Teacher behavior and student outcomes: Results of a European study

Anastasia Panayiotou · Leonidas Kyriakides · Bert P. M. Creemers · Léan McMahon · Gudrun Vanlaar · Michael Pfeifer · Galini Rekalidou · Matevž Bren

Abstract This study investigates the extent to which the factors included in the dynamic model of educational effectiveness are associated with student achievement gains in six different European countries. At classroom level, the dynamic model refers to eight factors relating to teacher behavior in the classroom: orientation, structuring, questioning, teaching-modeling, application, management of time, teacher role in

A. Panayiotou · L. Kyriakides (✉)
Department of Education, University of Cyprus, 20537, 1678 Nicosia, Cyprus
e-mail: kyriakid@ucy.ac.cy

A. Panayiotou
e-mail: panayiotou.anastasia@ucy.ac.cy

B. P. M. Creemers
Faculty of Behavioural and Social Sciences, Department of Pedagogy & Educational Science, University of Groningen, Groningen, The Netherlands
e-mail: b.p.m.creemers@rug.nl

L. McMahon
Economic and Social Research Institute (ESRI), Dublin, Ireland
e-mail: lean.mcmahon@esri.ie

G. Vanlaar
Katholieke Universiteit Leuven, Leuven, Belgium
e-mail: Gudrun.Vanlaar@ppw.kuleuven.be

M. Pfeifer
Technical University Dortmund, Dortmund, Germany
e-mail: Pfeifer@ifs.tu-dortmund.de

G. Rekalidou
Democritus University of Thrace, Komotini, Greece
e-mail: grekalid@psed.duth.gr

M. Bren
University of Maribor, Maribor, Slovenia
e-mail: matevz.bren@fvv.uni-mb.si

making classroom a learning environment, and classroom assessment. This paper presents results concerned with the impact of the teacher factors on student achievement. In each participating country (i.e., Belgium/Flanders, Cyprus, Germany, Greece, Ireland, and Slovenia), a sample of at least 50 primary schools (*n*=334) was drawn. Written tests in mathematics and science were administered to all grade 4 students (*n*= 10,742) at the beginning and at the end of the school year 2010–2011. Students were also asked to complete a questionnaire concerning the eight teacher factors of the dynamic model. Structural equation modeling techniques were used to test the construct validity of the student questionnaire. Both across and within country analyses revealed that student ratings are reliable and valid for measuring the functioning of the teacher factors included in the dynamic model. Multilevel analyses revealed that teacher factors are associated with student achievement gains in mathematics and science. Implications for the development of educational effectiveness research and for improving quality of teaching are drawn.

1 Introduction

Educational effectiveness research (EER) addresses the question of what works in education and why. Since the 1980s, EER has enjoyed rapid expansion in many countries. Methodological and technological advances have improved the power of estimation of teacher and school differences in student achievement (Goldstein 2003). However, EER lacks cross-cultural perspectives and has been criticized for showing strong ethnocentric tendencies (Reynolds 2000). Although EER publications have acknowledged seminal studies by Coleman et al. (1966), Edmonds (1979), Brookover et al (1979), Rutter et al. (1979), and Mortimore et al. (1988), a science of EER has not been developed. The absence of cross-national perspectives and intercultural relationships between educational effectiveness researchers is seen as intellectually damaging.

Specifically, over the last two decades, a trend has begun to emerge whereby education policy makers have attempted to apply simplistic suggestions for raising standards based on the notion of "transplanting" knowledge and ideas from one country, or one culture, to another. Examples of this practice include the proposed lengthening of the school day and shortening of school holidays discussed in some American states, following the popularity of the "time to learn" approach in Japan and the trend in British primary schools toward whole-class direct instruction, which is clearly based on the enthusiasm for this approach in the Pacific Rim countries (Reynolds 2000). Many EER researchers are voicing concern about the potential hazards of this practice of transplanting educational policies from one country or culture to another (e.g., Reynolds 2006; Scheerens 2013). The concerns are largely based on effectiveness studies that have shown how factors that seem successful in some countries may not be appropriate elsewhere. For example, some US studies have found positive association between "assertive principal instructional leadership" and student achievement gains (e.g., Hallinger and Heck 2011; Louis et al. 2010), but within-country research in Europe (e.g., The Netherlands, Cyprus) has not found the same level of support for this

factor (e.g., Kyriakides 2008; Scheerens and Bosker 1997). However, in an era when educational policy appears to be following the international route, it is unfortunate that EER appears to be lingering in a uni-culture research dimension, pursuing within-country studies rather than cross-national research. Yet, arguably, cross-national and multicultural studies on educational effectiveness are required in order to develop understanding about the complex structures of education policy across different countries and cultures and to explain how policies affect student outcomes in different settings.

The cross-national European study of educational effectiveness described in this paper aims to contribute to the development of the international dimension of EER and to provide a response to the knowledge and theoretical gaps in the field. More specifically, it investigates the extent to which the dynamic model of educational effectiveness (Creemers and Kyriakides 2008) can be used as a starting point for establishing a framework to identify the factors that work in different educational settings and help understand the reasons why. The dynamic model is multilevel in nature and refers to factors operating at four levels: student, teacher, school, and system. This paper is concerned with the generic nature of teacher factors and the implications for student achievement in mathematics and science in elementary education.

2 The dynamic model of educational effectiveness

The dynamic model is multilevel in nature and refers to factors operating at four levels: student, teacher, school, and system. The model emphasizes the teaching and learning situations and analyzes the roles played by the two main actors (i.e., teacher and student). According to the model, the teaching and learning situations are influenced by school-level factors, through the development and evaluation of school policy on teaching and the policy on creating a learning environment at the school. The model also shows how the teaching and learning situations are influenced by the system level through formal avenues, especially through developing and evaluating education policy at national/regional level. In addition, the model takes into account how the teaching and learning situations are influenced by the wider educational context in which students, teachers, and schools are expected to operate. Factors such as societal values for learning and the level of social and political importance attached to education play important roles both in shaping teacher and student expectations, and opinion formation of various stakeholders about what constitutes effective teaching practice.

3 The teacher factors of the dynamic model

Drawn from the main findings of EER (e.g., Brophy and Good 1986; Darling-Hammond 2000; Doyle 1990; Muijs and Reynolds 2000; Rosenshine and Stevens 1986; Scheerens and Bosker 1997), the dynamic model refers to the following eight factors that describe the teachers' instructional role and are associated with student outcomes: orientation, structuring, questioning, teaching-modeling, application, management of time, teacher role in making classroom a learning environment, and classroom assessment. The combination of these eight factors defines quality of

teaching, with reference to established teaching approaches such as the direct and active teaching model, and the constructivist approach. A short description of each teacher factor follows (for detailed information see Creemers and Kyriakides 2008).

1. *Orientation*: Refers to teacher behavior 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(s) the lesson involves a particular activity. It is anticipated that the orientation process can make tasks and lessons meaningful to students, which in turn encourages their active participation in the classroom (e.g., De Corte 2000; Paris and Paris 2001). Therefore, orientation tasks should take place in different parts of a lesson or series of lessons (e.g., introduction, core, and ending of the lesson) and in lessons that are expected to achieve different types of objectives. Moreover, orientation tasks are considered appropriate according to whether they are clear to students, whether students are encouraged to identify the purposes that can be achieved by carrying out a task, and whether all student views are taken into account.

2. *Structuring*: Rosenshine and Stevens (1986) pointed out that student achievement is maximized when teachers actively present materials and structure them by the following: (a) beginning with overviews and/or review of objectives, (b) outlining the content to be covered and signaling transitions between lesson parts, (c) calling attention to main ideas, and (d) reviewing main ideas at the end. Also, found to be important for student achievement is provision of summary reviews, since they integrate and reinforce the learning of major points (Brophy and Good 1986). These structuring elements facilitate memorizing of information and allow for its apprehension as an integrated whole with recognition of the relationships between parts. Moreover, student achievement levels tend to be higher when information is presented in the form of repeating and reviewing general views and key concepts. It is important to note that the structuring factor also refers to the ability of teachers to increase the difficulty level of their lessons or series of lessons gradually (Creemers and Kyriakides 2006).

3. *Questioning*: Based on the results of studies concerned with teacher questioning skills and their association with student achievement, this factor is defined in the dynamic model according to five elements. First, teachers are expected to offer a mix of product questions (i.e., expecting a single response from students) and process questions (i.e., expecting students to provide more detailed explanations) (Askew and William 1995; Evertson et al. 1980). Second, the length of pause following questions is taken into account, and it is expected to vary according to the level of difficulty of the questions. Third, question clarity is measured by investigating the extent to which students understand what is required of them, i.e., what the teacher expects them to do or find out. Fourth, the appropriateness of the level of difficulty of questions: It is expected that most questions should elicit correct answers and most of the other questions should elicit overt, substantive responses (incorrect or incomplete answers), rather than failure to respond at all (Brophy and Good 1986). Fifth, the way teachers deal with student responses to questions is investigated; correct responses should be acknowledged as such. In responding to students' partly correct or incorrect answers, effective teachers acknowledge whatever part may be correct, and if they consider, there is a good

prospect for success, they try to elicit an improved response (Rosenshine and Stevens 1986). Therefore, effective teachers are able to sustain the interaction with the original respondent by rephrasing the question and/or giving clues to its meaning, rather than terminating the interaction by providing the student with the answer or calling on another student to respond.

4. *Teaching-modeling*: Although there is a long tradition in research on teaching higher-order thinking skills and problem solving, these teaching and learning activities have received unprecedented attention during the last two decades due to the policy emphasis on the achievement of new goals of education. Thus, the teaching-modeling factor is associated with findings of effectiveness studies revealing that effective teachers are expected to help pupils use strategies and/or develop their own strategies that can help them solve different types of problems (Grieve 2010). Consequently, students are expected to develop skills that help them organize their own learning (e.g., self-regulation and active learning). In defining this factor, the dynamic model also addresses the properties of teaching-modeling tasks, and the role of teachers are expected to play in order to help students devise problem-solving strategies. Teachers may either present students with a clear problem-solving strategy, or they may invite students to explain how themselves would approach or resolve a particular problem and then use that information for promoting the idea of modeling. Recent research suggests that the latter approach may encourage students to not only use, but also develop their own problem-solving strategies (Aparicio and Moneo 2005; Gijbels et al. 2006).

5. *Application*: By taking into account the cognitive load theory, which supports that only a restricted amount of information can be processed in the working memory and can be retained in the short-term memory (Kirschner 2002; Paas et al. 2003), each lesson is expected to include application activities. Effective teachers are assumed to use seatwork or small-group tasks to provide students with practice and application opportunities (Borich 1992). Beyond looking at the number of application tasks given to students, the application factor investigates whether students are simply asked to repeat what has already been covered by the teacher or if the application task is set at a more complex level than the lesson. It also examines whether the application tasks are used as starting points for the next step of teaching and learning. Moreover, this factor refers to teacher behavior in monitoring and supervising and giving corrective feedback during application activities. Brophy and Good (1986) argue that once the students are released to work independently, effective teachers circulate to monitor progress and provide help and feedback.

6. *The classroom as a learning environment*: This factor comprises five elements, i.e., teacher-student interaction, student-student interaction, students' treatment by the teacher, competition between students, and classroom disorder. Classroom environment research has shown that the first two of these elements are important components of measuring classroom climate (for example, see Cazden 1986; Den Brok et al. 2004; Harjunen 2012). However, according to the dynamic model, what should be examined are the types of interactions that exist in a classroom, rather than how students perceive their teacher's interpersonal behavior. Specifically, the dynamic model is concerned with the immediate impact that teacher initiatives have on establishing relevant interactions in the classroom, and it investigates the

extent to which teachers are able to establish on-task behavior through promotion of interactions. The other three elements refer to teachers' attempts to create an efficient and supportive environment for learning in the classroom (Walberg 1986). These elements are measured by taking into account the teacher's behavior in establishing rules, persuading students to respect and use the rules, and the teacher's ability to maintain them in order to create and sustain an effective learning environment in the classroom.

7. *Management of time*: According to the dynamic model, effective teachers are able to organize and manage the classroom environment as an efficient learning environment and thereby maximize engagement rates. Thus, it is taken into account how much time is used for teaching per lesson and how much time is covered within the time framework. Additionally, the main interest of this factor is whether students are on task or off task and whether the teacher is able to deal effectively with any kind of classroom disorder without wasting the teaching time. It is also important to examine time attribution to different phases of the lesson according to their significance and the allocation of time for different groups of students. Therefore, management of time is considered an important indicator of teacher ability to manage the classroom effectively.

8. *Assessment*: Assessment is seen as an integral part of teaching (Stenmark 1992), and formative assessment, in particular, has been shown to be one of the most important factors associated with effectiveness at all levels, especially at the classroom level (e.g., De Jong et al. 2004; Kyriakides 2008; Shepard 1989). Therefore, information gathered from assessment is expected to be used to enable teachers to identify their students' needs, as well as to evaluate their own practice. In addition to the quality of the data emerging from teacher assessment (i.e., whether they are reliable and valid), the dynamic model is also concerned with the extent to which the formative rather than the summative purpose of assessment is achieved. The factor also refers to teacher skills in relation to each of the main phases of the assessment process (planning/construction of tools, assessment administration, recording, and reporting) and, at the same time, highlights the dynamic relationship between these phases (Black and Wiliam 2009).

The dynamic model is based on the assumption that, although there are eight different teacher effectiveness factors, each factor can be defined and measured by using five dimensions: frequency, focus, stage, quality, and differentiation. Frequency is a quantitative means of measuring the functioning of each effectiveness factor; most effectiveness studies to date have focused on this dimension only. The other four dimensions examine the qualitative characteristics of the functioning of the factors and help to describe the complex nature of effective teaching. A brief description of the four qualitative dimensions follows. The focus dimension takes into account both the specificity of the activities associated with the functioning of the factor and the number of purposes for which an activity takes place. The stage at which tasks associated with a factor take place is also examined; 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. Quality refers to properties of the specific factor itself, as they are discussed in the literature. Differentiation refers to the extent to which activities associated with a factor are implemented in the same way for all the subjects involved with it (e.g., all the

students, teachers, and 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 will ultimately maximize its effect on student learning outcomes (For further information on the conceptual background of the teacher factors of the dynamic model and the five measurement dimensions, see Creemers and Kyriakides 2008).

Although the framework underpinning the dynamic model is more complex than other models of effectiveness, it is based upon research evidence (Heck and Moriyama 2010; Hofman et al. 2010; Sammons 2009) and is empirically validated by three national studies testing the effects of classroom level factors upon student achievement on both cognitive and affective outcomes (Creemers and Kyriakides 2010; Kyriakides and Creemers 2008, 2009). However, since the factors included in the dynamic model are considered generic in nature, international studies are required to test the assumptions. This paper describes the results of a cross-national European study that helps us identify the extent to which the teacher factors contained in the dynamic model relate to student achievement in two different learning outcomes—mathematics and science—across six countries.

4 Methods

This cross-national European study investigated the effects of the teacher factors of the dynamic model of educational effectiveness on student outcomes in mathematics and science by collecting data from six countries (i.e., Belgium/Flanders, Cyprus, Germany, Greece, Ireland, and Slovenia). In each participating country, a sample of at least 50 primary schools was drawn ($n=334$), and written tests in mathematics and science were administered to all grade 4 students ($n=10,742$) at the beginning and at the end of school year 2010–2011. For the construction of the tests, permission was obtained from the International Association for the Evaluation of Educational Achievement (IEA) to use the released items of TIMSS 2007. The properties of each item and the relation with the curricula of grades 3 and 4 in each country were taken into account for developing four parallel types of test in each subject. Test-equating approaches were used to generate student scores in each administration period. During the school year 2010–2011, data were also collected on student-level and teacher, or classroom, factors.

Regarding student-level factors, four factors are examined: prior knowledge, socio-economic status, ethnicity, and gender. These factors have been found to explain most of the variance at student level (Scheerens and Bosker 1997; Teddlie and Reynolds 2000) and can be used to analyze differential effects of classroom factors (Strand 2010). Regarding classroom factors, all grade 4 students were asked to complete a questionnaire concerned with the behavior of their teacher in the classroom in regard to the eight factors of the dynamic model. For the development of the student questionnaire, we used an adapted version of the instrument that was developed for the studies mentioned above (i.e., Kyriakides and Creemers 2008; Creemers and Kyriakides 2010), which covered all eight factors and their dimensions. Specifically, students were asked to indicate the extent to which their teacher behaves in a certain way in their classroom, and a five-point Likert scale was used to collect the data. For example, an item concerned with the stage dimension of the structuring factor asked students to indicate whether at the beginning of the lesson the teacher explains how the new lesson relates

to previous ones; another item asked whether at the end of each lesson the teacher spends time reviewing the main ideas of the lesson. Another example is the item that was used to measure the differentiation dimension of the application factor: "The teacher of mathematics assigns to some students different exercises than to the rest of the students."

The original instrument was considered by the members of each country team, who were asked to express their views on the applicability and relevance of each item to their educational context and to assess whether the questions could be answered by young students in primary schools in their country. This process resulted in a substantial number of items being dropped from the original questionnaire. Specifically, we had to drop items of the questionnaire which referred to teaching materials and/or strategies that are used in the country where the original study took place but are not necessarily used in the other five countries. These items had to be removed in order to ensure that the questionnaire would be relevant to each context. Additionally, some of the items measuring the differentiation dimension had to be removed. This can probably be explained by the fact that differentiation is not used or interpreted the same way in all the countries. Specifically, in some countries, the differentiation items were perceived as inconsistencies to teaching behavior, instead of being realized as a response to the different educational needs of students. For example, in some countries, it was considered positive for the teacher to provide more time to certain students to complete their assessment assignment (e.g., slow learners), whereas in countries with a more centralized system, teachers may not be entitled to differentiate their assessment practice. Consequently, while the items of the new revised instrument were able to measure all eight factors, they did not account for all five measurement dimensions of each factor. In order to accommodate this shortfall, we classified the items of each factor into two broad categories concerned with the quantitative and qualitative characteristics of the functioning of each factor. The frequency and stage dimensions were treated as indicators of the quantitative characteristics of each factor, while the focus, quality, and differentiation dimensions were seen as indicators of the qualitative characteristics of the factors. The working version of the questionnaire was in English; the next step was to translate and back translate the questionnaire into four versions, i.e., Dutch, German, Greek, and Slovenian.

A generalizability study on the use of students' ratings was initially conducted. The results of the ANOVA analysis (see Kyriakides, Creemers, Panayiotou 2012) showed that the data can be generalized at the classroom level as, for all the questionnaire items, the between-group variance was higher than the within-group variance ($p < 0.05$).

5 Results

This section is subdivided into three parts: part A discusses the construct validity of the student questionnaire measuring the quality of teaching, which was tested using structural equation modeling (SEM). Part B examines the assumption of the dynamic model that teacher factors are related to each other, both across and within countries, and outlines the results of two sets of SEM analyses. Part C describes the results of multilevel modeling that was performed to explore the impact of teacher factors on student achievement gains in mathematics and science.

5.1 The construct validity of the student questionnaire

For the identification of the factor structure of the student questionnaire, SEM analyses were conducted using EQS software (Bentler 1995). Specifically, confirmatory factor analysis (CFA) was conducted for each teacher factor of the dynamic model using the EQS (Bryne 1994). CFA was appropriate as the objective was to test whether the data fit a hypothesized measurement model (in this case, the assumptions of the dynamic model in regard to quantitative and qualitative dimensions of each teacher factor). Each CFA model was estimated using maximum likelihood methods (ML). The ML estimation procedure was chosen because it does not require an excessively large sample size. Several fit indices were used to evaluate the extent to which the data fit the models tested, namely, the scaled chi-square, Bentler's (1990) comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA) (Brown and Mels 1990). Finally, the factor parameter estimates for the models with acceptable fit were examined to help interpret the models. The results of the CFA models provided support to the construct validity of the questionnaire. Although the scaled chi-square was statistically significant, the RMSEA values were less than 0.05 and the CFI values were above 0.95, thereby meeting the criteria for acceptable level of fit. Moreover, the standardized factor loadings were all positive and moderately high, with standardized values ranging from 0.48 to 0.84 and the majority higher than 0.65.

However, for measuring the time management factor, the dynamic model only takes into account the frequency dimension. Thus, for testing the validity of the questionnaire measuring this factor, CFA was not conducted, as there were only three items measuring the frequency dimension and the one-factor model is just identified (i.e., degrees of freedom=0). In the case of the management of time factor, exploratory factor analysis was therefore conducted, with satisfactory results. Specifically, the first eigenvalue was equal to 1.40 and explained almost 50 % of the total variance, while the second eigenvalue was less than 1 (i.e., 0.81). These results show that we can treat the three items as belonging to one factor, especially since all three items had relatively big loadings (i.e., greater than 0.67).

To test the construct validity of the questionnaire, within-country CFA analyses were also conducted. However, 9 out of 49 questionnaire items had to be removed in order to keep items with relatively high factor loadings when conducting both the across and the within-country analyses. Specifically, most of the items (n=4) measuring the differentiation dimension of the eight factors were removed. Similarly, most of the negative items (n=5) were removed. Finally, the items concerned with the classroom as a learning environment factor were found to belong to two different one-factor models measuring the types of interactions that exist in the classroom and the teacher ability to deal with student misbehavior (for more information about the CFA models that emerged from across- and within-country analyses, go to www.ucy.ac.cy/esf).

5.2 Searching for grouping of factors: A model describing quality of teaching

Since one of the main assumptions of the dynamic model is that the teacher factors are interrelated (see Kyriakides, Creemers and Antoniou 2009), the next step of the data analysis was to see how these effectiveness factors are related. Our assumption was that

the factors concerned with the following: (a) management of time, (b) teacher ability to deal with student misbehavior, and (c) the quantitative dimension of the questioning factor (measuring the extent to which teachers raise appropriate questions and avoid loss of teaching time) belong to one second-order factor, while the other factors can be grouped to another second-order factor. This assumption was initially tested by
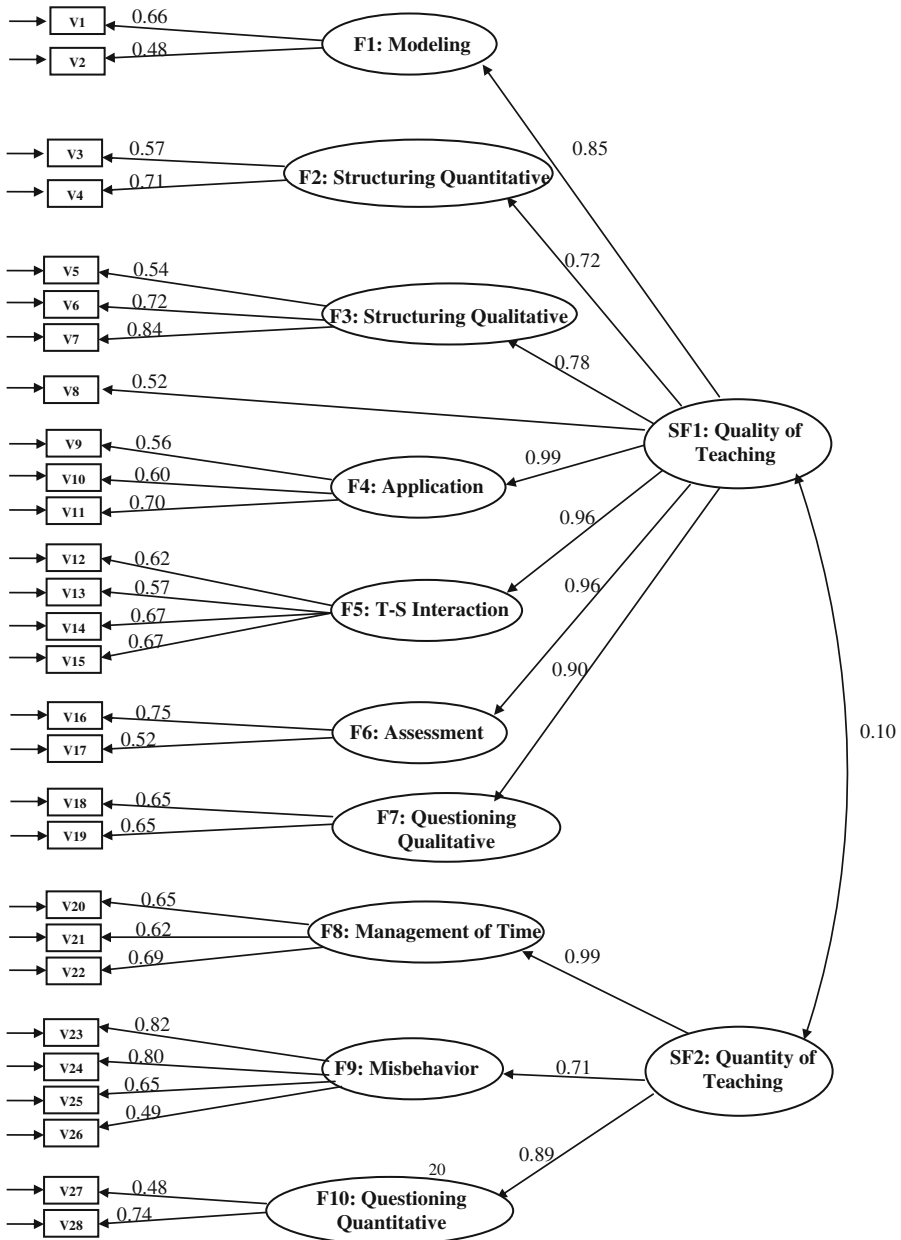

Fig. 1 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates

Table 1 Fit indices of the models used to test the factorial structure of the instrument emerged from the across- and within-country analyses

Models	X^2	df	χ^2/df	p	CFI	RMSEA
(A) Whole sample(N=9,967)						
Model 1	3,604	325	11.1	0.001	0.929	0.032
Model 2	16,507	350	47.1	0.001	0.648	0.068
Model 3	6,502	349	18.3	0.001	0.866	0.042
(B) Belgium(N=1,908)						
Model 1	731	297	2.4	0.01	0.929	0.028
Model 2	2,668	324	8.3	0.001	0.616	0.061
Model 3	1,395	323	4.3	0.001	0.824	0.042
(C) Cyprus (N=1,881)						
Model 1	825	317	2.6	0.01	0.943	0.029
Model 2	3,441	350	9.8	0.001	0.652	0.069
Model 3	1,584	349	4.3	0.001	0.861	0.043
(D) Greece(N=905)						
Model 1	560	312	1.8	0.01	0.944	0.030
Model 2	2,386	350	6.8	0.001	0.542	0.080
Model 3	1,285	349	3.7	0.001	0.789	0.054
(E) Ireland (N=2,140)						
Model 1	915	327	2.8	0.01	0.929	0.029
Model 2	2,416	350	6.9	0.001	0.752	0.053
Model 3	1,416	349	4.1	0.001	0.872	0.038
(F) Slovenia (N=2,049)						
Model 1	1,158	281	4.1	0.01	0.926	0.039
Model 2	4,573	324	14.1	0.001	0.640	0.080
Model 3	2,196	323	6.8	0.001	0.841	0.053
(G) Germany (N=1,072)						
Model 1	547	219	2.5	0.01	0.959	0.037
Model 2	3,472	275	12.6	0.001	0.599	0.104
Model 3	1,434	274	5.2	0.001	0.855	0.063

conducting across-countries SEM analysis. Our aim was to develop a model based on the data from all the countries and then to replicate the model by conducting relevant within-country analyses. Figure 1 illustrates the model with the two second-order factors. The fit statistics (scaled $X^2=3604$, $df=325$, $p<0.001$; RMSEA=0.032 and CFI=0.929) were acceptable. We also observe that the majority of the standardized path coefficients relating the first-order factors to the second-order factors were higher than 0.70. Thus, one second-order factor consists of three factors, i.e., management of time, teacher ability to deal with student misbehavior, and quantitative characteristics of the questioning factor. This second-order, or overarching, factor can be interpreted as an indication of the ability of teachers to maximize the use of teaching time (i.e., quantity of

teaching). The other seven factors were found to load on the other second-order factor, which can be interpreted as an indicator of the qualitative use of teaching time. The correlation coefficient between these two second-order factors is very small (see Fig. 1), implying that those teachers who are able to maximize the use of teaching time are not necessarily able to use the teaching time effectively.

Kline (1998, p. 212) argues that even when the theory is precise about the number of factors on a first- or second-order model, the researcher should determine whether the fit of a simpler model is better. Thus, we tested two alternative models to compare their fit to the data with the proposed model. In the first of these alternative models (model 2), all the items that were used for the SEM analysis were considered as belonging to a single first-order factor. This model was an attempt to see if the questionnaire items refer to a social desirability factor, and thereby, the questionnaire may not provide valid data. In the second alternative model (model 3), the 19 items concerned with the factors of the dynamic model measuring quality of teaching were considered as belonging to a single first-order factor, while the items concerned with the three factors of the dynamic model measuring quantity of teaching were expected to load on another first-order factor. If model 3 was found to fit to the data, this might cause doubts about the feasibility of including scores for each teacher factor in the dynamic model separately. However, the results show that model 1 provides the best fit to the data, and only the fit indices of model 1 can be considered satisfactory (Table 1).

Finally, we conducted six separate within-country SEM analyses and ran the three models on the data for each country. The results showed that the two second-order factor model (i.e., the theoretical model) fits well to the data for each country separately, whereas the two alternative models do not meet any requirements (Table 1). The fit indices of the within-country models indicate that, for each country, the two second-order factor model is the best fit. Moreover, all the within-country analyses revealed that the correlation between the two overarching factors is small. This implies that teachers who are effective in terms of their ability to maximize the use of teaching time may not also be effective in terms of using the teaching time appropriately.

5.3 Searching for the impact of teacher factors on student achievement

Separate multilevel modeling analyses for mathematics and science were conducted to identify the impact of teacher factors on student achievement in the two subjects. The results of these analyses are presented in Tables 2 and 3. The first step was to run a two-level model (classroom level and student level) with no explanatory variables (i.e., empty model) to determine the variance at each level. For both mathematics and science, the variance was found to be statistically significant at each level. Next, in model 1, two context variables concerned with students' prior achievement and the students' prior achievement at the classroom level were added to the empty model. Both variables had a statistically significant effect at level 0.05. The next step was to run different versions of model 2 (i.e., models 2a–i for mathematics, as shown in Table 2, and models 2a–g for science, as shown in Table 3)[1]. In each version of model 2, the first-order factor scores of

---

[1] The factors that did not have a statistically significant effect on achievement are not included on Tables 2 and 3.

Table 2 Parameter estimates and (standard errors) for the analysis of student achievement in mathematics (students within classes)

Teacher factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2 g	Model 2 h	Model 2i	Model 3
Fixed part (intercept)	330.3 (1.55)	39.2 (8.28)	27.1 (11.8)	21.6 (10.0)	28.6 (10.3)	12.2 (11.5)	14.9 (8.9)	−3.8 (12.4)	31.8 (8.4)	12.3 (12.7)	7.8 (11.6)	−7.4 (13.5)
Student level												
Context												
Prior achievement		0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)	0.7 (0.01)
School level												
Context												
Prior achievement		0.30 (0.03)	0.30 (0.03)	0.30 (0.03)	0.30 (0.03)	0.30 (0.03)	0.24 (0.03)	0.30 (0.03)	0.25 (0.03)	0.30 (0.03)	0.30 (0.03)	0.22 (0.03)
Modeling			5.8 (3.8)									
Structuring quantity				8.4 (2.9)								
Structuring quality					4.5 (2.9)							
Application						11.8 (3.7)						
Management time							19.2 (2.9)					
T-S interactions								18.4 (4.1)				
Misbehavior									10.3 (2.4)			
Questioning quality										10.7 (3.7)		
Assessment											12.1 (3.2)	
Overarching quality												13.7 (5.3)
Overarching quantity												21.1 (4.4)
Variance components												
Class	24.4 %	6.3 %	5.6 %	5.6 %	5.8 %	5.6 %	5.1 %	5.5 %	5.4 %	5.5 %	5.6 %	4.7 %
Student	75.6 %	47.1 %	46.7 %	46.8 %	46.8 %	46.8 %	46.7 %	46.8 %	46.8 %	46.8 %	46.8 %	46.7 %
Explained		46.6 %	47.7 %	47.6 %	47.4 %	47.6 %	48.2 %	47.7 %	47.8 %	47.7 %	47.6 %	48.6 %

Table 2 (continued)

Teacher factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2 g	Model 2 h	Model 2i	Model 3
Significance test												
Log-likelihood	103422	98606	96963	96963	96969	96961	96759	96951	96781	96791	96958	96765
Reduction		4816	1643	1643	1637	1645	1847	1655	1825	1815	1648	1841
Degrees of freedom		2	1	1	1	1	1	1	1	1	1	2
p value		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

For each alternative model 2 (i.e., models 2a up to 2i) and for model 3, the reduction is estimated in relation to the deviance of model 1

Table 3 Parameter estimates and (standard errors) for the analysis of student achievement in science (students within classes)

Teacher factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2 g	Model 3
Fixed part (intercept)	317.3 (1.6)	49.5 (8.59)	33.4 (10.24)	19.1 (12.2)	35.7 (9.5)	2.3 (12.8)	45.4 (8.9)	35.3 (13.8)	20.6 (12.2)	9.29 (14.7)
Student level										
Context										
Prior achievement		0.54 (0.01)	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)
School Level										
Context										
Prior achievement		0.36 (0.03)	0.35 (0.03)	0.35 (0.03)	0.32 (0.03)	0.32 (0.03)	0.33 (0.03)	0.35 (0.03)	0.35 (0.03)	0.31 (0.03)
Structuring quant.			8.8 (3.2)							
Application				13.7 (4.1)						
Management time					11.7 (3.5)					
T-S interactions						21.9 (4.6)				
Misbehavior							5.1 (2.6)			
Questioning qual.								5.8 (4.1)		
Assessment									11.8 (3.6)	
Overarching quality										14.5 (5.9)
Overarching quantity										12.9 (5.0)
Variance components (%)										
Class	31.4 %	9.8 %	9.1 %	9.0 %	8.9 %	8.7 %	9.1 %	9.1 %	9.0 %	8.1 %
Student	68.6 %	48.4 %	48.0 %	48.0 %	48.0 %	48.0 %	48.0 %	48.0 %	48.0 %	48.0 %
Explained		41.8 %	42.9 %	43.0 %	43.1 %	43.3 %	42.9 %	42.9 %	43.0 %	43.9 %

Table 3 (continued)

Teacher factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2 g	Model 3
Significance test										
Log-likelihood	99531	95955	94344	94341	94201	94330	94209	94211	94341	94198
Reduction		3576	1611	1614	1754	1625	1746	1744	1614	1757
Degrees of freedom		2	1	1	1	1	1	1	1	2
p value		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

For each alternative model 2 (i.e., models 2a up to 2g) and for model 3, the reduction is estimated in relation to the deviance of model 1

the SEM models, which refer to the teacher factors of the dynamic model, were added one by one to model 1. The fit of each model was tested against model 1. The likelihood statistic (X^2) shows a significant change between model 1 and each version of model 2 ($p<0.001$), which implies that variables measuring the teacher factors have significant effects on student achievement in mathematics and science. As can be seen in Table 2, in mathematics, most of the first-order factors have a statistically significant effect on achievement, with the exception of the factor concerned with the quantitative characteristics of questioning. Table 3 shows that most of the teacher factors were found to have a statistically significant effect on student achievement in science, with the exception of three factors concerned with the following: (a) modeling, (b) the quantitative characteristics of questioning, and (c) qualitative characteristics of structuring.

It was not possible to create a model containing all the first-order teacher factors of the quality and/or the quantity factors because the first-order teacher factors are highly correlated with each other. Therefore, to establish model 3, the two second-order, or overarching, factors (i.e., quality of teaching and quantity of teaching) were added to model 1 to test their impact on student achievement. Both overarching factors were found to have statistically significant effects on student achievement in each subject. The fit of model 3 was tested against model 1, and the likelihood statistic (X^2) shows a significant change between the two models ($p<0.001$). The likelihood statistic also shows that model 3 fits the data better than any of the models 2 (both in mathematics and science) where only one teacher factor is added. Model 3 explains approximately 50 % of the total variance of student achievement in mathematics and approximately 45 % of the total variance of student achievement in science.

6 Discussion

This paper contributes to existing knowledge on educational effectiveness by examining the effects of teacher behavior in the classroom on student achievement in mathematics and science. The results of the study presented in this paper demonstrate that the eight teacher factors included in the dynamic model of educational effectiveness have statistically significant effects on student achievement gains in all the participating countries. This implies that these eight factors can be considered as generic factors and are therefore relevant for policy making in each country. Thus, five key points are derived from the findings that have implications not only for the development of the theoretical framework of EER but may also help in policy formation on quality of teaching and, thus, contribute to teacher professional development.

First, the results of the cross-national European study show that primary students in grade 4 are capable of providing valid data on their teachers' behavior in the classroom in relation to the teacher factors included in the dynamic model. These results can be attributed to the fact that the eight teacher factors refer to observable behavior and to teaching actions that young students are capable of identifying and offering their opinions about whether and to what degree they take place. The questionnaire does not refer to inferences about the quality of their teacher in an abstract way; instead, students are asked to report on whether or not concrete actions take place in their classroom. For example, students were asked to indicate whether their teacher provides feedback when an answer is given by students and whether the lessons start and/or finish

on time. Since students are capable of observing and reporting on the behavior of their teachers in the classroom for a long period of time, the data they provide are likely to be reliable and valid. In addition, the questionnaire items are not concerned with the knowledge or training level of their teachers or with personality traits that would require some special knowledge and skills to evaluate. As mentioned in the introduction of the paper, the dynamic model is concerned with observable behavior of teachers based on many studies and meta-analyses that show teacher behavior in the classroom is more strongly associated with student achievement than any other teacher characteristic (e.g., Kyriakides 2008; Kyriakides, Creemers, Antoniou and Demetriou 2010; Seidel and Shavelson 2007).

Second, the SEM analyses (across-countries and within-country) revealed that effective teaching can be defined through two overarching factors that are very weakly related to each other. This weak relationship indicates that the teachers who are good at maximizing the use of teaching time are not necessarily good at making qualitative use of teaching time. Thus, overall effective teaching should be defined according to the teachers' skills in both areas of quality and quantity. The multilevel analyses also reveal the importance of quantity of teaching. Implications can be drawn for the development of national and/or school policy for effective teaching.

Third, the study highlights the need for further research to be conducted on measuring the five dimensions of the dynamic model. The cross-national study was not designed to produce data about each measurement dimension of teacher factors. This is partly due to the fact that the original instrument had to be adapted for use with a broad range of students coming from different countries and different educational contexts. In addition, student responses to most items concerned with the "differentiation" dimension were not found to be comparable from country to country, and thereby, all had to be removed. This finding in itself is an indication that the concept of differentiation is not interpreted in the same way by young students of different countries. For example, some students may consider it unfair that the teacher responds differently to different groups of students in specific teaching situations, while others may perceive such behavior appropriate (e.g., teachers giving different assessment tasks in a test or giving different feedback to students with different learning needs). Further research, using different mechanisms of measuring teacher behavior and using well-trained observers, may generate valid and reliable data about all five measurement dimensions. The other methodological drawback was that the project was not designed to collect data through the use of observation instruments, as were used to test the effect of the factors of the dynamic model in the earlier national studies (Kyriakides and Creemers 2008; Kyriakides and Tsangaridou 2008).[2] Nevertheless, despite these limitations, the results that emerged from the cross-national effectiveness study provide support for the dynamic model and can be seen as a starting point for further international studies that may employ a variety of approaches when measuring effectiveness of teaching.

Fourth, the SEM analyses show that factors associated with the active and direct teaching approach (e.g., structuring and application) are closely related to factors that align with the constructivist approach to learning (e.g., orientation and modeling).

---

[2] In the national studies, each dimension of each factor was measured, and the added value of taking into account all five dimensions of teacher factors to explain student achievement was demonstrated (see Creemers and Kyriakides 2008).

Consequently, it was possible for us to establish a second-order factor measuring quality of teaching that incorporates factors from teaching approaches that are seen as competing with each other. This implies that teachers who perform well in "traditional" instruction are likely to perform well in factors associated with the constructivist approach. Moreover, the results of the multilevel analyses show that teacher factors coming from different teaching approaches are associated with student achievement gains in different learning outcomes. Therefore, it can be claimed that the cross-national effectiveness study generates support for attempting to use an integrated approach in defining quality of teaching. The study also provides some support for the generic nature of the eight factors since the data coming from different countries show that they are related with student achievement gains in two different learning outcomes—mathematics and science.

Finally, the study emphasizes the importance of comparative studies in EER for the development of its theoretical framework. It supports claims that effectiveness of teaching can be described in a similar way in different contexts and can explain achievement gains in different learning outcomes. The study was conducted in six different European countries, and one could suggest that the findings could be attributed to the underlying similarities of European cultures. In order to explore or refute this suggestion, further comparative studies collecting data from countries within and outside Europe are required to test the generic nature of teacher factors. These studies may provide evidence about the cultural impact on education and quality of teaching and thereby the factors and dimensions contained in the dynamic model. Such comparative studies may also provide evidence about the impact of teacher factors in different learning outcomes in subjects other than mathematics and science, in other domains (e.g., affective and psychomotor), and in metacognition. Not only would they test the generic nature of the teacher factors, but they would also contribute to the broader debate about the importance of factors associated with different teaching approaches. Finally, studies searching for the generic nature of teacher factors included in the dynamic model of educational effectiveness may help policy makers to develop policies on quality of teaching and contribute to teacher professional development.

References

Aparicio, J. J., & Moneo, M. R. (2005). Constructivism, the so-called semantic learning theories, and situated cognition versus the psychological learning theories. *Spanish Journal of Psychology, 8*(2), 180–198.

Askew, M., & William, D. (1995). *Recent research in mathematics education 5–16*. London: Office for Standards in Education.

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

Bentler, P. M. (1995). *EQS: structural equations program manual*. California: Multivariate Software Inc.

Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Borich, G. D. (1992). *Effective Teaching Methods* (2nd ed.). New York: Macmillan.

Brookover, W. B., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School systems and student achievement: schools make a difference*. New York: Praeger.

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.

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

Bryne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Newbury Park: Sage.

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 432–463). New York: Macmillan.

Coleman, J. S., Campbell, E. Q., Hobson, C. F., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington: U.S. Government Printing Office.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366.

Creemers, B. P. M., & Kyriakides, L. (2008). The Dynamics of Educational Effectiveness. A contribution to Policy, Practice and Theory in contemporary schools. London: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2010). Explaining stability and changes in school effectiveness by looking at changes in the functioning of school factors. *School Effectiveness and School Improvement, 21*(4), 409–427.

Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives, 8* (1). http://epaa.asu.edu/epaa/v8n1/. Accessed 14 Jan 2014.

De Corte, E. (2000). Marrying theory building and the improvement of school practice: a permanent challenge for instructional psychology. *Learning and Instruction, 10*(3), 249–266.

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

Den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement, 15*(3/4), 407–442.

Doyle, W. (1990). Classroom knowledge as a foundation for teaching. *Teachers College Record, 91*(3), 347–360.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–27.

Evertson, C. M., Anderson, C., Anderson, L., & Brophy, J. (1980). Relationships between classroom behaviour and student outcomes in junior high math and English classes. *American Educational Research Journal, 17*, 43–60.

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.

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

Grieve, A. M. (2010). Exploring the characteristics of "teachers for excellence": teachers' own perceptions. *European Journal of Teacher Education, 33*(3), 265–277.

Hallinger, P., & Heck, R. H. (2011). Conceptual and methodological issues in studying school leadership effects as a reciprocal process. *School Effectiveness and School Improvement, 22*(2), 149–173.

Harjunen, E. (2012). Patterns of control over the teaching–studying–learning process and classrooms as complex dynamic environments: a theoretical framework. *European Journal of Teacher Education, 34*(2), 139–161.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: a regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.

Hofman, R. H., Hofman, W. H., & Gray, J. M. (2010). Institutional contexts and international performances in schooling: comparing patterns and trends over time in international surveys. *European Journal of Education, 45*(1), 153–173.

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1–10.

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

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: a step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement, 19*(4), 429–446.

Kyriakides, L. & Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19*(2), 183–205.

Kyriakides, L., & Creemers, B. P. M. (2009). The effects of teacher factors on different outcomes: two studies testing the validity of the dynamic model. *Effective Education, 1*(1), 61–86.

Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*(1), 12–23.

Kyriakides, L., Creemers, B. P. M., & Panayiotou, A. (2012). Report of the Data Analysis of the Student Questionnaire Used to Measure Teacher Factors: Across and Within Country Results (ESF project: Establishing a knowledge base for quality in education: testing a dynamic theory for education 08-ECRP-012). Nicosia, Cyprus: University of Cyprus.

Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *British Educational Research Journal, 36*(5), 807–830.

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

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement, 21*(3), 315–336.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: the junior years*. Somerset: Open Books.

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*(3), 273–303.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist, 38*(1), 1–4.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.

Reynolds, D. (2000). School effectiveness: the international dimension. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 232–256). London: Falmer Press.

Reynolds, D. (2006). World class schools: some methodological and substantive findings and implications of the International School Effectiveness Research Project (ISERP). *Educational Research and Evaluation, 12*(6), 535–560.

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.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: secondary schools and their effects on children*. Cambridge: Harvard University Press.

Sammons, P. (2009). The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools. *School Effectiveness and School Improvement, 20*(1), 123–129.

Scheerens, J. (2013). *School leadership effects revisited: review and meta-analysis of empirical studies*. Dordrecht: Springer.

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

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis research. *Review of Educational Research, 77*, 454–499.

Shepard, L. A. (1989). Why we need better assessments. *Educational Leadership, 46*(2), 4–8.

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

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement, 21*(3), 289–314.

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

Walberg, H. J. (1986). What works in a nation still at risk. *Educational Leadership, 44*(1), 7–10.