Teaching and Teacher Education 25 (2009) 12–23

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development

L. Kyriakides [a,*], B.P.M. Creemers [b], P. Antoniou [a]

[a] Department of Education, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
[b] Faculty of Behavioural and Social Sciences, University of Groningen, the Netherlands

ARTICLE INFO

Article history:
Received 29 November 2007
Received in revised form 1 March 2008
Accepted 18 June 2008

Keywords:
Teacher effectiveness research
Teaching
Teacher education
Evaluation

ABSTRACT

The study reported here examines whether teaching skills included in the dynamic model of educational effectiveness can be grouped into types of teacher behaviour and whether these types are related with different student outcome measures. The data stem from a study which was conducted in order to test the validity of the dynamic model. Results reveal that teaching skills can be grouped into five types of teacher behaviour which are discerned in a distinctive way and move gradually from skills associated with direct teaching to more advanced skills concerned with new teaching approaches and differentiation of teaching. Teachers exercising more advanced types of behaviour have better student outcomes. Suggestions for research on teacher education and professional development are drawn.

1. Introduction

Educational Effectiveness Research (EER) addresses the questions on what works in education and why. Over the last two decades, studies conducted in different countries revealed that the classroom effect is more important than the school effect in explaining variation on student achievement in both cognitive and affective outcomes (Teddlie & Reynolds, 2000). Effectiveness studies also reveal that quality of teaching is the most important factor at classroom level (Brophy & Good, 1986; Fraser, Walberg, Welch, & Hattie, 1987). Moreover, researchers in the area of effectiveness developed theoretical models which attempt to explain why these teaching factors are important for learning and learning outcomes (Creemers, 1994). However, an important constraint of the existing approaches of modelling educational effectiveness is the fact that the whole process does not contribute significantly to the improvement of teaching practice (Scheerens, Glas, & Thomas, 2003). In this context, a dynamic model of educational effectiveness, which attempts to establish stronger links between effectiveness research and improvement of policy and practice, has been developed (Creemers & Kyriakides, 2008).

The dynamic model of educational effectiveness is based on two main assumptions. First, the fact that most of the effectiveness studies are exclusively focused on language or mathematics rather than on the whole school curriculum aims (cognitive, metacognitive and affective) reveals that the models of educational effectiveness should take into account the new goals of education and related to this their implications for teaching and learning. This means that the outcome measures should be defined in a more broad way rather than restricting to the achievement of basic skills. It also implies that new theories of teaching and learning are used in order to specify variables associated with the quality of teaching. Specifically, the dynamic model is based on traditional views on learning and instruction such as direct learning and teaching which emphasise not only the role of teacher as instructor responsible for providing knowledge and skills but also the specific behaviours he/she should apply (e.g., Brophy & Good, 1986; Rosenshine, 1983). The model also takes into account new ideas on learning and instruction associated with constructivism which give emphasis to independent learning and the construction of knowledge by the learner (Simons, van der Linden, & Duffy, 2000). This implies that in the later case the role of the teacher gradually moves from instructing to coaching and modeling learning. In the next part of this paper, the teacher factors of the dynamic model and their association with different approaches to teaching are described. Second, the dynamic model should not only be parsimonious but also be able to describe the complex nature of educational effectiveness. This implies that the model could be based on specific theory but at the same time some of the factors included in the major constructs of the model are interrelated within and/or between levels.

Based on the rationale of the dynamic model presented above, the essential characteristics of the model are as follows: First, the dynamic model, shown in Fig. 1, refers to multiple factors of

* Corresponding author. Tel.: +357 22752913; fax: +357 22753702.
E-mail address: kyriakid@ucy.ac.cy (L. Kyriakides).

0742-051X/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tate.2008.06.001

effectiveness which operate at different levels (i.e., student, classroom, school and system). Second, it is expected that some factors which operate at the same level are related to each other. It is therefore considered important to specify groupings of factors. In this way, specific strategies for improvement could be provided. Third, although there are different effectiveness factors and groupings of factors, 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.

In regard to the use of the model for improvement purposes at teacher level (e.g., initial training and professional development), it is assumed that teaching factors refer to knowledge and skills associated with different types of teacher behaviour in the classroom. These types of behaviour may not necessarily stem from a specific approach to teaching such as direct instruction, active teaching, differentiated teaching or more constructivist approaches to teaching. The dynamic model is not promoting specific approaches but concentrates on instructional behaviour of teachers by describing them through the use of specific teaching factors. Moreover, the model is based on the assumption that teaching factors are not separate entities but some of them are interrelated (Campbell, Kyriakides, Muijs, & Robinson, 2003; Creemers, 2007; Johnson, 1997). This implies that teachers may demonstrate types of behaviour that are based on different combinations of the various teaching approaches which can be related to student outcome measures. The model also attempts to describe the complex nature of effectiveness by pointing out not only the importance of specific factors but also by searching for grouping of factors (i.e., types of teacher behaviour). This implies that the model is based on the assumption that improvement of teacher effectiveness can be focused not on the acquisition of isolated skills/competencies (Gilberts & Lignugaris-Kraft, 1997) but on helping teachers exercise and/or develop types of teacher behaviour that are more effective than others. Since not much empirical evidence on the validity of this assumption is available, the study reported in this paper attempts to find out the types of behaviour that teachers demonstrate and the extent to which the use of specific type of behaviour can explain variation on student outcome measures. When these assumptions can be confirmed, a question can be posed concerned

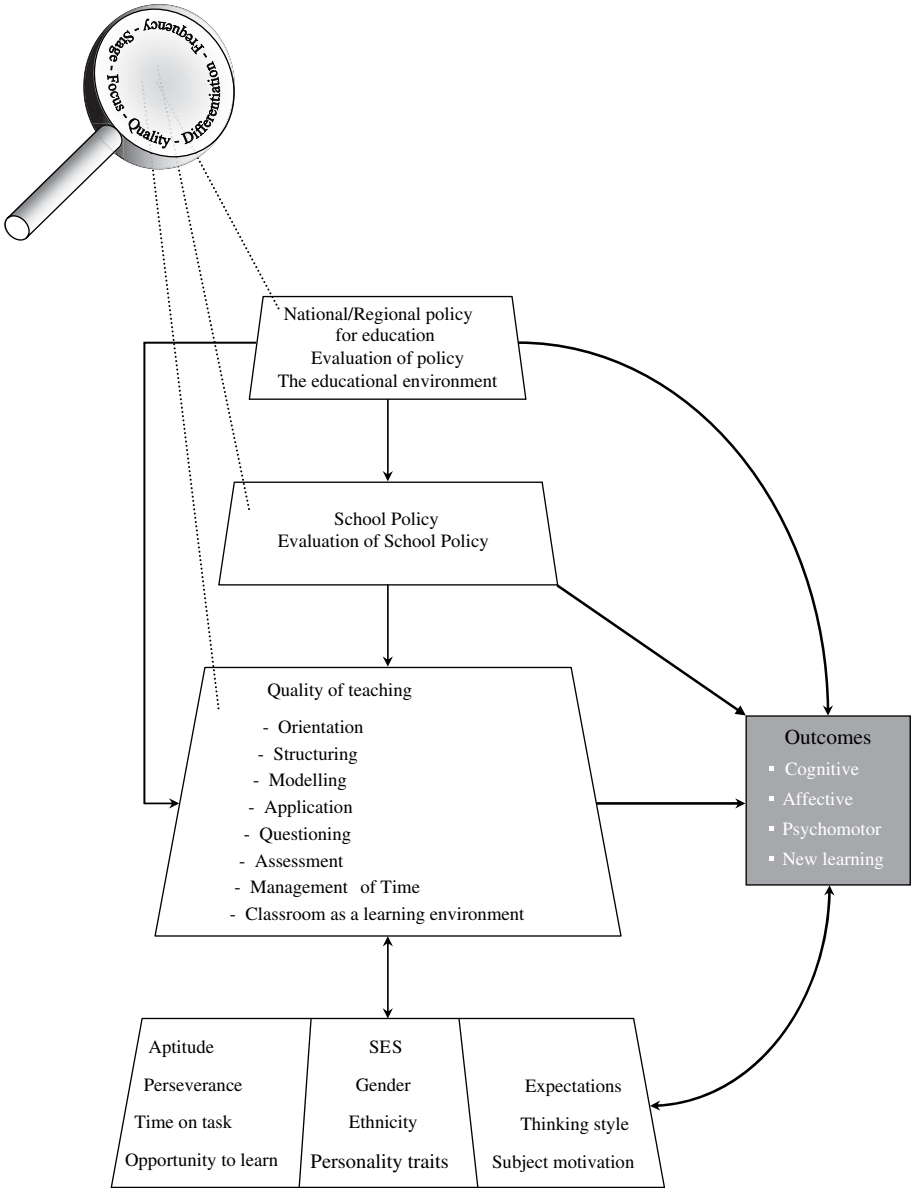

Fig. 1. The dynamic model of educational effectiveness.

with how teachers can acquire and/or develop more effective types of teacher behaviour. Although this question is beyond the scope of this research, we will elaborate on this issue and provide suggestions for further research on teacher professional development at the last section of this paper.

2. Factors of effective teaching in the dynamic model

Based on the main findings of teacher effectiveness research (e.g., Brophy & Good, 1986; Fraser, Walberg, Welch, & Hattie, 1987; Muijs & Reynolds, 2001; Opdenakker & Van Damme, 2000; Rosenshine & Stevens, 1986), the dynamic model refers to factors which describe teachers' instructional role and are associated with student outcomes. These factors refer to observable instructional behaviour of teachers in the classroom rather than on factors that may explain such behaviour (e.g., teacher beliefs and knowledge and interpersonal competences). The eight factors included in the model are as follows: orientation, structuring, questioning, teaching modelling, application, management of time, teacher role in making classroom a learning environment, and classroom assessment. These eight factors do not refer only to one approach of teaching such as structured or direct teaching (Joyce, Weil, & Calhoun, 2000) or to approaches associated with constructivism (Schoenfeld, 1998). An integrated approach in defining quality of teaching is adopted. Specifically, 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 theories of teaching associated with constructivism (Brekelmans, Sleegers, & Fraser, 2000). Moreover, the collaboration technique is included under the overarching factor contribution of teacher to the establishment of classroom learning environment. Furthermore, studies investigating differential teacher effectiveness revealed that these factors may have stronger impact for the learning of specific groups of students but should be treated as generic in nature since they were found to be related with achievement of each group of students (Campbell et al., 2003).

The dynamic model is also based on the assumption that although there are different effectiveness factors, each factor can be defined and measured using the following five dimensions: *frequency, focus, stage, quality, and differentiation.* These dimensions are supposed to contribute to the effects that a factor is expected to have on student outcome measures. Moreover, they help us describe in a better way the functioning of a factor. Specifically, frequency is a quantitative way to measure the functioning of each effectiveness factor, whereas the other four dimensions examine the qualitative characteristics of the functioning of the factor operating at the system/school/classroom level. The dimensions are not only important for a measurement perspective but also and even more for a theoretical point of view. Actions of teachers associated with each factor can be understood from different perspectives and not only by giving emphasis to the number of cases the actions occur in teaching. In addition, the use of these dimensions may help us develop strategies for improving teaching since the feedback given to teachers could refer not only to quantitative but also to qualitative characteristics of their teaching practice. The importance of taking each dimension into account is also illustrated below by explaining how one of the factors included in the model namely orientation is defined.

Orientation refers to teacher behaviour in providing the objectives for which a specific task or lesson or series of lessons take(s) place and/or challenging students to identify the reason for which an activity takes place in the lesson. The engagement of students with orientation tasks may encourage them to actively participate in the classroom since the tasks that take place are meaningful for them. As a consequence, the *frequency* dimension is measured by taking into account the number of orientation tasks that take place in a typical lesson as well as how long each orientation task takes place. These two indicators help us identify the importance that the teacher attached to this factor. The effectiveness factors are also measured by taking into account the *focus* of the activities which are associated with each factor. Two aspects of focus for each factor are 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. Thus, according to the dynamic model, the first aspect of the focus dimension of each factor addresses *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. 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 (Pellgrino, 2004). In the case of orientation, this aspect of focus is measured by examining the extent to which an activity is restricted to finding one single reason for doing a task or finding the multiple reasons for doing a task. The second aspect of this dimension refers to the specificity of the activities which can range from specific to general. The specificity of the orientation tasks is measured by taking into account that an orientation task may refer to a part of a lesson or to the whole lesson or even to a series of lessons (e.g., a lesson unit).

Activities associated with a factor can be measured by taking into account the *stage* at which they take place. It is supported 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 (Creemers, 1994). 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 (e.g., 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 (Slater & Teddlie, 1992). 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. In the case of orientation, it is taken into account that orientation tasks may take place in different parts of a lesson or series of lessons (e.g., introduction, core, ending of the lesson). Effective teachers are expected to offer orientation tasks at different parts of lessons (Killen, 2007). Further, it is expected that effective teachers are able to take others perspectives into account during this orientation phase. For example, students may come with suggestions for the reasons for doing a specific task, which an effective teacher is expected to take into account (Gijbels, Van de Watering, Dochy, & Van den Bossche, 2006).

The *quality* dimension refers to the properties of the specific factor itself, as these are discussed in the literature. This implies that the quality dimension deals with the process of teaching and is not concerned with the effects of teaching in terms of student outcomes. We assume that this dimension, as well as all the others,

may help us explain variation on student outcomes and for this reason are included in the model. The importance of using this dimension also arises from the fact that looking at the quantity elements of a factor ignores the fact that the functioning of the factor may vary. The measurement of the dimension quality refers to the properties of the orientation task and especially whether it is clear for the students. It also refers to the impact that the task has on student engagement in the learning process. For example, teachers may present the reasons of doing a task simply because they have to do it and is part of their teaching routine without having much effect on student participation, whereas others may encourage students to identify the purposes that can be achieved by doing a task and therefore to increase their motivation towards a specific task/lesson/series of lessons.

The dynamic model takes into account the findings of research into differential educational effectiveness (Campbell et al., 2003). Specifically, it is acknowledged that the impact of teaching factors on different groups of students 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. It is expected that adaptation to the specific needs of each group of students will increase the successful implementation of a factor and ultimately maximise its effect on student 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 group of students is recognised. Thus, 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, socio-economic status, ability, thinking style, and personality type (Kyriakides, 2007). However, the differentiation dimension does not imply that these groups of students are not expected to achieve the same purposes. On the contrary, adapting the functioning of each factor to the special needs of each group of 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 der Grift, & Creemers, 2004; Noble, 2004). However, the use of differentiation as a measurement dimension does not imply that all instruction has to be individualised since findings on Aptitude Treatment Interaction research reveal that in real classroom situations is neither feasible nor effective to offer only individual tasks during the whole teaching time (Corno & Snow, 1986; Good & Stipek, 1983). On the contrary, all the factors of the dynamic model and their measurement dimensions can be observed irrespective of the use of specific classroom organisation procedures and the majority of the factors can easily take place in whole class teaching.

In the case of orientation, differentiation is measured by looking at the extent to which teachers provide different types of orientation tasks to students according to their learning needs and especially by taking into account differences in the personal and background characteristics of students. Using different orientation tasks is expected to help all students to find out the reasons for which specific tasks take place in their classroom. Moreover, taking into account the different types of objectives that are supposed to be covered during the instruction, teachers are also expected to use different orientation tasks in order to introduce students to the importance of different objectives that have to be acquired (Gijbels, Van de Watering, Dochy, & Van den Bossche, 2006). Finally, teachers may differentiate the orientation tasks in relation to the organisational and cultural context of their school or classroom in order to facilitate their understanding of the purposes of learning tasks (Kyriakides, 2007).

The dynamic model attempts to describe the complex nature of effective teaching by pointing out not only the importance of specific factors and dimensions but also explaining how the functioning of each factor can be defined. The model is also based on the assumption that these factors and their dimensions may be interrelated and the importance of grouping of specific factors for explaining student achievement is stressed. In this way, not only the complex nature of effective teaching is illustrated but also specific strategies for teacher improvement may emerge. It should be acknowledged that although the effect of these factors upon achievement in different outcomes has been demonstrated (Kyriakides & Creemers, 2008), the assumption concerned with the grouping of factors is tested and reported below. Specifically, the study reported here attempts to find out whether the five dimensions of the eight teacher factors included in the dynamic model can be grouped into different types of teacher behaviour. We also investigate the extent to which the use of specific type of teacher behaviour can explain variation on different types of student outcome measures (i.e., cognitive and affective).

3. Methods

3.1. Participants

Stratified sampling 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 ($X^2 = 0.84$, d.f. $= 1$, $p = 0.42$). 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 ($t = 1.21$, d.f. $= 107$, $p = 0.22$). Although this study refers to other variables such as the socio-economic status of students and their achievement levels in different outcomes of schooling, there is 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. The teacher sample was also found to be nationally representative in terms of their background characteristics. Specifically, the t-test did not reveal any statistically significant difference between the teacher sample and the population in terms of their years of teaching experience ($t = 0.85$, d.f. $= 453$, $p = 0.39$). Moreover, the chi-square test did not reveal any statistically significant difference between the teacher sample and the population in terms of teachers' sex ($X^2 = 0.25$, d.f. $= 1$, $p = 0.62$).

3.2. Dependent variables: Student achievement in mathematics, Greek language and religious education

Data on student achievement in mathematics, Greek language, and Religious Education (RE) were collected by using external forms of assessment designed to assess knowledge and skills in mathematics, Greek language, and RE 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 RE was also measured. 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

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 RE, and also to student attitudes towards RE 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. Thus, for each student, four different scores for his/her achievement at the beginning of 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.

3.3. Explanatory variables at student level

3.3.1. 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 RE) and prior attitudes towards RE emerged from student responses to the external forms of assessment administered to students at the beginning of the school year (i.e., baseline assessment). The baseline test for each subject was in line with the national curriculum for year 4 students, whereas the tests used for measuring achievement at the end of year 5 were representative to the content of the year 5 curriculum (Ministry of Education, 1994).

3.3.2. 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 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 standardized values of the above five variables were calculated, resulting in the SES indicator.

3.4. Explanatory variables at 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 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, and (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 after the action of the teacher.

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). 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 generates data that 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, 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 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 practice 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 high-inference observation instrument covers the five dimensions of all eight factors of the model, and observers are expected to complete a Likert scale in order to indicate how often each teacher behaviour was observed. For example, an item concerned with the quality dimension of orientation is asking observers to indicate the extent to which the orientation activities that were organised during the lesson helped students understand the new content.

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 2004–2005, 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. Since more than 8% of the lessons were observed by pairs of observers, the inter-rater reliability coefficient (ρ^2) was estimated. For each subject, the coefficient was found to be 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. A Generalisability Study (Shavelson, Webb, & Rowley, 1989) on the use of students' ratings was conducted. It was found that the data emerged from almost all the questionnaire items could be used for measuring the quality of teaching of each teacher in each subject separately. Thus, the score for each teacher in each of the questionnaire item 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 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 concerned with each classroom level factor of the dynamic model in relation to each subject provided support to the construct validity of the proposed five measurement dimensions of most effectiveness factors (Kyriakides & Creemers, 2008). The two exceptions which were identified reveal the difficulty of defining the quality dimension. 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 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) standardized trait loadings, in comparison to the relatively lower (i.e., lower than .40) standardized method loadings. These findings support the use of multimethod techniques for increasing measurement validity, construct validity, and thus, stronger support for the validity of subsequent results. Thus, based on the results of the SEM analyses, 44 factor scores for the performance of each teacher in teaching each subject were estimated. Each factor score was estimated by calculating the average score emerged from the various methods used to measure the factor (i.e., the observation instruments and the student questionnaire).

4. Results

4.1. The scaling and structure of teaching skills included in the dynamic model

4.1.1. Using the Rasch model to specify the hierarchy of teaching skills' difficulties

Having established the construct validity of the framework used to measure the functioning of the teacher level factors of the dynamic model, it was decided to use the Rasch model in order to identify the extent to which the five dimensions of these factors (i.e., the 44 first order factor scores) could be reducible to a common unidimensional scale. The Rasch model does not test only the unidimensionality of the scale but also is able to find out whether the tasks can be ordered according to the degree of their difficulty. At the same time the people who carry out these tasks can be ordered according to their performance in the construct under investigation. This procedure is justified theoretically and is used in studies on teacher evaluation (e.g., Burry & Shaw, 1988; Wang & Cheng, 2001; Wright & Linacre, 1989). Specifically, the Rasch model puts people and tasks on the same scale and enables the researcher to examine the range of the teaching practice scale to see if the items/tasks within it form a continuum of teaching practice from "easy to perform" to "difficult to perform" that is devoid of gaps in construct coverage (Green & Frantom, 2002). Furthermore, the reliability of persons and items is calculated, indicating how well the scale discriminates among people based on their estimated teaching practice and how well items/tasks can be discriminated from one another on the basis of their difficulty (Andrich, 1988). Finally, Rasch analysis provides a basis for insight into the validity of a measurement tool and provides information that may limit the reliability and validity of measures made with the instrument (Sampson & Bradley, 2004). In the case of this study, specifying the position of one factor score (i.e., teaching skill) on the scale provides exact information about the individuals (teachers) who can perform sufficiently (i.e., those scoring higher than the position of this teaching skill on the scale) or insufficiently (those scoring lower than the position of this teaching skill). This analysis also makes it possible to make statements about the relative difficulty of each teaching skill. Similarly, specifying individual teacher's position on this continuum provides information about the probability of this teacher to show teaching competence below or above this position (Bond & Fox, 2001).

Thus, the Rasch model was applied on the whole sample of teachers and all 44 measures concerned with their teaching skills together using the computer program Quest (Adams & Khoo, 1996). It is important to note that we treated teacher behaviour in each subject separately meaning that 324 person estimates (i.e., for each of the 108 classrooms, three estimates of the performance of teachers to teach each subject) were generated. Two teaching skills (i.e., the focus dimension of the structuring factor and the quality dimension of time management) did not fit the model. The results of the various approaches used to test the fitting of Rasch model to our data revealed that there was a good fit to the model when teachers' performance in the other 42 teaching skills was analyzed. Information about the fitting of our data to the model is given in footnote[1].Moreover, by using the Rasch model to analyse teacher performance in relation to these 42 teaching skills included in the dynamic model, it was found that these skills were well targeted against the teachers' measures since teachers' scores range from −2.96 to 3.04 logits and the difficulties of the 42 teaching skills range from −2.69 to 3.05 logits. Moreover, the indices of persons and of teaching skills separation were found to be higher than 0.93 indicating that the separability of the scale is satisfactory. This implies that the reliability of the scale is very high and above this indicates that five levels representing different types of teacher behaviour could be discerned (Bond & Fox, 2001).

4.1.2. Using cluster analysis to specify types of teacher behaviour

Having established the reliability of the scale, one might ask if the various teaching skills are systematically grouped into levels of difficulty that may be taken to stand for types of teacher behaviour which move from relatively easy to more difficult and span across the five dimensions of the eight teacher level factors included in the dynamic model. As such, the procedure for detecting pattern clustering in measurement designs developed by Marcoulides and Drezner (1999) was used. This procedure enables us to segment the

[1] Using the Rasch model to analyse teacher estimates in the 42 teaching skills, the scale statistics which emerged revealed that the infit mean squares and the outfit mean squares were nearly one and the values of the infit *t*-scores and the outfit *t*-scores were approximately zero. It was also found that all teaching skills have item infit with the range 0.83–1.20, and item outfit with the range of 0.71–1.42. In addition, all the values of infit *t* for both teachers and teaching skills were greater than −2 and smaller than 2. Moreover, the procedure proposed by Yen (1993) was used to test for local independence. It was found that local independence was generally not violated, but if a perfect score was given to the teaching skill concerned with quality of application, the difficulty parameter of the teaching skill concerned with the quality of assessment was decreased by 0.32. However, this model violation did not result in substantial bias estimates of teaching skill parameters. Finally, the fitting of the Rasch model to the data was tested against alternative item response theory models but the improvement of fit to the existing data made by the 2PL over the Rasch model was not statistically significant. Specifically, in our study, the 2PL adds 42-item discrimination parameters (one for each teaching skill) as compared with the Rasch model, but the X^2 of 39.4 reveals that the 2PL model does not fit significantly better than the Rasch model. Similarly, the difference between the three parameter logistic model and the 2PL is not significantly better since the improvement of BILOG's chi-square index of comparative fit by 31.1 at the cost of 42 additional parameters is not statistically significant (Zimowski, Muraki, Mislevy, & Bock, 1996). It can therefore be claimed that the results of the various approaches used to test the fitting of the Rasch model to our data revealed that there was a good fit to the model.

observed measurements into constituent groups (or clusters) so that the members of any one group are similar to each other, according to selected criterion that stands for difficulty. Applying this method to segment the 42 teaching skills on the basis of their difficulties that emerged from the Rasch model showed that they are optimally clustered into the five clusters shown in Table 1. Specifically, the cumulative D for the five-cluster solution was 53%, whereas the sixth gap adds only 4%. According to the literature in cluster analysis, the five-cluster solution explaining 53% of the observed variance is considered satisfactory (Romesburg, 1984). These five clusters are further explored and specified by using the Saltus model described below.

4.1.3. Using the Saltus model to specify the structure of teaching abilities

How deep is the divide separating the five types of teacher behaviour that emerged from cluster analysis and which can be ordered into different levels according to their difficulty? The Rasch model and the clustering method used so far cannot answer this question. Wilson developed a variant of the Rasch model, the so-called Saltus model (Mislevy & Wilson, 1996; Wilson, 1989), as a method that can differentiate between different levels. Specifically, the Saltus model allows the researcher to differentiate between major and less pervasive changes in moving from one level to the other without sacrificing the idea of one common underlying

Table 1
Rasch and Saltus parameter estimates for factor scores measuring the classroom level factors of the dynamic model of educational effectiveness

Classroom level factors	Rasch	Implied within-stage difficulty (Saltus)				
	All	Level 1	Level 2	Level 3	Level 4	Level 5
Frequency management time	−2.69	−3.76	−3.76	−3.76	−3.76	−3.76
Stage management of time	−2.62	−3.65	−3.65	−3.65	−3.65	−3.65
Frequency structuring	−2.58	−3.45	−3.45	−3.45	−3.45	−3.45
Frequency application	−2.45	−3.35	−3.35	−3.35	−3.35	−3.35
Frequency assessment	−2.40	−3.00	−3.00	−3.00	−3.00	−3.00
Frequency questioning	−2.38	−2.96	−2.96	−2.96	−2.96	−2.96
Frequency teacher–student relation	−2.16	−2.50	−2.50	−2.50	−2.50	−2.50
Stage structuring	−1.56	−1.40	−2.34	−2.31	−2.28	−2.30
Quality application	−1.50	−1.36	−2.32	−2.22	−2.23	−2.28
Stage questioning	−1.48	−1.30	−2.19	−2.12	−2.06	−2.17
Frequency student relations	−1.42	−1.35	−2.26	−2.15	−2.16	−2.20
Focus application	−1.37	−1.37	−2.29	−2.09	−2.08	−2.21
Stage application	−1.33	−1.25	−2.25	−2.19	−2.09	−2.20
Quality of questions	−1.30	−1.21	−2.20	−2.08	−2.00	−2.15
Stage student relations	−0.74	−0.29	−1.10	−1.89	−1.82	−2.03
Stage teacher–student relation	−0.71	−0.22	−0.94	−1.86	−1.75	−1.97
Stage assessment	−0.62	−0.12	−0.88	−1.83	−1.74	−1.82
Frequency teaching modelling	−0.60	0.08	−0.92	−1.80	−1.68	−1.70
Frequency orientation	−0.50	0.15	−0.75	−1.93	−1.60	−1.63
Focus student relations	−0.36	0.29	−0.63	−1.73	−1.43	−1.50
Quality: feedback	−0.32	0.24	−0.64	−1.55	−1.45	−1.55
Focus questioning	−0.31	0.25	−0.55	−1.39	−1.32	−1.52
Focus teacher–student relation	−0.31	0.18	−0.72	−1.62	−1.51	−1.54
Quality structuring	−0.29	0.26	−0.64	−1.53	−1.40	−1.53
Quality assessment	−0.26	0.30	−0.48	−1.38	−1.30	−1.48
Differentiation structuring	0.59	1.02	0.36	0.28	−1.12	−1.21
Differentiation time management	0.61	1.08	0.42	0.34	−1.05	−1.15
Differentiation questioning	0.71	1.09	0.39	0.31	−1.04	−1.09
Differentiation application	0.88	1.12	0.43	0.35	−1.02	−1.12
Focus assessment	0.94	1.06	0.36	0.29	−1.06	−1.17
Differentiation assessment	1.17	1.13	0.43	0.40	−1.01	−1.09
Stage teaching modelling	1.21	1.19	0.49	0.42	−0.97	−1.05
Stage orientation	1.29	1.23	0.53	0.44	−0.95	−1.03
Quality teacher–student relation	2.32	2.10	1.50	1.40	1.12	−0.93
Quality student relations	2.39	2.21	1.61	1.42	1.15	−0.82
Dif teacher–student relation	2.50	2.25	1.64	1.44	1.17	−0.78
Differentiation student relations	2.72	2.38	1.77	1.63	1.31	−0.69
Focus orientation	2.89	2.27	1.66	1.57	1.21	−0.75
Quality orientation	2.95	2.42	1.82	1.72	1.41	−0.59
Differentiation orientation	3.00	2.55	2.00	1.88	1.58	−0.42
Quality of teaching modelling including differentiation	3.04	2.78	2.19	1.99	1.69	−0.32
Focus teaching modelling	3.05	2.91	2.21	2.10	1.80	−0.10

The Saltus parameter estimates (i.e. τ values)

Item class	Examinee stage				
	1	2	3	4	5
1	0.00*	0.00*	0.00*	0.00*	0.00*
2	0.00*	0.94	0.85	0.81	0.90
3	0.00*	0.85	1.78	1.65	1.76
4	0.00*	0.69	0.76	2.14	2.23
5	0.00*	0.61	0.75	1.05	3.03

Note 1: Empty lines are used to separate the five levels/types of teacher behaviour emerged by cluster analysis.
*Fixed at zero for model identification.

continuum. It is beyond the scope of this paper to provide a comprehensive account of the Saltus model. However, readers interested in the technical details are referred to footnote[2].

To apply the Saltus model in the present study, we assumed that the 42 teaching skills included in the dynamic model are structured in the five groups of teaching skills identified through the cluster analysis. It was found that the Saltus solution represents a better fit to the actual data rather than the Rasch model and offers a statistically significant improvement over the Rasch model which is equal to 1121 chi-square units at the cost of 30 additional parameters (i.e. 16 τs, five means, five standard deviations, and four independent proportions). Table 1 presents the difficulty parameters of the 42 teaching skills for teachers in the easiest type of teacher behaviour (i.e., Level 1 shown in column 3) and the implied within level difficulty (i.e., columns 4–7). The Saltus parameter estimates (i.e., τ values) are shown in the bottom part of the table. The following observations arise from this table:

First, difficulty parameters of teaching skills for teachers in Level 1 (i.e., the values, shown in the third column of Table 1) are more spread out than those of the Rasch Model (shown in the second column). This finding reveals that for teachers in Level 1 a large gap between the teaching skills of Level 1 and the skills in Levels 2–5 can be observed. On the other hand, for teachers who belong to Level 2, skills of Level 2 are as easy as skills of Level 1. In regard to the difficulties of the skills of Level 3, these skills are relatively difficult for Level 2 teachers but for Level 3 teachers these skills are as easy as Level 2 skills. Similar observations can be made in relation to skills of Levels 4 and 5. Second, using the figures of Table 1 and calculating the asymmetry and segmentation indices we observe that the gappiness between Levels 1 and 2 and between Levels 2 and 3 is much smaller than the gappiness between Levels 3 and 4 and Levels 4 and 5. This implies that the transition from one level to the other is not linear and moreover the transition from Level 3 to 4 and from Level 4 to 5 is much more difficult than the transition among the first three levels. A description of the different levels/types of teacher behaviour is given below.

4.1.3.1. Type 1 of teacher behaviour: Basic elements of direct teaching. The seven teaching skills situated in this type of teacher behaviour (see Table 1) refer to quantitative characteristics of factors associated with the direct teaching approach. All of them but one skill are concerned with the frequency dimension. The stage dimension of management of time is also quantitative in character and closely associated with the frequency dimension of this factor. It is interesting to note that the first two skills with the lowest difficulty estimates are concerned with management of time. This could be attributed to the fact that quantity of teaching is a prerequisite for instruction. Moreover, these seven teaching skills reveal that teachers demonstrating this type of behaviour are able to use effectively the daily routines in teaching such as keeping students on task, structuring the content of the lesson, asking questions and giving application tasks and administering assessment tasks.

4.1.3.2. Type 2: Putting aspects of quality in direct teaching and touching on active teaching. In the second type of teacher behaviour, skills which are concerned with qualitative aspects of three factors associated with the direct teaching approach (i.e., structuring, application, and questioning) are situated. Specifically, three dimensions of the application factor are included in this type of behaviour indicating that teachers of this level are able to demonstrate competences in relation to each aspect of application factor but differentiation. This indicates that application is a basic and relatively simple teacher competence. The other factor situated in this type of teacher behaviour is concerned with the questioning skills of teachers. Teachers are expected not only to use questions across the lessons but also to phrase both process and product questions appropriately. Finally, a factor concerning the role of teachers in establishing interactions among students is situated in this level. Although this factor is not exclusively associated with direct teaching, only the frequency dimension of this factor is included in this level. This implies that teachers of this level are not only able to put aspects of quality in the easiest factors associated with the direct teaching approach but are also able to encourage interactions among students which may encourage active involvement of students in learning.

4.1.3.3. Type 3: Acquiring quality in active teaching and reaching out. The 11 teaching skills situated in this type of behaviour mainly refer to qualitative characteristics of active teaching which reveal that teachers at this level are able to engage students actively in the teaching and learning processes. Moreover, teachers can create a learning environment in their classroom since all the dimensions of the two aspects of this overarching factor, but differentiation, are part of this type of teacher behaviour. Furthermore, teachers provide constructive feedback to student answers and this dimension of the questioning factor also contributes to the establishment of the classroom as an active learning environment. Similar observations can be made in relation to the quality dimension of assessment which reveals that teachers conduct assessment for formative reasons and thereby integrate assessment into teaching and learning. A new element of this level is concerned with the frequency dimension of two factors associated with the new teaching approach namely teaching modelling and orientation. This implies that teachers at this level are not only able to effectively use strategies related to direct and active

[2] Formally, when comparing two groups of persons, the Saltus model states that the difficulty parameter changes by a certain amount for a subtest of items in one group, $P(X_{ij}=1)=f(\theta_j-\beta_i+\tau)$ in which τ denotes the change in difficulty, also called the Saltus parameter, and f is the logistic distribution function (Wilson, 1989). A positive value of τ implies that all of the items to which the Saltus parameter pertains become easier to the same extent in that group. For the other items in this group and for all the items in the other group, the difficulty remains unchanged. Hence, when two groups of persons who are assumed to be at different levels of development are compared a positive and significant value of the τ parameter for a subset of items in the more developed group may reflect a discontinuity in development that may reflect some kind of qualitative change. The change consists in part of the items being easier. It is situated on the same dimension and supplements a progression along the same latent scale. It can be claimed that the jump made by the group at the higher development level reflects a kind of reorganization of the processes underlying the handling of items that renders them easier than before. In most theories dealing with stages of development, developmental sequences involve levels that are qualitatively discrete from each other and they follow a constant order of succession. In the Saltus model, these two aspects of development are summarized by the twin concepts of "gappiness" and "rigidity", respectively. Gappiness is indicated in the Saltus by a segmentation or a break, so to speak, of the logit scale. That is, segmentation is specified as the distance between the most difficult item of Level 1 and the easiest item of Level 2. Technically speaking, segmentation is measured through two segmentation indices, one for each person group. The difference between these two segmentation indices is called the asymmetry index. Wilson (1985) points out that asymmetry refers to the relative difference in difficulty of the item types from the perspectives of the two person groups. When the asymmetry index is zero, the Saltus model is equivalent to the Rasch model, which can be interpreted to mean that the difference in difficulty between the two item types is the same for both person groups. On the other hand, if the asymmetry index is positive, the Level 1 persons perform on the items as being further apart in difficulty than do students in Level 2. This pattern indicates rigidity and is typical, therefore, of hierarchical development. That is, the upper stage items are near to impossible for persons at the lower stage, but persons at the upper stage can solve the items of both stages, although finding some difficulty in dealing with the upper stage items and making some errors in dealing with the lower stage items is reasonable. This diminishes the observed difference in difficulty of the item types. This pattern is also manifested in a jump in the predicted probability of success at the border between the two groups that is not present when the asymmetry index is zero.

teaching but also use techniques in their instruction associated with constructivism.

4.1.3.4. Type 4: Differentiation of teaching and putting aspects of quality in new teaching. The eight teaching skills situated in this level are mainly concerned with the differentiation dimension of factors associated with direct teaching. Teachers at this level are able to differentiate their teaching practice according to their students' needs and offer appropriate application, and structuring tasks for each group of students. In addition, different questions and assessment techniques are offered to each group of students which are in line with their learning needs. Another element of this level is concerned with the stage dimension of two factors associated with the new teaching approach. Thus, teachers at this level are not only able to differentiate their instructions but also to incorporate some qualitative characteristics of teaching modelling and orientation. Specifically, they are not only able to provide enough tasks associated with these two factors but they also offer them at appropriate occasions.

4.1.3.5. Type 5: Achieving quality and differentiation in teaching using different approaches. Finally, the nine teaching skills of this level are concerned with the most difficult qualitative characteristics of factors related to both active teaching and the new teaching approach. Specifically, the first four skills are concerned with the quality and differentiation dimensions of the classroom as a learning environment factor stressing both teacher–student and student–student interactions (den Brok, Bergen, Stahl, & Brekelmans, 2004). The other five skills are associated with the focus, quality, and differentiation dimensions of the new teaching approach. Therefore, teachers of this level use effectively a variety of teaching approaches and are also able to incorporate the qualitative characteristics of these approaches in their teaching practice. One may assume that teachers at this level are the most effective and this assumption is tested in the next part of this section.

Looking at the description of these five types of teacher behaviour in terms of the teaching skills and approaches situated in each type, we can see that the first three levels are mainly related to the direct and active teaching approach by moving from the basic requirements concerning quantitative characteristics of teaching routines to the more advanced requirements concerning the appropriate use of these skills as these are measured by the qualitative characteristics of these factors. One could also observe that these skills gradually move from the use of teacher-centre approaches to the active involvement of students in teaching and learning. The last two types of teacher behaviour are more demanding since teachers are expected to differentiate their instruction and also to demonstrate their ability to use instructional techniques associated with the new teaching approach. Again, a progression from quantitative characteristics of factors associated with the new teaching approach to their qualitative aspects can be observed in Levels 4 and 5. The content description of these five types of teacher behaviour and the distinction between Levels 1–3 versus Levels 4 and 5 can be seen as a justification of the results emerged from the Saltus model which shows the gap between the levels/types of teacher behaviour in general and also the relatively higher gappiness in moving from type 3 to type 4 and from type 4 to type 5 of teacher behaviour.

4.2. The added value of classifying teachers into levels of teaching competences: Explaining variation on student achievement in different outcomes

Not only the construct validity of the developmental scale which refers to the teaching skills included in the dynamic model should be demonstrated but also its significance and relevance to the field of teacher effectiveness should be investigated. For this reason, it was decided to examine the extent to which the classification of teachers into these five levels explains variation in achievement in each of the four types of outcomes of schooling. Separate multilevel analysis for each dependent variable was performed. The first step in each analysis was to determine the variance at individual, class, and school level without explanatory variables (i.e., baseline 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. Grouping variables were entered as dummies with one of the groups as baseline (e.g., boys = 0). The models presented in Table 2 were estimated without the variables that did not have a statistically significant effect at the 0.05 level.

A comparison of the baseline 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 RE. Moreover, the teacher effect was found to be higher on achievement of cognitive rather than affective aims of RE. In model 1, the context variables at student, classroom, and school levels were added to the baseline model. The following observations arise from the figures of model 1 in 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 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 in RE. 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 mathematics. Finally, prior knowledge (i.e., aptitude) has the strongest effect in predicting student achievement at the end of the school year. Aptitude is the only contextual variable which had a consistent effect on achievement when aggregated either at the classroom or the school level.

At the next step of the analysis, we examined whether classification of teachers into the five levels presented above can help us explain variance of student achievement in each outcome of schooling. For this reason, teachers at Level 3 were treated as reference group and four dummy variables were entered in model 1. We can observe that the students of teachers at Level 1 had the lowest achievement in each outcome measure whereas students of teachers at Levels 4 and 5 had higher achievement than students of the first three levels. In each model 2, one can also observe that students of teachers who were found to belong to higher levels performed better than students of teachers at lower levels. The only exception to this rule is concerned with the fact that in mathematics students of teachers of Level 3 did not outperform students of Level 2. In RE, no teacher was found to belong to Level 5 and it was therefore not possible to compare the performance in RE of students of teachers who belong to Level 4 with students of teachers who belong to Level 5.

5. Discussion

This study provides some support to the assumption of the dynamic model that teacher level factors are interrelated and should not be treated as isolated. Moreover, the use of specific measurement dimensions to describe not only quantitative but also qualitative characteristics of these factors helps us define 42 teaching skills that are grouped into five types of teacher behaviour which move from relatively easy to more advanced. These five types of behaviour are also described in a distinctive way. Furthermore, taking student outcomes as criteria of effectiveness, it was found

Table 2
Parameter estimates and (standard errors) for the analyses of achievement in Greek language, mathematics and religious education (cognitive and affective aims)

Factors	Greek language			Mathematics			Religious education (cognitive)			Religious education (affective)		
	Model 0	Model 1	Model 2	Model 0	Model 1	Model 2	Model 0	Model 1	Model 2a	Model 0	Model 1	Model 2
Fixed part (intercept)	−0.39 (.08)	−0.33 (.08)	−0.30 (.08)	0.36 (.05)	0.30 (.05)	0.13 (.02)	−0.79 (.11)	−0.63 (.09)	−0.61 (.08)	0.61 (.08)	0.50 (.07)	0.43 (.07)
Student level												
Prior knowledge		0.49 (.05)	0.48 (.05)		0.71 (.12)	0.70 (.12)		0.51 (.05)	0.49 (.05)		0.41 (.10)	0.40 (.10)
Sex (boys = 0, girls = 1)		0.23 (.10)	0.19 (.09)		−0.18 (.07)	−0.15 (.07)		0.23 (.09)	0.19 (.09)		0.18 (.07)	0.15 (.07)
SES		0.32 (.06)	0.27 (.05)		0.60 (.25)	0.55 (.24)		0.12 (.05)	0.10 (.05)		NSS	NSS
Classroom level												
Context												
Average prior knowledge		0.15 (.05)	0.10 (.04)		0.31 (.11)	0.28 (.10)		0.25 (.07)	0.21 (.07)		0.21 (.08)	0.18 (.07)
Average SES		0.09 (.04)	0.06 (.03)		0.15 (.04)	0.13 (.04)		0.09 (.04)	0.08 (.04)		NSS	NSS
Percentage of girls		NSS*	NSS		−0.05 (.02)	−0.05 (.02)		NSS	NSS		0.05 (.02)	0.04 (.02)
Quality of teaching/teacher behavior												
Type 1			−0.22 (.05)			−0.24 (.07)			−0.19 (.04)			−0.18 (.03)
Type 2			0.12 (.04)			NSS			−0.10 (.04)			−0.11 (.05)
Type 4			0.16 (.06)			0.18 (.04)			0.15 (.06)			0.12 (.04)
Type 5			0.29 (.05)			0.28 (.05)			NA**			NA
School level: context												
Average SES		NSS	NSS		NSS	NSS		NSS	NSS		NSS	NSS
Average prior knowledge		0.13 (.05)	0.11 (.05)		0.11 (.05)	0.08 (.04)		0.13 (.05)	0.12 (.05)		0.08 (.02)	0.06 (.02)
Percentage of girls		NSS	NSS		NSS	NSS		NSS	NSS		NSS	NSS
Variance components												
School	9.5%	7.7%	7.6%	11.5%	8.1%	7.5%	8.0%	7.7%	7.6%	7.5%	7.0%	6.7%
Class	15.2%	11.1%	5.7%	15.4%	9.3%	6.0%	13.2%	11.1%	7.5%	10.4%	9.3%	6.3%
Student	75.3%	31.5%	28.9%	73.1%	30.9%	29.7%	78.8%	34.5%	29.3%	82.1%	32.6%	31.7%
Explained		49.7%	57.8%		51.7%	56.8%		46.7%	55.6%		51.1%	55.3%
Significance test												
X^2	1015.6	686.7	521.5	1224.3	984.9	875.9	1823.6	1457.1	1307.6	1024.5	835.1	725.2
Reduction		328.9	165.2		239.4	119.0		366.5	149.5		189.4	109.9
Degrees of freedom		6	4		7	3		6	3		5	3
p-Value		.001	.001		.001	.001		.001	.001		.001	.001

*NSS = No statistically significant effect; **NA = Since there was no teacher of RE who was situated in type 5, this dummy variable was not entered in model 1.

out that teachers who use more advanced types of behaviour were more effective than those demonstrating the relatively easy types. This association is found for achievement in different subjects and for both cognitive and affective outcomes. Therefore, in order to understand the complex nature of effective teaching not only specific factors and their measurement dimensions should be defined but also the concept of grouping of factors should be used in theoretical frameworks concerned with effective teaching. It can also be claimed that this study provides strong support to the validity of the dynamic model since not only the importance of the eight factors and their dimensions is demonstrated but also the added value of searching for grouping of factors and their dimensions has been illustrated.

Moreover, the results of this study provide support to the attempt of the dynamic model to describe effective teaching using an integrated approach. Specifically, skills associated with both the direct teaching and the new teaching approaches were found to belong to the same levels/types of teacher behaviour (see for instance the description of Levels 3–5). Moreover, the types of teacher behaviour which were discerned support the idea of combining teaching skills within each type of behaviour rather than treating each skill or factor in an isolated way. Therefore, this study provides further support to those indicating the limitations of using exclusively either the direct teaching approach (e.g., Steffe & Gale, 1995) or approaches associated with constructivism (e.g., Kirschner, Sweller, & Clark, 2006) to describe effective teaching.

Given that the dynamic model was designed in order to contribute to the improvement of education practice at the levels of classroom, school, and system, this study can be seen as a step towards the development of a research programme searching for ways of using the model for teacher professional development purposes. In this context, the last section of this paper provides suggestions for research on teacher education and professional development. First, it is pointed out that this study reveals that the five types of teacher behaviour, which emerged from the dynamic model, vary in difficulty and move gradually from relatively simple types of behaviour to more advanced types. This finding implies that we should examine whether teachers may also move gradually from one type of teacher behaviour to a more complicated type of behaviour. Second, the results of the Saltus model indicate that the transition from one type to the other may not be linear since smaller gaps among the first three types of behaviour were identified whereas the gaps between Levels 3 and 4 and between Levels 4 and 5 were much bigger. This finding implies that longitudinal studies should be conducted in order to find out whether moving from one type to the other is relatively easy in the case of the first three types of behaviour and more difficult in moving from Levels 3 to 5. Finally, the results of this study indicate that teachers at more advanced levels are more effective in terms of both cognitive and affective outcome measures. Therefore, the results of this study reveal the importance of conducting longitudinal studies to find out how teachers can move from an easier type of behaviour to a more advanced level in order to become more effective.

Although the results of this study could be interpreted as providing support to the use of stage models of professional development (e.g., Berliner, 1994; Dreyfus & Dreyfus, 1986; Sternberg et al., 2000), we acknowledge the limitations of these models and especially the fact that there are no empirical evidence supporting that progression occurs in a stepwise manner from one stage to the next (Dall' Alba & Sandberg, 2006). It is also important to note that cross-sectional studies are very likely to give rise to a stage notion of development because they focus on measuring skills at different levels of experience. However, finding differences among teachers in their teaching skills does not necessarily imply that transition from one level to the other can occur in a stepwise manner. Problems are likely to arise when cross-sectional studies which do not explore the development of teaching skills over time are used in making assumptions about how development occurs. Given that the aim of the study reported here was to test the validity of the dynamic model and illustrate the importance of grouping teaching factors into types of teacher behaviour, teaching skill acquisition was not investigated over time. Therefore, the results of this study cannot be used in order to make claims on how development occurs. Thus, a question that arises from this study is whether stepwise development of types of teacher behaviour can be achieved and if so what type of programmes of teacher professional development should be offered in order to improve teacher effectiveness.

It is also acknowledged that the notion that skill development can be conceptualised as a defined body of knowledge and skills has been questioned from within a range of research approaches (e.g., Ball & Cohen, 1999; Billett, 2001; Dall'Alba & Sandberg, 1996; Schon, 1987). A principal critique of this approach is that a focus on stages veils more fundamental aspects of development; it directs attention away from the skill that is being developed. Dall'Alba and Sandberg (2006) argue that a fundamental dimension of professional skill development concerned with the understanding of, and, in practice is overlooked in stage models. It is also claimed that understanding should not be seen as limited to cognitive content or activity but should be embedded in dynamic, intersubjective practice (Dall'Alba, 2004). In this literature, knowledge and beliefs are seen as "filters in the mind" for each situation that is encountered and for learning from these situations (Borko et al., 1997; Day, 1993). However, this begs the question of how the gap between the content of the mind and professional practice is bridged. This implies that further research is needed to extend and deepen our understanding of the ways in which embodied understanding of professional practice is interrelated with performance and development of that practice. Longitudinal studies are also needed in order to find out whether specific approaches to teacher professional development are more effective than others. For example, one could argue that in order to help teachers develop their teaching skills moving from an easier type of behaviour to a more advanced level, an overemphasis on the skills that have to be developed may underestimate the importance of teacher understanding of, and, in practice and that development of practice could emerge through emphasising not only the acquisition of specific teaching skills but also through advocating that reflection and inquiry are needed (e.g., Calderhead & Shorrock, 1997; Ponte, Ax, Beijard, & Wubbels, 2004; Van Huizen, Van Oers, & Wubbels, 2005). The dynamic model may help in providing such opportunities for understanding of, and, in practice and for helping teachers improve their effectiveness. However, further research is needed on how learning environments could be established in order to give opportunities to teachers call into question and extent their understanding of, and, in practice.

Finally, suggestions for further research on testing the generalisability of the findings of this study can be formulated. The results of this study are based on research conducted in a single country and thereby studies in other countries as well as in other subject areas using various outcome measures could help us test the generalisability of our findings. Case studies can also be conducted to find out the difficulties that teachers experience in moving to the next level up and clarify the barriers associated with the gappiness between levels as well as the difficulties of promoting teacher professional development. Furthermore, using this theoretical framework and conducting longitudinal studies on teacher professional development might also help us consider the dynamic character of effectiveness since not only teachers moving to a higher level might be identified but also teachers who drop to a lower level for several reasons (including burnout) could be identified. Such findings may also reveal that in helping teachers improve their skills

other factors such as their efficacy beliefs and attitudes towards the teaching profession should be considered.

References

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

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

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: toward a practice-based theory of professional education. In L. Darling-Hammond, & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Berliner, D. (1994). Expertise: the wonder of exemplary performances. In J. Mangieri, & C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives* (pp. 161–186). Fort Worth, TX: Harcourt Brace College.

Billett, S. (2001). Knowing in practice: reconceptualising vocational expertise. *Learning and Instruction, 11*, 431–452.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education, 13*, 259–278.

Brekelmans, M., Sleegers, P., & Fraser, B. (2000). Teaching for active learning. In P. R. J. Simons, J. L. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 227–242). Dordrecht: Kluwer Academic Publishers.

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.

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

Calderhead, J., & Shorrock, S. B. (1997). *Understanding teacher education*. London: Falmer Press.

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.

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). (pp. 605–629) New York: MacMillan.

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

Creemers, B. P. M. (2007). Combining different ways of learning and teaching in a dynamic model of educational effectiveness. *Journal of Basic Education, 17*(1).

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

Dall'Alba, G. (2004). Understanding professional practice: investigations before and after an educational program. *Studies in Higher Education, 29*, 679–692.

Dall'Alba, G., & Sandberg, J. (1996). Educating for competence in professional practice. *Instructional Science, 24*, 411–437.

Dall'Alba, G., & Sandberg, J. (2006). Unveiling professional development: a critical review of stage models. *Review of Educational Research, 76*, 383–412.

Day, C. (1993). The importance of learning biography in supporting teacher development: an empirical study. In C. Day, J. Calderhead, & P. Denicolo (Eds.), *Research on teacher thinking: understanding professional development* (pp. 221–232). London: Falmer Press.

den Brok, P., Bergen, T., Stahl, R. J., & Brekelmans, M. (2004). Students' perceptions of teacher control behaviours. *Learning and Instruction, 14*, 425–433.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.

Flanders, N. (1970). *Analyzing teacher behavior*. Reading, MA: Addison-Wesley.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*, 145–252.

Gijbels, D., Van de Watering, G., Dochy, F., & Van den Bossche, P. (2006). New learning environments and constructivism: the students' perspective. *Instructional Science, 34*(3), 213–226.

Gilberts, G. H., & Lignugaris-Kraft, B. (1997). Classroom management and instruction competencies for preparing elementary and special education teachers. *Teaching and Teacher Education, 13*(6), 597–610.

Good, T. L., & Stipek, D. J. (1983). Individual differences in the classroom: a psychological perspective. In G. D. Fenstermacher, & J. I. Goodlad (Eds.), *Individual differences and the common curriculum (82nd yearbook of the National Society for the Study of Education, part 1* (pp. 9–14). Chicago: University of Chicago Press.

Gray, J., Hopkins, D., Reynolds, D., Wilcox, B., Farrell, S., & Jesson, D. (1999). *Improving schools: Performance and potential*. Buckingham, UK: Open University Press.

Green, K. E., & Frantom, C. G. (2002). Survey development and validation with the Rasch model. Paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing. Charleston, SC, November 14–17, 2002.

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*(3), 337–376.

Johnson, B. (1997). An organizational analysis of multiple perspectives of effective teaching: implications of teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 69–87.

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

Killen, R. (2007). *Effective teaching strategies: Lessons from research and practice* (4th ed.). Thomson: Social Science Press.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry based teaching. *Educational Psychologist, 41*(2), 75–86.

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., & 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.

Marcoulides, G. A., & Drezner, Z. (1999). A procedure for detecting pattern clustering in measurement designs. In M. Wilson, & G. Engelhard, Jr. (Eds.), *Objective measurement: Theory into practice, Vol. 5*. Ablex Publishing Corporation.

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

Mislevy, R. J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrica, 61*, 41–71.

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.

Opdenakker, M. C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: similarities and differences between school outcomes. *School Effectiveness and School Improvement, 11*(2), 65–196.

Pellgrino, 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.

Ponte, P., Ax, J., Beijard, D., & Wubbels, T. (2004). Teacher's development of professional knowledge through action research and the facilitation of this by teacher educators. *Teaching and Teacher Education, 20*, 571–588.

Romesburg, H. C. (1984). *Cluster analysis for researchers*. BelmontCalifornia: Lifetime Learning Publication and Wadsworth Inc.

Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal, 83*(4), 335–351.

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.

Sampson, S. O., & Bradley, K. D. (2004). Measuring factors impacting educator supply and demand: an argument for Rasch analysis. Paper presented at the American Educational Research Association Conference, San Diego, CA, April 2004.

Scheerens, J., Glas, C., & Thomas, S. (2003). *Educational evaluation, assessment and monitoring: a systemic approach*. Lisse: Swets & Zweitlinger Publishers.

Schon, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.

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*. Dordrecht, the Netherlands: Kluwer.

Slater, R. O., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement, 3*(4), 247–257.

Steffe, L., & Gale, J. (Eds.). (1995). *Constructivism in education*. Hillsdale, N.J: Lawrence Erlbaum Associates, Inc.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, A. J., Wagner, R. K., Williams, W. M., et al. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.

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

Van Huizen, P., Van Oers, B., & Wubbels, T. (2005). A Vygotskian perspective on teacher education. *Journal of Curriculum Studies, 37*(3), 267–290.

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

Wilson, M. (1985). *Measuring stages of growth: a psychometrical model of hierarchical development*. (ACER occasional paper no. 19). Hawthorn, Australia: Australian Council for Educational Research.

Wilson, M. (1989). Saltus: a psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105*(2), 276–289.

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal: measurements, however, must be interval. *Archives of Physical Measurement and Rehabilitation, 70*(12), 857–860.

Yen, W. (1993). Scaling and performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187–213.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG*. Chicago, IL: SSI.