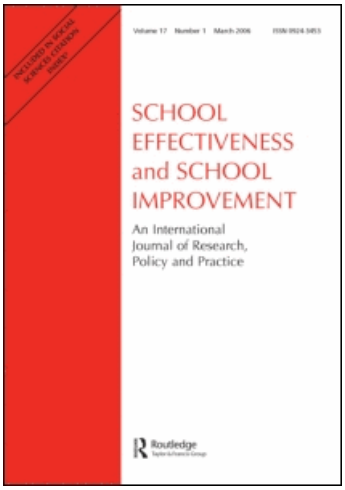

School Effectiveness and School Improvement
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t714592801

Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of the dynamic model

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

To cite this Article Kyriakides, Leonidas and Creemers, Bert 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
To link to this Article: DOI: 10.1080/09243450802047873
URL: http://dx.doi.org/10.1080/09243450802047873

School Effectiveness and School Improvement
Vol. 19, No. 2, June 2008, 183–205

Routledge
Taylor & Francis Group

Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of the dynamic model

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

[a]*Department of Education, University of Cyprus, Nicosia, Cyprus;* [b]*Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands*

(*Received 3 April 2007; final version received 24 October 2007*)

The dynamic model does not only refer to different effectiveness factors and groupings of factors operating at different levels but also supports that each factor can be defined and measured using 5 dimensions: frequency, focus, stage, quality, and differentiation. The importance of taking each dimension into account is raised in this paper. Moreover, empirical support to the model and the use of this measurement framework is provided. Specifically, the paper refers to the methods and results of a study conducted in Cyprus which investigates the validity of the model at the classroom level by measuring teacher effectiveness in mathematics, language, and religious education. It is shown that the proposed measurement framework can be used to describe each classroom-level factor. The added value of using these 5 dimensions of the classroom-level factors to explain variation on student achievement is also identified. Finally, implications for the development of the dynamic model are drawn.

Keywords: modelling educational effectiveness; teacher effectiveness; measuring teacher behaviour; dimensions of teacher factors

Introduction

One of the most important criticisms of educational effectiveness research (EER) is that there is a shortage of rational models from which researchers can build theory. The problem is aggravated by infrequent use of whatever models exist (Bosker & Scheerens, 1994). However, in the 1990s, researchers in the area of effectiveness attempted to integrate the findings of school effectiveness research, teacher effectiveness research, and the early input-output studies. The resulting theoretical models of educational effectiveness (e.g., Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992) have a multilevel structure, where schools are nested in contexts, classrooms are nested in schools, and students are nested in classrooms or teachers. Nevertheless, none of these models explicitly refers to the measurement of each effectiveness factor. On the contrary, it is often assumed that these factors represent unidimensional constructs. Considering effectiveness factors as multidimensional constructs provides a better picture of what makes teachers and schools effective and may help us develop specific strategies for improving educational practice. In this context, a dynamic model of educational

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

ISSN 0924-3453 print/ISSN 1744-5124 online

DOI: 10.1080/09243450802047873
http://www.informaworld.com

effectiveness has been developed (Creemers & Kyriakides, 2008) which takes into account the major criticisms of the current models of educational effectiveness and illustrates the dimensions upon which the measurement of each effectiveness factor should be based. Although the model refers to different effectiveness factors and groupings of factors operating at different levels, it is assumed that each factor can be defined and measured using similar dimensions. This is a way to consider each factor as a multidimensional construct and at the same time to be in line with the parsimonious nature of the model. However, many theories in the area of social sciences die, not because of any demonstrated lack of merit, but because even their creators failed to provide any evidence at all supporting even some of the ideas included in their theory. Thus, this paper illustrates the results of the first phase of a study conducted in Cyprus in order to test the validity of the dynamic model at the classroom level. Since one of the main differences of the dynamic model from the current models of educational effectiveness has to do with the use of a multidimensional approach in measuring the effectiveness factors, emphasis is given to the identification of the importance of using different dimensions to measure classroom-level factors. Moreover, the proposed measurement framework is described in the second section of this paper. In the next two sections, the methods used to test the validity of the dynamic model and the main results of the study are illustrated. Finally, implications of findings for the development of the model are drawn.

Dimensions measuring effectiveness factors

In most effectiveness studies, no clear distinction is made between the different aspects of an effectiveness factor which were found to be associated with student achievement. Unless researchers explain how they attempted to measure each factor and point out which aspects of the functioning of each factor were found to be related with student achievement, we cannot conduct quantitative syntheses of effectiveness studies in a systematic way which will help us generate and/or test theoretical models of effectiveness. For example, one study (Reezigt, Guldemond, & Creemers, 1999) testing the validity of the comprehensive model of educational effectiveness (Creemers, 1994) was looking at the frequency dimension of school evaluation policy to identify the effect of this factor on achievement and revealed both negative and positive effects, whereas another study (Kyriakides, 2005) was looking at the emphasis given to the formative aspect of evaluation and revealed positive effects. Unless the dimensions used to measure this factor are taken into account, these results can be seen as contradicting each other, whereas the second study revealed the importance of treating quality as a measurement dimension of evaluation. Thus, one of the essential characteristics of the dynamic model is concerned with its attempt to define effectiveness factors by using the following five measurement dimensions: frequency, focus, stage, quality, and differentiation. Frequency is a quantitative way to measure the functioning of each effectiveness factor, whereas the other four dimensions examine qualitative characteristics of the functioning of each effectiveness factor at the system/school/classroom level. Using this measurement framework implies that each factor should not only be examined by measuring how frequently the factor is present in the system/school/class (i.e., through a quantitative perspective) but also by investigating specific aspects of the way the factor is functioning (i.e., looking at qualitative characteristics of the functioning of the factor). The importance of taking each dimension into account is discussed below.

Frequency

The frequency dimension refers to how often an activity associated with an effectiveness factor is present in a system, a school, or a classroom. This is probably the easiest way to measure the effect of a factor on student achievement, and, consequently, most effectiveness studies used this dimension to define effectiveness factors. Aiming to explain how this dimension is used in order to measure the functioning of effectiveness factors, two examples are given below. First, the dynamic model of educational effectiveness is based on the assumption that student achievement is maximised when teachers not only actively present materials but structure it by: (a) beginning with overviews and/or review of objectives, (b) outlining the content to be covered and signalling transitions between lesson parts, (c) calling attention to main ideas, and (d) reviewing main ideas at the end (Rosenshine & Stevens, 1986). Summary reviews are also important since they integrate and reinforce the learning of major points (Brophy & Good, 1986). It can be claimed that these structuring elements not only facilitate memorising of the information but allow for its apprehension as an integrated whole with recognition of the relationships between parts. Moreover, achievement is higher when information is presented with a degree of redundancy, particularly in the form of repeating and reviewing general views and key concepts. Therefore, the frequency dimension of structuring is measured by taking into account the number of structuring tasks that take place in a typical lesson, as well as how long each structuring task takes place. These two indicators help us identify the importance that the teacher attached to this factor. Second, evaluation is seen as an integral part of teaching (Stenmark, 1992) and the dynamic model treats teacher evaluation as an important effectiveness factor at the classroom level. In order to measure the frequency dimension of this factor, the number of evaluative tasks and the time when they take place is taken into account.

Creemers and Kyriakides (2008) claim that researchers should examine whether there is a linear or a nonlinear relation between the frequency dimension of each effectiveness factor and student achievement. For example, it is expected that there is a curvilinear relation between the frequency of teacher evaluation and student outcomes, since an overemphasis to evaluation might reduce the actual time spent on teaching and learning, whereas teachers who do not collect any information are not able to adopt their teaching to student needs (Creemers & Kyriakides, 2006).

Focus

The effectiveness factors are measured by taking into account the focus of the activities which reveal the function of the factors at the classroom, school, and system level. Two aspects of focus for each factor can be measured. First, it is taken into account that each task associated with the functioning of an effectiveness factor may not take place by chance but for some reasons. For example, it is very likely that when teachers and/or the headteacher of a primary school attempt to establish their school policy on quality of teaching, they expect to achieve through this activity some specific purpose(s) (e.g., improve the quality of teaching at their school). This implies that researchers measuring qualitative characteristics of the functioning of a factor should try to identify the purposes that are expected to be achieved through an activity. Thus, according to the dynamic model, the first aspect of the focus dimension of each factor should address *the purpose(s)* for which an activity takes place. It is taken into account that an activity may be expected to achieve single or multiple purposes. For example, in the case of establishing a policy on

parental involvement, the activities might be restricted to a single purpose (e.g., parents visit schools to get information about student progress) or might address more than one purpose (e.g., parents visit the school to exchange information about children's progress and to assist teachers in and outside the classroom). The importance of measuring this aspect of focus dimension can be attributed to research findings which reveal that if all the activities are expected to achieve a single purpose, then the chances of achieving the purpose are high, but the effect of the factor might be small due to the fact that other purposes are not achieved and/or synergy may not exist since the activities are isolated (Schoenfeld, 1998) On the other hand, if all the activities are expected to achieve multiple purposes, there is a danger that specific purposes are not addressed in such a way that they can be implemented successfully (Pellegrino, 2004).

The second aspect of the focus dimension refers to the *specificity* of the activities, which can range from specific to general. For example, in the case of school policy on parental involvement, the policy could either be more specific in terms of concrete activities that are expected to take place (e.g., the school policy may refer to specific hours that parents can visit the school) or more general (e.g., it informs parents that they are welcome to the school but without giving them specific information about what, how, and when). The dynamic model is based on the assumption that the measurement of the focus of an activity, either in terms of its specificity or in terms of the number of purposes that it is expected to achieve, may be related in a nonlinear way with student achievement. For example, guidelines on parental involvement which are very general may not be helpful at all in establishing good relations between parents and teachers, which, when good, can result in supporting student learning. On the other hand, a school policy which is very specific in defining activities may restrict teachers and parents from being productively involved and creating their own ways for implementing the school policy. The above example also reveals the importance of investigating whether for some effectiveness factors an interaction between these two aspects of their focus dimension may exist. An issue that can be raised is whether the focus dimension can be measured based on the activities observed or needs further interpretation. For the purposes of the study reported here, we are mainly concerned with observable tasks and try to find out whether a single or multiple purposes were addressed.

Stage

The activities associated with a factor can be measured by taking into account the stage at which they take place. It is assumed 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. This assumption is partly based on the fact that evaluations of programmes aiming to improve educational practice reveal that the extent to which these intervention programmes have any impact on educational practice is partly based on the length of time that the programmes are implemented in a school (Gray et al., 1999). Moreover, the importance of using the stage dimension to measure each effectiveness factor arises from the fact that it has been shown that the impact of a factor on student achievement partly depends on the extent to which activities associated with this factor are provided throughout the school career of the student (e.g., Creemers, 1994; Slater & Teddlie, 1992). For example, school policy on quantity of teaching, which refers to policy on cancellation of lessons and absenteeism, is expected to be implemented throughout the year and not only through specific regulations announced at a specific point of time (e.g., at the beginning of the school year). It is also expected that the continuity will be achieved when

the school is flexible in redefining its own policy and adapting the activities related to the factor by taking into account the results of its own self-evaluation mechanism (Creemers & Kyriakides, 2008). Although measuring the stage dimension gives information about the continuity of the existence of a factor, activities associated with the factor may not necessarily be the same. Therefore, using the stage dimension to measure the functioning of a factor can help us identify the extent to which there is constancy at each level and flexibility in using the factor during the period that the investigation takes place.

Quality

The quality dimension refers to the properties of the specific factor itself, as these are discussed in the literature. The importance of using this dimension arises from the fact that looking at the quantity element of a factor ignores the fact that the functioning of the factor may vary. Moreover, the literature has shown that only using certain activities associated with a factor has positive effects on student outcomes. For example, the classroom factor concerned with teacher evaluation can be measured by looking at the properties of the evaluation instruments used by the teacher, such as the validity, the reliability, the practicality, and the extent to which the instruments cover the teaching content in a representative way. This dimension is also measured by investigating the type of feedback that a teacher gives to the students and the way students use the teacher feedback. Specifically, research has shown that effective teachers provide constructive feedback, which has positive implications for teaching and learning (Black & Wiliam, 1998; Harlen & James, 1997; Kyriakides, 2002; Muijs & Reynolds, 2001). This implies that, in our attempt to measure the teacher evaluation factor, we should not only look at the frequency dimension of this factor but also at the extent to which teachers attempt to achieve the formative rather than the summative purpose.

Differentiation

Although the dynamic model is expected to be a generic model, it takes into account the findings of research into differential educational effectiveness (Campbell, Kyriakides, Muijs, & Robinson, 2003). Specifically, effectiveness factors are seen as generic in nature, but it is acknowledged that their impact on different groups of students/teachers/schools may vary. As a consequence, differentiation is treated as a measurement dimension and is concerned with 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 the specific needs of each subject or group of subjects will increase the successful implementation of a factor and ultimately maximise its effect on student learning outcomes. Although differentiation could be considered a property of an effectiveness factor, it was decided to treat differentiation as a separate dimension of measuring each effectiveness factor rather than incorporate it into the quality dimension. In this way, the importance of taking into account the special needs of each subject or group of subjects is recognised. It is finally important to note that the dynamic model is based on the assumption that it is difficult to deny that persons of all ages learn, think, and process information differently.

One way to differentiate instruction is for teachers to teach according to individual student learning needs as these are defined by their background and personal characteristics such as gender, socioeconomic status (SES), ability, thinking style, and personality type (Kyriakides, 2007). For example, effective teachers provide more active

instruction and feedback, more redundancy, and smaller steps with a higher success rate to their low-SES or low-achieving students (Brophy, 1986). On the other hand, they are aware of the fact that high-SES or high-achieving students thrive in an atmosphere that is academically stimulating and somewhat demanding, and they create such a learning environment for them. Warmth and support, in addition to good instruction, is provided to low-achieving students, who are more frequently encouraged for their efforts (Muijs, Campbell, Kyriakides, & Robinson, 2005). A similar argument can be made in relation to the way teachers should be treated by their school leaders. For example, instructional leadership should not be seen as equally important for all the teachers of a school. Effective principals are expected to adapt their leadership to the specific needs of the teachers by taking into account the extent to which they are ready to implement a task (Hersey & Blanchard, 1993). Similarly, policy-makers are expected to adapt their general policy to the specific needs of groups of schools and encourage teachers to differentiate their instruction. Research into differential educational effectiveness reveals that teachers' objectives, as well as organisational and cultural factors, should be taken into account when the dimension of differentiation is measured (Dowson & McInerney, 2003; Hayes & Deyhle, 2001).

However, the differentiation dimension does not imply that the subjects are not expected to achieve the same purposes. On the contrary, adapting the policy to the special needs of each group of schools/teachers/students may ensure that all of them will become able to achieve the same purposes. This argument is partly supported by research into adaptive teaching and the evaluation projects of innovations concerned with the use of adaptive teaching in classrooms (e.g., Houtveen, Van de Grift, & Creemers, 2004; Noble, 2004; Reusser, 2000). Therefore, policy-makers should make explicit to teachers what they are expected to achieve through differentiating their instruction and through responding to the different needs of their students. This is particularly crucial for establishing an effective policy on equal opportunities since research has shown that some existing educational practices are maladaptive (e.g., Kyriakides, 2004; Peterson, Wilkinson, & Hallinan, 1984). Therefore, the differentiation dimension helps policy-makers not only establish a policy on equal opportunities but also provide support to the schools where teaching practice is maladaptive and help them act in such a way that differentiation of instruction does not result in holding lower achievers back and increasing individual differences (Kyriakides, 2007).

Research aims

The importance of taking each dimension into account is raised above, but it should also be acknowledged that studies investigating the validity of the proposed measurement framework of effectiveness factors are needed. Thus, this paper refers to the results of a study investigating the validity of the proposed measurement framework. One of the main differences of the dynamic model from all the existing theoretical models is concerned with its attempt to show that effectiveness factors are multidimensional constructs and can be measured in relation to specific dimensions. Therefore, it is considered important to identify whether the proposed factors are multidimensional constructs and the five dimensions can be used to measure each one. It is also important to identify the added value of using these five dimensions of the effectiveness factors to explain variation on student achievement. Not only the construct validity of the measurement framework should be demonstrated but also its significance and relevance to the field of EER should be investigated. Thus, two are the major aims of this study. First, it is examined whether

each dimension of the classroom-level factors of the dynamic model is associated with student achievement. Second, since the dynamic model is considered as a generic model of educational effectiveness, the effects of effectiveness factors upon different outcomes of schooling (both cognitive and affective) are examined.

The study reported here is concerned with the effects of the five dimensions of the classroom effectiveness factors. The choice to test the validity of the dynamic model at classroom level first, rather than at any of the upper levels, is based on the fact that studies on EER show that this level is more significant than the school and the system level (e.g., Kyriakides, Campbell, & Gagatsis, 2000; Scheerens & Bosker, 1997; Yair, 1997). In addition, defining factors at the classroom level is seen as a prerequisite for defining the school and the system level (Creemers, 1994).

In this paper, not much emphasis on describing the classroom level of the dynamic model is given. However, it is pointed out that, based on the main findings of EER (e.g., Brophy & Good, 1986; Darling-Hammond, 2000; Doyle, 1990; Kyriakides, Campbell, & Christofidou, 2002; Muijs & Reynolds, 2000; Rosenshine & Stevens, 1986; Scheerens & Bosker, 1997; Wang, Haertel, & Walberg, 1993), the dynamic model refers to eight effectiveness factors which describe teachers' instructional role: orientation, structuring, questioning, teaching modelling, applications, management of time, teacher role in making classroom a learning environment, and classroom assessment (see Creemers & Kyriakides, 2006). These eight factors do not refer only to one approach of teaching such as the direct teaching model or the constructivist approach. An integrated approach in defining quality of teaching is adopted. Therefore, the dynamic model does not refer only to skills associated with direct teaching and mastery learning such as structuring and questioning but also to orientation and teaching modelling, which are in line with new theories of teaching. In recent years, constructivists and others who support the "new learning" approach (e.g., Choi & Hannafin, 1995; Collins, Brown, & Newman, 1989; Savery & Duffy, 1995; Simons, Van der Linden, & Duffy, 2000; Vermunt & Verschaffel, 2000) have developed a set of instructional techniques that are supposed to enhance the learning disposition of students, such as modelling, coaching, scaffolding and fading, articulating, reflection, exploration, generalisation, collaboration, provision of anchors, goal orientation, and self-regulated learning. Creemers and Kyriakides (2006) explain in detail how the five dimensions of the model can be used to measure the classroom-level factors. It is also shown that the eight factors of the dynamic model cover at least partly the main approaches to learning and teaching. For example, the collaboration technique is included under the overarching factor contribution of teacher to the establishment of the classroom-learning environment. Moreover, most of these approaches are subsumed in the factors teaching modelling and orientation. Therefore, it is important to identify the extent to which each dimension of these eight classroom-level factors is associated with student achievement in different outcomes of schooling.

Methods

Participants

Stratified sampling (Cohen, Manion, & Morrison, 2000) was used to select 52 Greek Cypriot primary schools, but only 50 schools participated in the study. All the Year 5 students ($n = 2503$) 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 ($t = 1.21$, $df = 107$, $p = 0.22$). Although this study refers 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 Greek Cypriot students of Year 5. Therefore, it was not possible to examine whether the sample was nationally representative in terms of any other characteristic than students' sex and the size of class. However, it can be claimed that a nationally representative sample of Greek Cypriot Year 5 students in terms of these two characteristics was drawn.

Dependent variables: student achievement in mathematics, Greek language, and religious education

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 (Ministry of Education, 1994). Student achievement in relation to the affective aims included in the Cyprus curriculum for religious education was also measured. The written tests are available upon request from the first author. But since effectiveness in religious education (RE) is rarely measured, some information about the RE test is given in the endnote.[1] The three written tests in mathematics, Greek language, and RE were administered to all Year 5 students of the school sample at the beginning and at the end of the school year 2004–2005. 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 analyse the emerging data in each subject separately, and four scales, which refer to student knowledge in mathematics, Greek language, and religious education, and also to student attitudes towards religious education, were created and analysed for reliability, fit to the model, meaning, and validity. Analysis of the data revealed that each scale had satisfactory psychometric properties (see Creemers & Kyriakides, 2008). Thus, for each student, four different scores for his/her achievement at the beginning of the school year were generated by calculating the relevant Rasch person estimate in each scale. The same approach was used to estimate student achievement at the end of the school year in relation to these four outcomes of schooling.

Since one of the issues that can be raised in measuring achievement in affective aims of RE is the fact that it is unclear whether student responses reveal attitudes or knowledge about religion, we searched for correlations between achievement in the two Rasch scales (i.e., achievement of cognitive and achievement of affective aims in RE). Although we found out statistically significant correlations between the student estimates in the two Rasch scales of religious education which emerged both at the beginning ($r = 0.29$, $n = 2503$, $p < .001$) and at the end ($r = 0.27$, $n = 2503$, $p < .001$) of the school year 2004–2005, the relatively small values of these two correlation coefficients reveal that the two scales which emerged from both measurement periods refer to two different constructs (Cronbach, 1990).

Explanatory variables at the student level

Aptitude

Aptitude refers to the degree in which a student is able to perform the next learning task. 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 at the beginning of the school year (i.e., baseline assessment).

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 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 economical 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 (34%), occupations held by middle class (36%), and occupations held by upper-middle class (30%). Relevant information for each child was taken from the school records. Then standardised values of the above five variables were calculated, resulting in the SES indicator.

Explanatory variables at the classroom level: quality of teaching

The eight factors 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. The observation instruments and the guidelines for the observers are published on a disk and are available for research purposes. It is shown below that the two low-inference observation instruments generate data for all eight factors and their dimensions. Specifically, one of the low-inference observation instruments is based on Flanders' system of interaction analysis (Flanders, 1970). However, we developed a classification system of teacher behaviour which is based on the way each factor of the dynamic model is measured. For example, in order to measure the quality dimension of teacher behaviour in dealing with disorder, which is an element of the classroom as a learning environment factor, the observers are asked to identify any of the following types of teacher behaviour in the classroom: (a) the teacher is not using any strategy at all to deal with a classroom disorder problem; (b) the teacher is using a strategy, but the problem is only temporarily solved; (c) the teacher is using a strategy that has a long-lasting effect. The distinction between temporarily (i.e., category b) and long-lasting effect (i.e., category c) is based on observations on what is happening during the lesson and after the action of the teacher. Similarly, in order to measure the focus dimension of the way the teacher deals with the negative aspects of competition, the following two types of teacher behaviour were given specific codes: (a) the teacher is dealing only with the specific problem that arises and which is associated with the negative effects of competition and (b) the teacher puts the problem in a more general perspective in order to help students see the positive aspects of competition and avoid the negative ones.

Moreover, we developed a classification system of student behaviour, and the observer is not only expected to classify student behaviour when it appears but also to identify the students who are involved in each type of behaviour. Thus, the use of this instrument enables us to generate data about teacher–student and student–student interaction. For example, the focus dimension of teacher–student interactions is measured by classifying each observed teacher–student interaction according to the purpose(s) that was expected to serve (i.e., managerial reasons, social encounter, learning). Moreover, the quality

dimension of this factor is measured by investigating the immediate impact that each teacher initiative has on establishing relevant interactions and especially whether the teacher was able to establish on task behaviour through the interactions she/he promoted. The measurement of the impact of teacher activity is based on observations of students' reactions and not on interpretation of the quality of teacher activity. As far as the measurement of the stage is concerned, the instrument generated data enable us to take into account at which phase of the lesson each interaction took place.

The second low-inference observation instrument refers to five factors of the model (i.e., orientation, structuring, teaching modelling, questioning techniques, and application). This instrument was designed in a way that enables us to collect more information in relation to the quality dimension of these five factors. For each factor, quality and focus dimension is defined in a specific way. For example, in regard to the measurement of the quality of an application task, observers have to indicate whether the teacher is: (a) asking students to practise in using a specific process/algorithm to solve a number of similar exercises or (b) expecting students to activate certain cognitive processes in order to find the solution of more complex tasks and/or algorithms. The following two examples illustrate the difference between the two types of application task. First, after discovering the formula that gives the area of rectangles, students are given the dimensions (width and length) of 10 rectangles and are asked to find their area. Second, students are asked to find how much money they will need to paint the ceiling of their classroom if paint comes into buckets of 3 liters and each of them costs $10. As far as the measurement of the focus dimension of structuring is concerned, three types of activities are discerned. First, a structuring task may refer to the day lesson activities only without establishing any links with other lessons. Second, the teacher might relate the day lesson activities with the previous lessons. Finally, the teacher might not only show the relation of the day lesson with the previous lessons but may also explain how the lesson is related to lessons in the future.

In regard to the other three dimensions, similar measurement ways are used irrespective of whether an activity belongs to one factor or another. Specifically, observers are asked to give an ordinal number to each observed activity. For example, if at the beginning of the lesson the teacher asks students to practise on the content of the lesson that was taught the day before and then he/she comments on the structure of the lesson of the day, the first observed task is an application one and the second a structuring task. By giving ordinal numbers to the activities, we could establish a score for measuring the stage dimension of each factor. Moreover, the observers are asked to report the time (in minutes) that was used for each activity. Therefore, the quantity dimension of each factor was measured by identifying not only how many activities associated with a factor were observed but also by calculating the total time that was used for all the activities associated with this factor. In regard to the measurement of the differentiation dimension, observers are asked to indicate whether there is any type of differentiation in the observed task. For example, in the case of an orientation activity, a teacher may clarify further the aims of the lesson to a certain group of students (e.g., "the less able ones"). Similarly, in the case of an application task, the teacher may assign to the less able students more application exercises or give them more time to solve them.

The high-inference observation instrument covers the five dimensions of all eight factors of the model, and observers are expected to complete a Likert scale to indicate how often each teacher-behaviour was observed. For example, an item concerned with the frequency dimension of orientation is asking observers to indicate how much time the teacher spent to explain the objectives of the lesson. In order to measure the quality

dimension of this factor, one of the items of the high-inference observation instrument is asking observers to indicate the extent to which the orientation activities that were organised during the lesson helped students understand the new content. Similarly, the quality dimension of the application factor is measured through items asking the observers to identify the extent to which the observed tasks were nothing else but replication of the activities that were organised during the presentation of the new content or whether the application tasks were used by the teacher as starting points for teaching new concepts.

Observations were carried out by six members of the research team who attended a series of seminars on how to use the three observation instruments. During the school year, the external observers visited each class nine 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 0.81.

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. For example, an item concerned with the stage dimension of the structuring factor was asking students to indicate whether at the beginning of the lesson the teacher explains how the new lesson is related to previous ones, whereas another item was asking whether at the end of each lesson they spend some time in reviewing the main ideas of the lesson. Similarly, the following item was used to measure the differentiation dimension of the application factor: "the teacher of Mathematics assigns to some pupils different exercises than to the rest of the pupils". A Generalisability Study (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson, Webb, & Rowley, 1989) on the use of students' ratings was conducted. It was found that the data which emerged from almost all the questionnaire items could be used for measuring the quality of teaching of each teacher in each subject separately (see Creemers & Kyriakides, 2008). However, three items of the questionnaire concerned with assessment in religious education and one item concerned with the differentiation dimension of learning strategies in both Greek language and religious education had to be removed. Thus, the score for each teacher in each of the questionnaire items found to be generalisable was the mean score emerged from the responses of the students of his/her class.

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 presented here. Specifically, for each subject, the first-order factor model which was found to be the most appropriate for describing each classroom-level factor is shown in Table 1. Moreover, Table 1 illustrates the second-order factors which were found to fit reasonably well with MTMM data in relation to some classroom-level factors. This table reveals that this study provides support for the construct validity of the five measurement dimensions of most effectiveness factors. The few exceptions which were identified reveal the difficulty of defining the quality dimension. For example, in the case of questioning, aspects of quality were found to belong to two separate factors, whereas in the case of teaching modelling, the differentiation and the quality dimensions were found to belong to the same factor. Moreover, the results of this study seem to reveal that the classroom as a learning environment cannot be treated as a single factor but as two interrelated factors in the learning environment concerning relations among students and relations between a

Table 1. Goodness-of-fit indices for the best fitting structural equation models used to test the validity of the proposed framework for measuring each classroom-level effectiveness factor in each subject.

SEM Models	Greek Language					Mathematics					Religious Education				
	X^2	*df*	CFI	RMSEA	X^2/df	X^2	*df*	CFI	RMSEA	X^2/df	X^2	*df*	CFI	RMSEA	X^2/df
Structuring															
1) 5 correlated traits, 3 correlated methods	248.0	137	.947	.03	1.81	253.4	137	.942	.03	1.85	261.7	137	.935	.04	1.91
2) 2 correlated second-order general, 3 correlated methods	346.1	139	.936	.05	2.49	404.5	139	.930	.06	2.91	304.4	139	.939	.05	2.19
Orientation															
1) 5 correlated traits, 3 correlated methods	253.4	137	.941	.03	1.85	246.6	137	.940	.04	1.80	260.3	137	.935	.04	1.90
2) 2 correlated second-order general, 3 correlated methods	318.3	139	.938	.05	2.29	390.6	139	.930	.06	2.81	297.5	139	.939	.05	2.14
Questioning															
1) 6 correlated traits, 4 correlated methods	553.7	301	.947	.03	1.84	562.9	301	.946	.03	1.87	574.9	301	.943	.04	1.91
2) 2 correlated second-order general, 4 correlated methods	580.2	307	.942	.04	1.89	610.9	307	.940	.05	1.99	684.6	307	.935	.06	2.23
Application															
5 correlated traits, 3 correlated methods	231.5	137	.965	.02	1.69	226.1	137	.969	.02	1.65	261.7	137	.938	.04	1.91
Teaching Modelling															
4 correlated traits, 3 correlated methods	251.0	141	.953	.03	1.78	245.3	141	.952	.03	1.74	262.3	141	.942	.04	1.86
Management of Time															
4 correlated traits, 3 correlated methods	126.1	66	.942	.05	1.91	122.1	66	.948	.03	1.85	112.2	66	.953	.03	1.70
Teacher evaluation															
5 correlated traits, 2 correlated methods	28.1	14	.936	.05	2.01	30.9	14	.930	.06	2.21	28.7	14	.945	.05	2.05
Classroom as a learning environment															
2 correlated second-order, 5 correlated methods	770.9	352	.930	.06	2.19	700.5	352	.932	.05	1.99	707.5	352	.935	.04	2.01

teacher and his/her students. Furthermore, 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) standardised trait loadings, in comparison to the relatively lower (i.e., lower than .40) standardised method loadings (see Creemers & Kyriakides, 2008). These findings support the use of multimethod techniques for increasing measurement validity, construct validity, and thus, stronger support for the validity of subsequent results.

Results

Having established the construct validity of the framework used to measure the dimensions of the eight effectiveness factors of the dynamic model, it was decided to examine the extent to which the first-order factors which were established through the Structural Equation Modelling (SEM) analyses (see Table 1) show the expected effects upon each of the four dependent variables. The analyses were performed separately for each variable. Specifically, the dynamic model was tested using "MLwiN" (Rasbash, Steele, Browne, & Prosser, 2005). The first step in the analysis was to determine the variance at the individual, class, and school level without explanatory variables (empty model). In subsequent steps, explanatory variables at different levels were added. Explanatory variables, except grouping variables, were centred as Z scores with a mean of 0 and a standard deviation of 1. This is a way of centring 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 2 and 3 were estimated without the variables that did not have a statistically significant effect at 0.05 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 teacher (classroom) effect was found to be higher on achievement of cognitive rather than affective aims of religious education. In Model 1, the context variables at the 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 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. Third, the effects of all contextual factors at the student level (i.e., SES, prior knowledge, sex) are significant, but SES was not found to be associated with achievement of affective aims in religious education. Moreover, gender was not found to be consistently associated with student achievement in each outcome. Girls were found to have better results in relation to every outcome except from mathematics. Finally, prior knowledge (i.e., aptitude) has the strongest effect in predicting student achievement at the end of the school year. Moreover, aptitude is the only contextual variable that had a consistent effect on student achievement when aggregated either at the classroom or at 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 factor scores of SEM models

Table 2. Parameter estimates and (standard errors) for the analyses of Greek language and of mathematics achievement.

	Greek Language							Mathematics						
Factors	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)	−0.39 (.08)	−0.33 (.08)	−0.29 (.08)	−0.30 (.08)	−0.31 (.08)	−0.27 (.08)	−0.30 (.08)	0.36 (.05)	0.30 (.05)	0.13 (.02)	0.11 (.02)	0.10 (.02)	0.10 (.02)	0.10 (.02)
Student Level														
Prior knowledge		0.49 (.05)	0.47 (.05)	0.46 (.05)	0.45 (.05)	0.48 (.05)	0.48 (.05)		0.71 (.12)	0.70 (.12)	0.70 (.12)	0.69 (.12)	0.70 (.12)	0.68 (.12)
Sex (boys = 0, girls = 1)		0.23 (.10)	0.21 (.09)	0.20 (.09)	0.22 (.09)	0.21 (.09)	0.19 (.09)		−0.18 (.07)	−0.15 (.07)	−0.14 (.07)	−0.16 (.07)	−0.14 (.07)	−0.13 (.07)
SES		0.32 (.06)	0.28 (.05)	0.29 (.05)	0.25 (.05)	0.26 (.05)	0.27 (.05)		0.60 (.25)	0.55 (.24)	0.50 (.24)	0.51 (.24)	0.53 (.24)	0.52 (.24)
Classroom Level														
Context														
Average prior knowledge		0.15 (.05)	0.11 (.04)	0.12 (.05)	0.13 (.04)	0.12 (.04)	0.10 (.04)		0.31 (.11)	0.28 (.10)	0.26 (.11)	0.25 (.11)	0.23 (.11)	0.24 (.11)
Average SES		0.09 (.04)	0.08 (.04)	0.09 (.04)	0.08 (.04)	0.07 (.03)	0.06 (.03)		0.15 (.04)	0.13 (.04)	0.10 (.04)	0.09 (.04)	0.09 (.04)	0.10 (.04)
Percentage of girls		N.S.S.*	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.		−0.05 (.02)	−0.05 (.02)	−0.05 (.02)	−0.04 (.02)	−0.05 (.02)	−0.05 (.02)
Quality of teaching														
Frequency Structuring			0.09 (.02)							0.07 (.02)				
Frequency Orientation			0.07 (.02)							0.11 (.02)				
Frequency Questioning			0.08 (.02)							0.08 (.02)				
(Frequency Questioning)^2			−0.02 (.01)							N.S.S.				
Frequency Application			0.06 (.03)							0.08 (.03)				
Freq/cy Teaching Modelling			N.S.S.							N.S.S.				
Frequency Assessment			0.06 (.02)							N.S.S.				
(Frequency Assessment)^2			−0.02 (.01)							N.S.S.				
Freq/cy management time			0.10 (.03)							0.10 (.03)				
Fr. Teacher–student relation			N.S.S.							0.03 (.01)				
Frequency student relations			0.10 (.02)							0.10 (.02)				
Stage Structuring				0.04 (.02)							0.04 (.02)			
Stage Orientation				0.02 (.01)							0.02 (.01)			
Stage Questioning				N.S.S.							N.S.S.			
Stage Application				0.07 (.02)							0.07 (.02)			
Stage Teaching Modelling				N.S.S.							N.S.S.			
Stage Management of time				N.S.S.							N.S.S.			
Stage Assessment				N.S.S.							N.S.S.			
St. teacher–student relations				N.S.S.							0.03 (.01)			
Stage student relations				0.03 (.01)							N.S.S.			
Focus Structuring					N.S.S.							0.03 (.01)		
Focus Orientation					0.07 (.02)							0.05 (.02)		
Focus Questioning					N.S.S.							N.S.S.		
Focus Application					N.S.S.							0.04 (.02)		
Focus Teaching Modelling					0.09 (.03)							0.11 (.03)		
Focus Assessment					0.03 (.01)							0.03 (.01)		
Foc teacher–student relation					N.S.S.							N.S.S.		
Focus student relations					0.03 (.01)							N.S.S.		
Quality Structuring						0.05 (.02)							0.05 (.02)	

(*continued*)

Table 2. (*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
Quality Orientation						N.S.S.							0.02 (.01)	
Quality of questions						0.04 (.02)							0.04 (.02)	
Quality of feedback						0.06 (.02)							0.05 (.02)	
Quality Application						N.S.S.							N.S.S.	
Quality Assessment						0.06 (.03)							0.06 (.02)	
Quality time management						N.S.S.							N.S.S.	
Q. teacher–student relation						0.03 (.01)							0.03 (.01)	
Quality student relations						0.04 (.02)							N.S.S.	
Differentiation Structuring							N.S.S.							N.S.S.
Differentiation Orientation							N.S.S.							N.S.S.
Differentiation Questioning							0.04 (.02)							0.05 (.02)
Differentiation Application							0.06 (.02)							0.08 (.02)
Differentiation Assessment							N.S.S.							N.S.S.
Dif/tion time management							N.S.S.							N.S.S.
Dif teacher–student relation							N.S.S.							0.04 (.02)
Dif/tion student relations							0.03 (.01)							0.03 (.01)
Quality of teaching modeling including differentiation							0.03 (.01)							0.05 (.02)
School Level: Context														
Average SES		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.
Average prior knowledge		0.13 (.05)	0.12 (.05)	0.11 (.05)	0.10 (.05)	0.12 (.05)	0.11 (.05)		0.11 (.05)	0.08 (.04)	0.07 (.04)	0.08 (.04)	0.06 (.03)	0.08 (.04)
Percentage of girls		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.
Variance components														
School	9.5%	7.7%	7.6%	7.6%	7.7%	7.5%	7.6%	11.5%	8.1%	7.5%	7.5%	7.9%	7.5%	7.4%
Class	15.2%	11.1%	8.8%	9.2%	9.3%	8.7%	8.9%	15.4%	9.3%	7.3%	6.9%	6.7%	6.0%	6.2%
Student	75.3%	31.5%	28.3%	28.6%	28.5%	28.2%	28.5%	73.1%	30.9%	29.7%	30.0%	30.3%	29.5%	30.0%
Explained		49.7%	55.3%	54.6%	54.5%	55.6%	55.0%		51.7%	55.5%	55.6%	55.1%	57.0%	56.4%
Significance test														
X^2	1015.6	686.7	428.8**	558.8	579.2	497.4	581.5	1224.3	984.9	795.9	885.7	883.6	821.4	861.4
Reduction		328.9	257.9	127.9	107.5	189.3	105.2		239.4	189.0	99.2	101.3	163.5	123.5
Degrees of freedom		6	9	4	4	6	4		7	7	4	5	6	5
p value		.001	.001	.001	.001	.001	.001		.001	.001	.001	.001	.001	.001

*N.S.S. = No statistically significant effect **For each alternative Model 2 (i.e., Models 2a up to 2e) the reduction is estimated in relation to the deviance of Model 1.

Table 3. Parameter estimates and (standard errors) for the analyses of achievement in religious education (cognitive and affective outcomes).

	Religious Education (Cognitive aims)							Religious Education (Affective aims)						
Factors	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)	−0.79 (.11)	−0.63 (.09)	−0.61 (.08)	−0.60 (.08)	−0.62 (.08)	−0.64 (.08)	−0.60 (.08)	0.61 (.08)	0.50 (.07)	0.43 (.07)	0.41 (.07)	0.42 (.07)	0.40 (.07)	0.44 (.07)
Student Level														
Prior knowledge/attitude		0.51 (.05)	0.49 (.05)	0.48 (.05)	0.50 (.05)	0.46 (.05)	0.50 (.05)		0.41 (.10)	0.40 (.10)	0.40 (.10)	0.39 (.10)	0.40 (.10)	0.38 (.10)
Sex (boys = 0, girls = 1)		0.23 (.09)	0.19 (.09)	0.20 (.09)	0.21 (.09)	0.21 (.09)	0.20 (.09)		0.18 (.07)	0.15 (.07)	0.15 (.07)	0.16 (.07)	0.14 (.07)	0.15 (.07)
SES		0.12 (.05)	0.10 (.05)	0.09 (.04)	0.11 (.05)	0.10 (.05)	0.08 (.04)		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.
Classroom Level														
Context														
Average prior knowledge		0.25 (.07)	0.21 (.07)	0.22 (.07)	0.23 (.07)	0.22 (.07)	0.20 (.07)		0.21 (.08)	0.18 (.07)	0.16 (.07)	0.15 (.07)	0.19 (.07)	0.20 (.18)
Average SES		0.09 (.04)	0.08 (.04)	0.09 (.04)	0.08 (.04)	0.07 (.03)	0.06 (.03)		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.
Percentage of girls		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.		0.05 (.02)	0.04 (.02)	0.04 (.02)	0.04 (.02)	0.04 (.02)	0.03 (.01)
Quality of teaching														
Frequency Structuring			0.11 (.02)							0.09 (.02)				
Frequency Orientation			N.S.S.							N.S.S.				
Frequency Questioning			0.10 (.02)							0.09 (.02)				
Frequency Application			0.06 (.03)							N.S.S.				
Freq/cy Teaching Modelling			N.S.S.							N.S.S.				
Frequency Assessment			N.S.S.							N.S.S.				
Freq/cy management time			0.10 (.03)							0.10 (.02)				
Fr. teacher–student relation			N.S.S.							0.03 (.01)				
Frequency student relations			0.05 (.02)							0.07 (.02)				
Stage Structuring				0.05 (.02)							0.06 (.03)			
Stage Orientation				0.02 (.01)							N.S.S.			
Stage Questioning				N.S.S.							N.S.S.			
Stage Application				0.08 (.02)							0.07 (.02)			
Stage Teaching Modelling				N.S.S.							N.S.S.			
Stage Management of time				N.S.S.							N.S.S.			
Stage Assessment				N.S.S.							N.S.S.			
St. teacher–student relation				0.05 (.02)							0.03 (.01)			
Stage student relations				0.03 (.01)							N.S.S.			
Focus Structuring					0.07 (.02)							0.03 (.01)		
Focus Orientation					N.S.S.							0.05 (.02)		
Focus Questioning					0.03 (.01)							0.04 (.02)		
Focus Application					0.05 (.02)							0.11 (.03)		
Focus Teaching Modeling					N.S.S.							N.S.S.		
Focus Assessment					N.S.S.							N.S.S.		
Foc teacher–student relation					N.S.S.							0.03 (.01)		
Focus student relations					0.04 (.02)							N.S.S.		
Quality Structuring						0.05 (.02)							0.05 (.02)	
Quality Orientation						N.S.S.							0.02 (.01)	
Quality of questions						0.09 (.02)							0.04 (.02)	

(*continued*)

Table 3. (*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 of feedback						0.08 (.02)							0.05 (.02)	
Quality Application						0.04 (.02)							N.S.S.	
Quality Assessment						N.S.S.							0.06 (.02)	
Quality time management						N.S.S.							N.S.S.	
Quality teacher–student relation						0.03 (.01)							0.03 (.01)	
Quality student relations						0.04 (.02)							N.S.S.	
Differentiation Structuring							N.S.S.							N.S.S.
Differentiation Orientation							N.S.S.							N.S.S.
Differentiation Questioning							0.07 (.02)							0.05 (.02)
Differentiation Application							0.08 (.02)							0.08 (.02)
Differentiation Assessment							N.S.S.							N.S.S.
Dif/tion time management							N.S.S.							N.S.S.
Dif teacher–student relation							0.02 (.01)							0.04 (.02)
Dif/tion student relations							0.03 (.01)							0.03 (.01)
Quality of teaching modeling including differentiation							0.03 (.01)							0.05 (.02)
School Level: Context														
Average SES		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.
Average prior knowledge		0.13 (.05)	0.12 (.05)	0.11 (.05)	0.10 (.05)	0.12 (.05)	0.11 (.05)		0.08 (.02)	0.06 (.02)	0.05 (.02)	0.07 (.02)	0.06 (.02)	0.07 (.02)
Percentage of girls		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.		N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.	N.S.S.
Variance components														
School	8.0%	7.7%	7.6%	7.6%	7.7%	7.5%	7.6%	7.5%	7.0%	6.7%	6.6%	6.7%	6.4%	6.5%
Class	13.2%	11.1%	8.8%	8.2%	8.3%	7.7%	8.9%	10.4%	9.3%	6.3%	7.2%	6.7%	6.4%	6.5%
Student	78.8%	34.5%	30.3%	29.6%	30.5%	29.2%	29.5%	82.1%	32.6%	31.7%	31.4%	31.3%	31.1%	31.0%
Explained		46.7%	53.3%	54.6%	53.5%	55.6%	54.0%		51.1%	55.3%	54.8%	55.3%	56.1%	56.0%
Significance test														
X^2	1823.6	1457.1	1337.6*	1309.2*	1359.6*	1277.6*	1331.9*	1024.5	835.1	705.2	758.9	719.8	700.6	711.6
Reduction		366.5	119.5	147.9	97.5	179.5	125.2		189.4	129.9	76.2	115.3	134.5	123.5
Degrees of freedom		6	5	5	4	6	5		5	5	3	5	6	5
p value		.001	.001	.001	.001	.001	.001		.001	.001	.001	.001	.001	.001

*N.S.S. = No statistically significant effect **For each alternative Model 2 (i.e., Models 2a up to 2e) the reduction is estimated in relation to the deviance of the Model 1.

which refer to the same dimension of measuring the classroom-level effectiveness factors of the dynamic model were added to Model 1. Thus, the fitting of these five models was tested against Model 1, and the likelihood statistic (X^2) shows a significant change between Model 1 and each version of Model 2 ($p < .001$). This implies that variables measuring the five dimensions of the classroom effectiveness factors have significant effects on student achievement in the four outcomes of schooling.

The following observations arise from the figures of Model 2a, which refer to the impact of the frequency dimension of the effectiveness factors on each of the four dependent variables. First, teaching modelling is the only factor which did not have any statistically significant effect on student achievement. On the other hand, the structuring and the management of time were found to be associated with student achievement in each of the four dependent variables. Second, although curvilinear relations were assumed to exist between most of the frequency factors and student achievement (see Creemers & Kyriakides, 2006), only two such relations were identified, and both of them refer to student achievement in Greek language. Specifically, a curvilinear relation between achievement in Greek language and the frequency dimension of the factor concerned with asking questions was identified. Moreover, the frequency dimension of teacher assessment is related in a nonlinear way to language achievement.

As far as the figures of the models which refer to the impact of the stage dimension of the classroom-level factors are concerned, the stage dimension of two factors (i.e., structuring and application) is associated with each outcome measure, whereas the stage dimension of three other factors (i.e., questioning, assessment, and management of time) is not associated with student achievement. Moreover, the effect of the stage dimension of the application factor was found to be the strongest. The figures of Model 2c reveal that the focus dimension of at least four factors is associated with student achievement in each dependent variable. However, there is no factor which has a statistically significant effect across the four outcomes or any factor that is not associated with at least one dependent variable. The figures of Model 2d refer to the impact of the quality dimension of each effectiveness factor upon student achievement. We can observe that, 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 student achievement. Finally, for each outcome, the figures of Model 2e reveal that the differentiation dimension of three factors (i.e., questioning, application, and classroom learning environment) is consistently related to student achievement, whereas the differentiation dimension of all the other factors is not associated with student achievement on any outcome measure.

At the next stage of the analysis, we attempted to identify the amount of variance which can be explained when researchers take into account the effects of the frequency dimensions of the classroom-level factors and the effects of at least one other dimension. For this reason, four alternative models were created which took into account a combination of frequency dimension with another dimension of the eight factors. Each model was compared with Model 2a which takes into account only the frequency dimension. The likelihood statistics for each model justify the inclusion of more than one dimension of factors in the model. Table 4 illustrates the total explained variance of Model 2a and of the 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 only by considering the frequency dimension. Moreover, the model with a combination of frequency and quality dimensions of the classroom-level factors explains more total variance than any other

Table 4. Percentage of explained variance of student achievement for each student outcome provided by each alternative model testing the effect of the frequency dimension of the classroom-level factors and the effect of combinations of frequency dimensions with each of the other dimensions.

Alternative Models	Greek Language	Mathematics	Cognitive Rel. Educ.	Affective Rel. Educ.
Model 2a (frequency dimension of classroom-level factors)	55.3%	55.5%	53.3%	51.1%
Model 2f (frequency and stage dimensions)	59.2%	57.8%	56.7%	54.7%
Model 2g (frequency and focus dimensions)	58.7%	56.8%	55.9%	53.9%
Model 2h (frequency and quality dimensions)	59.7%	59.1%	57.1%	55.2%
Model 2i (frequency and differentiation dimensions)	58.9%	58.1%	56.2%	54.9%
Model 3 (all five dimensions of classroom-level factors)	60.9%	60.1%	59.0%	57.3%

combination of the frequency with each of the other three dimensions. Finally, the model combining all five dimensions explains most of the variance and was found to fit better than any other alternative model. It is important to note that this model is able to explain more than 70% of the variance at the classroom 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 classroom level. However, none of these models explains more than about 60% of the total variance. Nevertheless, this can be attributed to the fact that we did not take into account the impact of any school-level factor. Moreover, only some factors at the student level were taken into account.

Implications of findings for the development of the dynamic model

In the last section of this paper, implications of findings for the development of the dynamic model are drawn. First, 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 the model empirically. For example, it can be claimed that the model is not parsimonious since it contains more factors and more dimensions than previous models, and it is therefore not possible to illustrate priorities for educational improvement. Moreover, the inclusion of different dimensions for measuring each factor complicates the data collection and the analysis. However, this study seems to reveal that the dynamic model is a theoretical model that can be put into testing. Specifically, the results of the study provide support for the construct validity of the five measurement dimensions of most effectiveness factors at the classroom level. It has been shown that the classroom-level effectiveness factors cannot be considered as unidimensional constructs since neither the one first-order factor model nor the one second-order factor model was considered appropriate for measuring any of the eight effectiveness factors.

Second, this study reveals the added value of using five dimensions to measure the classroom-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. 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 classroom-level factors increases the explained variance on student achievement. Furthermore, there are factors which were found to have no statistically significant effect on student achievement by measuring the impact of their frequency dimension but which 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 classroom level. For example, in this study, the frequency dimension of teaching modelling was not associated with student achievement on any outcome, but the quality dimension of this factor had an impact on achievement. 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 studies in the past.

When these findings are supported and expanded upon by other studies, directions can be given to link effectiveness research with educational practice and especially improvement efforts by indicating ways of improving education other than just increasing the presence of effective factors in the classroom. More specifically, teachers may realise that more emphasis should be given not only to the quantitative presence of each factor but also to some qualitative characteristics of the tasks associated with each factor. For example, the importance of the stage dimension reveals that it is not enough to use a significant number of structuring and/or orientation tasks in each lesson but also to distribute these tasks at different phases of a lesson and/or series of lessons. In addition, when further studies provide support to the importance of the differentiation dimension, teachers and other stakeholders may be encouraged to adopt their teaching practice to the needs of different groups of students. Therefore, further research testing the generalisability of the findings of this study may not only provide support to the validity of the dynamic model but may also help teachers and other stakeholders develop specific strategies to improve their teaching practice.

Notes

1. In regard to the measurement of student achievement in the affective aims of RE, we, as evaluators, took into account the fact that the major affective aim of the Cyprus curriculum is to establish positive attitudes towards the liturgical life of the Greek Christian Orthodox Church (Ministry of Education, 1994). For this reason, an open-ended question of the RE test was asking students to indicate how frequently and especially why they go to church. Based on student responses, it was possible to classify them into four groups which enable us to measure their attitudes towards the liturgical life using an ordinal scale. Specifically, students of the lowest level pointed out that they do not participate in the liturgical life of the church any more, and they do not do that because they cannot see any meaning for participation in church activities. Students of the second level pointed out that they participate because someone else is asking them to do it (e.g., ''I go to Church to do a favour to my grandfather, whom I like a lot'') or even for other reasons which do not reveal that they consider it important to participate in the liturgical life of the church (e.g., ''I go to Church to meet my friends and play football when the liturgy is over''). Students at the third level considered participation at the liturgical life as a duty (e.g., ''I go to Church because every good Christian has to do so''). Finally, the responses of students of the highest level revealed that they participate in the liturgical life because they like it and/or because it makes them feel nice (e.g., ''When I go to Church I pray to God and take the Holy Communion and this makes me feel nice'').

   Since teaching RE is expected to contribute to the moral development of students, another task was formulated in order to measure student attitudes towards diversity within religions.

Students were asked to express their agreement or disagreement with the fact that in some icons of Christians in African countries Jesus is painted as a black person. Based on student responses, we could classify them into three levels of an ordinal scale. Responses such as "This is bad because the black colour is the colour of evil" revealed racist attitudes and were representatives of the lowest level of the scale. Students of the middle level were those who did not consider these icons acceptable because Jesus was born in Jerusalem and was not a black person. This implies that students of the second level used criteria associated with the formal knowledge as contained in the curriculum to judge this activity. Finally, students of the highest level were in favour of this activity and attributed this behaviour to the fact that when people want to express their love to someone they use means familiar to their own life style to do that.

On the other hand, questions measuring achievement in the cognitive aims of RE were concerned with the aims of the specific units that teachers of Year 5 are expected to teach in Greek Cypriot primary schools (Ministry of Education, 1994). For example, a four-alternative multiple-choice question was asking students to find out who was the last prophet that was born before Jesus. Short-answer (completion and fill in the blank) questions were also used. One of them was asking students to name the best friend of Jesus, whereas another question was asking them to explain what the first Christians used to do when they met each other.

Notes on contributors

Dr Leonidas Kyriakides is Assistant 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. He is a member of the editorial board and the book review editor of the *School Effectiveness and School Improvement* journal. In 2008, he published together with Bert Creemers a book entitled *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*.

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 Effectivenes 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.

References

Andrich, D. (1988). A general form of Rasch's Extended Logistic Model for partial credit scoring. *Applied Measurement in Education*, *1*(4), 363–378.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: King's College London School of Education.

Bosker, R.J., & Scheerens, J. (1994). Alternative models of school effectiveness put to test. *International Journal of Educational Research*, *21*(2), 159–180.

Brophy, J. (1986). Teacher influence on student achievement. *American Psychologist*, *41*(10), 1069–1077.

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: Macmillan.

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

Campbell, R.J., Kyriakides, L., Muijs, R.D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education*, *29*(3), 347–362.

Choi, J.I., & Hannafin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational Technology Research and Development*, *43*(2), 53–69.

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

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L.B. Resnick (Ed.), *Knowing learning and instruction* (pp. 453–495). Hillsdale, MI: Lawrence Erlbaum.

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*, 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.
Cronbach, L.J. (1990). *Essentials of psychological testing* (3rd ed.). New York: Harper & Rowe.
Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability scores and profiles*. New York: Wiley.
Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, *8*(1). Retrieved March 16, 2005, from http://epaa.asu.edu/epaa/v8n1/
Dowson, M., & McInerney, D.M. (2003). What do students say about motivational goals?: Towards a more complex and dynamic perspective on student motivation. *Contemporary Educational Psychology*, *28*(1), 91–113.
Doyle, W. (1990). Classroom knowledge as a foundation for teaching. *Teachers College Record*, *91*(3), 347–360.
Flanders, N. (1970). *Analyzing teacher behaviour*. Reading, MA: Addison-Wesley.
Gray, J., Hopkins, D., Reynolds, D., Wilcox, B., Farrell, S., & Jesson, D. (1999). *Improving schools: Performance and potential*. Buckingham, UK: Open University Press.
Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, *4*(3), 365–379.
Hayes, M.T., & Deyhle, D. (2001). Constructing difference: A comparative study of elementary science curriculum differentiation. *Science Education*, *85*(3), 239–262.
Hersey, P., & Blanchard, K. (1993). *Management of organizational behavior: Utilizing human resources* (6th ed.). Englewood Cliffs, NJ: Prentice Hall.
Houtveen, A.A.M., Van de Grift, W.J.C.M., & Creemers, B.P.M. (2004). Effective school improvement in Mathematics. *School Effectiveness and School Improvement*, *15*, 337–376.
Kyriakides, L. (2002). A research-based model for the development of policy on baseline assessment. *British Educational Research Journal*, *28*(6), 805–826.
Kyriakides, L. (2004). Differential school effectiveness in relation to sex and social class: Some implications for policy evaluation. *Educational Research and Evaluation*, *10*, 141–161.
Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement*, *16*, 103–152.
Kyriakides, L. (2007). Generic and differentiated models of educational effectiveness: Implications for the improvement of educational practice. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 41–56). Dordrecht, The Netherlands: Springer.
Kyriakides, L., Campbell, R.J., & Christofidou, E. (2002). Generating criteria for measuring teacher effectiveness through a self-evaluation approach: A complementary way of measuring teacher effectiveness. *School Effectiveness and School Improvement*, *13*, 291–325.
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*, 501–529.
Ministry of Education. (1994). *The new curriculum*. Nicosia, Cyprus: Author.
Muijs, D., Campbell, R.J., Kyriakides, L., & Robinson, W. (2005). Making the case for differentiated teacher effectiveness: An overview of research in four key areas. *School Effectiveness and School Improvement*, *16*, 51–70.
Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: Some preliminary findings from the evaluation of the mathematics enhancement programme (primary). *School Effectiveness and School Improvement*, *11*, 273–303.
Muijs, D., & Reynolds, D. (2001). *Effective teaching: Evidence and practice*. London: Sage.
Noble, T. (2004). Integrating the revised Bloom's taxonomy with multiple intelligence: A planning tool for curriculum differentiation. *Teachers College Records*, *106*(1), 193–211.
Pellegrino, J.W. (2004). Complex learning environments: Connecting learning theory, instructional design, and technology. In N.M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instructional design* (pp. 25–49). Mahwah, NJ: Lawrence Erlbaum Associates.
Peterson., P., Wilkinson, L.C. & Hallinan, M. (Eds.). (1984). *The social context of instruction: Group organization and group processes*. New York: Academic Press.

Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2005). *A user's guide to MLwiN: Version 2.0*. Bristol, UK: Centre for Multilevel Modelling, University of Bristol.

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*, 193–216.

Reusser, K. (2000). Success and failure in school mathematics: Effects of instruction and school environment. *European Child & Adolescent Psychology*, *9*, 17–26.

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

Savery, J.R., & Duffy, T.M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, *35*(5), 31–38.

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

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

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*(6), 922–932.

Simons, R.J., Van der Linden, J., & Duffy, T. (2000). New learning: Three ways to learn in a new balance. In R.J. Simons, J. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 1–20). Dordrecht, The Netherlands: Kluwer.

Slater, R.O., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement*, *3*, 247–257.

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

Stenmark, J.K. (1992). *Mathematics assessment: Myths, models, good questions and practical suggestions*. Reston, VA: NCTM.

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.

Vermunt, J., & Verschaffel, L. (2000). Process-oriented teaching. In R.J. Simons, J. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 209–225). Dordrecht, The Netherlands: Kluwer.

Wang, M.C., Haertel, G.D., & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*(3), 249–294.

Yair, G. (1997). When classrooms matter: Implications of between-classroom variability for educational policy in Israel. *Assessment in Education*, *4*(2), 225–248.