not simply develop a policy at a certain specific stage but constant attention on this aspect of SLE should be given, especially since the stage dimension of partnership policy (implying that different activities should be carried out across the school year) was found to have a stronger effect than any other dimension of this factor. Another example refers to the importance of the quality dimension of the evaluation factors. The findings of this study reveal that it is not sufficient for schools to develop evaluation mechanisms but that they should focus on the formative purposes of evaluation by providing feedback that can be used for improving their own practice both for teaching and the SLE.

Based on the findings of this study, it is also possible to provide feedback to each individual school on the functioning of each school factor. The feedback can be specific by referring to the items that were used to measure each factor and its dimension. For example, in order to measure school policy on teaching, we refer to three specific aspects of this policy (i.e., quantity of teaching, provision of learning opportunities, and quality of teaching). For

each aspect, items include specific activities that schools should consider in designing their policy for teaching. A typical example is the items that measure school policy on management of teaching time, which investigate the extent to which specific activities are carried out in order to maximize teaching time (e.g. lessons start on time and finish on time; there are no interruptions of lessons for staff meetings and/or for preparation of school festivals and other events). Thus, suggestions for designing school policy on teaching may emerge by looking at the items that are used to measure the factors and their dimensions in relation to the way each school is functioning. This implies that further research testing the dynamic model may not only address the theoretical and empirical implications formulated above, but may also explore further the practical suggestions for policy and practice that arise from the model and the studies testing the model in different contexts.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Andrich, D. (1988). A general form of Rasch's Extended Logistic Model for partial credit scoring. *Applied Measurement in Education, 1*(4), 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.* California: Multivariate Software Inc.

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

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: SAGE.

Byrne, B. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Campbell, R.J., Kyriakides, L., Muijs, R.D., & Robinson, W. (2004). *Assessing teacher effectiveness: A differentiated model.* London: Routledge/Falmer.

Cohen, D., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge/Falmer.

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

Creemers, B.P.M., & Kyriakides, L. (2006). A critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 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: Routledge.

de Jong, R., Westerhof, K.J., & Kruiter, J.H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement, 15*(1), 3–31.

den Brok, P., Brekelmans, M., Levy, J., & Wubbels, T. (2002). Diagnosing and improving the quality of teachers' interpersonal behavior. *International Journal of Educational Management, 4,* 176–184.

Driessen, G., & Sleegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement, 11*(1), 57–79.

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

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., et al. (1998). *A user's guide to MLWIN.* London: Institute of Education.

Grosin, L. (1993). School effectiveness research as a point of departure for school evaluation. *Scandinavian Journal of Educational Research, 37*(4), 317–330.

Gustafsson, J.E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioural Research, 28*(4), 407–434.

Harris, A. (2001). Building the capacity for school improvement. *School Leadership and Management, 21*(3), 261–270.

Hopkins, D. (2001). *School improvement for real.* London: RoutledgeFalmer.

Hopkins, D., Reynolds, D., & Gray, J. (1999). Moving on and moving up: Confronting the complexities of improvement. *Educational Research and Evaluation, 5*(1), 22–40.

Kline, R.H. (1998). *Principles and practice of structural equation modeling.* London: Guilford Press.

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 school effectiveness and school improvement* (pp. 107–124). Lisse: Swets and Zeitlinger.

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement, 16*(2), 103–152.

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*(4), 429–446.

Kyriakides, L., Campbell, R.J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of Creemers' comprehensive model of educational effectiveness. *School Effectiveness and School Improvement, 11*(4), 501–529.

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*(2), 183–205.

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

Kyriakides, L., & Tsangaridou, N. (2008). Towards the development of generic and differentiated models of educational effectiveness: A study on school and teacher effectiveness in physical education. *British Educational Research Journal, 34*(6), 807–838.

Leithwood, K., & Jantzi, D. (2006). Transformational school leadership for large-scale reform: Effects on students, teachers, and their classroom practices. *School Effectiveness and School Improvement, 17,* 201–227.

Levine, D.U., & Lezotte, L.W. (1990). *Unusually effective schools: A review and analysis of research and practice.* Madison: National Center for Effective Schools Research and Development.

Linnakyla, P., Malin, A., & Taube, K. (2004). Factors behind low reading literacy achievement. *Scandinavian Journal of Educational Research, 48*(3), 231–249.

Marsh, H.W., & Roche, L.A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias and utility. *American Psychologist, 52*(11), 1187–1197.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Ministry of Education. (1994). *The new curriculum.* Nicosia, Cyprus: Ministry of Education.

Muthen, L.K., & Muthen, B.O. (2001). *MPlus user's guide.* Los Angeles: Muthen & Muthen.

Reezigt, G.J., Guldemond, H., & Creemers, B.P.M. (1999). Empirical validity for a comprehensive model on educational effectiveness. *School Effectiveness and School Improvement, 10*(2), 193–216.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness.* Oxford: Pergamon.

Scheerens, J., Seidel, T., Witziers, B., Hendriks, M., & Doornekamp, G. (2005). *Positioning and validating the supervision framework.* Enschede/Kiel: University of Twente, Department of Educational Organisational and Management.

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

Stringfield, S. (1994). A model of elementary school effects. In D. Reynolds, B.P.M. Creemers, P.S. Nesselrodt, E.C. Schaffer, S. Stringfield, & C. Teddlie (Eds.), *Advances in school effectiveness research and practice* (pp. 153–187). Oxford: Pergamon Press.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer Press.

Torres, R.T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation, 22*(3), 387–395.

Witziers, B., Bosker J.R., & Kruger L.M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly, 39,* 398–425.