



Using student ratings to measure quality of teaching in six European countries

Leonidas Kyriakides, Bert P.M. Creemers, Anastasia Panayiotou, Gudrun Vanlaar, Michael Pfeifer, Gašper Cankar & Léan McMahon

To cite this article: Leonidas Kyriakides, Bert P.M. Creemers, Anastasia Panayiotou, Gudrun Vanlaar, Michael Pfeifer, Gašper Cankar & Léan McMahon (2014) Using student ratings to measure quality of teaching in six European countries, *European Journal of Teacher Education*, 37:2, 125-143, DOI: [10.1080/02619768.2014.882311](https://doi.org/10.1080/02619768.2014.882311)

To link to this article: <http://dx.doi.org/10.1080/02619768.2014.882311>



Published online: 28 Jan 2014.



Submit your article to this journal [↗](#)



Article views: 257



View related articles [↗](#)



View Crossmark data [↗](#)

Using student ratings to measure quality of teaching in six European countries

Leonidas Kyriakides^{a*}, Bert P.M. Creemers^b, Anastasia Panayiotou^a, Gudrun Vanlaar^{c,d}, Michael Pfeifer^e, Gašper Cankar^f and Léan McMahon^g

^aDepartment of Education, University of Cyprus, Nicosia, Cyprus; ^bFaculty of Behavioural and Social Sciences, Department of Pedagogy & Educational Science, University of Groningen, Groningen, The Netherlands; ^cCenter for Education Policy Analysis, Stanford University, Stanford, CA, USA; ^dDepartment of Educational Sciences, Katholieke Universiteit Leuven, Leuven, Belgium; ^eInstitute for School Development Research (IFS), Technical University Dortmund, Dortmund, Germany; ^fNational Examinations Centre, Ljubljana, Slovenia; ^gEconomic and Social Research Institute (ESRI), Dublin, Ireland

This paper argues for the value of using student ratings to measure quality of teaching. An international study to test the validity of the dynamic model of educational effectiveness was conducted. At classroom level, the model consists of eight factors relating to teacher behaviour: orientation, structuring, questioning, teaching modelling, application, management of time, teacher role in making classroom a learning environment and assessment. In each participating country (i.e. Belgium/Flanders, Cyprus, Germany, Greece, Ireland and Slovenia), a sample of at least 50 primary schools was used and all grade 4 students ($n = 9967$) were asked to complete a questionnaire concerning the eight factors of the dynamic model. Structural equation modelling techniques were used to test the construct validity of the questionnaire. Both across- and within-country analyses revealed that student ratings are reliable and valid for measuring the functioning of the teacher factors of the dynamic model. Implications for teacher education are drawn.

Keywords: quality of teaching; evaluation of teaching; international study; educational effectiveness; structural equation modelling

Introduction

Student ratings of teacher performance are frequently used in higher education, although not without criticism. As direct recipients of the teaching–learning process, students are in a key position to provide information about teachers' behaviour in the classroom. Moreover, student ratings constitute a main source of information regarding the development of motivation in the classroom, opportunities for learning, degree of rapport and communication developed between teacher and student, and classroom equity (Carle 2009; Kyriakides 2005; Marsh 1987). Students are considered good sources of information about their instructors for the following reasons: they know their own situation well; they have closely and recently observed a number of teachers; they uniquely know how students think and feel and they directly benefit from good teaching (Creemers, Kyriakides, and Sammons 2010).

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

However, in spite of the widespread use of, and reliance on, student ratings in higher education, such ratings of teacher performance remain suspect as a means of evaluating instructional effectiveness. While it is generally accepted that there are several strong reasons for using student ratings to evaluate teachers (Lasagabaster and Sierra 2011), still little effort has gone into the development of principles and practices in relation to this source of data at the K-12/primary level. Moreover, most studies investigating the quality of student ratings have focused on identifying factors affecting students' ratings of the effectiveness of their teachers (Kyriakides and Creemers 2008). While such research is useful for investigating the validity of student ratings, very little emphasis has been given to the material content of the student questionnaires used to measure teacher effectiveness. Consequently, very few studies have examined the construct validity of the questionnaires or address the theoretical foundation upon which student questionnaires should be based. Furthermore, very few studies have used different methodological approaches to evaluate the various forms of reliability and validity of student ratings. By way of commendation, however, it should be acknowledged that several researchers examining the reliability of student ratings have attempted to investigate the stability of student ratings across time, across courses and across instructors (e.g. Carle 2009; Marsh 2007). Although there is a considerable body of literature questioning the reliability of student ratings, recent studies indicate just the opposite. The stability of student ratings from one year to the next resulted in relatively high correlations (>0.80) and the correlations between student ratings of the same instructors and courses ranged from 0.70 to 0.87.

However, irrespective of how reliable measurements may be, they lack utility if they are not designed to fulfil some desired purpose. As validity of student ratings has far-reaching implications for using student ratings to measure teachers' behaviour in the classroom, researchers should not only investigate the reliability of instruments used to measure teacher performance through student views, they must also pay heed to designing measurement instruments and investigating their construct validity.

In order to identify the extent to which student ratings are used in the field of educational effectiveness, we conducted a review of papers published in the following eight journals which have a special interest in educational effectiveness and/or quality of teaching: (a) *Teaching and Teacher Education*, (b) *European Journal of Teacher Education*, (c) *School Effectiveness and School Improvement*, (d) *Effective Education*, (e) *Oxford Review of Education*, (f) *British Educational Research Journal*, (g) *Journal of Research on Educational Effectiveness* and (h) *American Educational Research Journal*. In total, these journals contained reports on 450 effectiveness studies. However, we found that only 44 of these reports used student ratings to measure quality of teaching. These 44 studies took place in more than 20 countries: almost three-quarters of them (73%) were conducted during the last two decades and approximately two-thirds (65%) related to secondary education. This implies that there is a growing interest in using student ratings to measure quality of teaching, but researchers are reluctant to use data from younger students. With regard to the testing of the construct validity of the instruments used to measure the quality of teaching, only 36% of the studies systematically investigated the validity of the questionnaire by using either structural equation modelling techniques or models of item response theory. Finally, the majority of the studies (93%) that use

student ratings only use data from students in one country and do not compare data from different jurisdictions.

Research aims

This paper investigates the extent to which primary school students can provide valid data on quality of teaching, which can be used for conducting both national and international studies. Specifically, we used data from an international study that aimed to test the validity of the dynamic model of educational effectiveness (Creemers and Kyriakides 2008). The study used student ratings to measure the teacher factors of the dynamic model, which are concerned with teachers' behaviour in the classroom. It is important to note that the study drew on earlier studies testing the validity of the dynamic model and the use of student ratings to measure quality of teaching. A study conducted by Kyriakides and Creemers (2008) used measures from students in grade 5 and external observers to assess quality of teaching and concluded that student ratings provided valid and reliable measures of teacher factors. Moreover, student ratings were found to be highly correlated with data provided by external observers (see Creemers and Kyriakides 2010; Kyriakides and Creemers 2008). Another study conducted in Canada used primary school students' ratings to measure the eight teacher factors included in the dynamic model. Almost all items were found to be generalisable at the teacher level and student ratings also provided support for the construct validity of the questionnaire (Janosz, Archambault, and Kyriakides 2011). In this paper, we move a step further and investigate the extent to which younger students (9- and 10-year olds) from different European countries can provide valid data about the teacher factors included in the dynamic model. The next section outlines the main elements of the dynamic model and the teacher factors. Following that, the methods and main results of the study are presented and discussed.

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 teaching and learning situation is emphasised and the roles of the two main actors (i.e. teacher and student) are analysed. Above these two levels, the dynamic model also refers to school-level factors. It is expected that school-level factors influence the teaching and learning situation by developing and evaluating the school policy on teaching and the policy on creating a learning environment at the school. The system level refers to the influence of the educational system through more formal avenues, especially through the development and evaluation of educational policy at the national/regional level. The model also takes into account the fact that the teaching and learning situation is influenced by the wider educational context in which students, teachers and schools are expected to operate. Factors such as the 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, as well as in the opinion formation of various stakeholders about what constitutes effective teaching practice.

The teacher factors of the dynamic model

Based on the main findings of educational effectiveness research (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 which describe teachers' instructional role and are associated with student outcomes: *orientation, structuring, questioning, teaching modelling, application, management of time, teacher role in making classroom a learning environment* and *classroom assessment*. These eight factors comprise an integrated approach to defining quality of teaching, which refers to different teaching approaches, such as the direct and active teaching model and the constructivist approach. A short description of each teacher factor follows.

Orientation

It 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(s) for which an activity takes place in the lesson. It is anticipated that the orientation process makes tasks/lessons meaningful for students, which in turn may encourage their active participation in the classroom (e.g. De Corte 2000; Paris and Paris 2001).

Structuring

Rosenshine and Stevens (1986) point out that student achievement is maximised when teachers not only actively present materials, but also structure them by: (a) beginning with an overview and/or review of objectives; (b) outlining the content to be covered and signalling transitions between lesson parts; (c) calling attention to main ideas and (d) reviewing main ideas at the end. Summary reviews are also important, since they integrate and reinforce the learning of major points (Brophy and Good 1986). These structuring elements facilitate memorising of the information and also allow for its apprehension as an integrated whole with recognition of the relationships between parts. Moreover, achievement levels tend to be higher when information is presented with a degree of redundancy, particularly in the form of repeating and reviewing general views and key concepts. Finally, 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).

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 the following five elements. Firstly, 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), but it has been found that effective teachers ask more process questions (Askew and William 1995; Everson et al. 1980). Secondly, 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. Thirdly, question clarity is measured by investigating the extent to which students understand what is required of them, that is, what the teacher expects them to do/find out. Fourthly, another element of this factor is the appropriateness of the difficulty level of the question. Most questions should elicit correct answers and that 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). In addition, optimal question difficulty should vary with context, for example, basic skills instruction requires a great deal of drill and practise and thus requires frequent fast-paced review in which most questions are answered rapidly and correctly. However, when teaching complex cognitive content or trying to get students to generalise, evaluate or apply their learning, effective teachers may raise questions that few students can answer correctly. Finally, the way teachers deal with student responses to questions is investigated. Correct responses should be acknowledged as such because even if the respondent may know that the answer is correct, some other students in the classroom may not know. In responding to students' partially correct or incorrect answers, effective teachers acknowledge whatever part may be correct, and if they consider there is a good prospect of success, they may try to elicit an improved response (Rosenshine and Stevens 1986). Therefore, effective teachers are more likely than other teachers 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.

Teaching modelling

Although there is a long tradition in research on teaching higher order thinking skills and, especially problem-solving, these teaching and learning activities have received more attention during the last two decades due to the emphasis given through policy on the achievement of new educational goals. Thus, the teaching modelling factor is associated with findings of effectiveness studies revealing that effective teachers are likely to help pupils use strategies and/or develop their own strategies that can help them solve different types of problem (Grieve 2010; Kyriakides, Campbell, and Christofidou 2002). Consequently, it is more likely that students will develop skills that can help them to organise their own learning (e.g. self-regulation and active learning). In defining this factor, the dynamic model also addresses the properties of teaching modelling tasks and especially the role teachers are expected to play in order to help students use a strategy to solve problems. Teachers may either present a clear problem-solving strategy or they may invite students to explain how they would approach or solve a particular problem and then use that information to promote the idea of modelling. Recent research has suggested that the latter may encourage students to not only use, but also to develop their own problem-solving strategies (Aparicio and Moneo 2005; Gijbels et al. 2006).

Application

Effective teachers also 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 also 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 that of the lesson. It also examines whether the application tasks are used as starting points for the next step of teaching and learning.

The classroom as a learning environment

Five elements of ‘the classroom as a learning environment’ are taken into account: *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 elements are important aspects of measuring classroom climate (e.g. see Cazden 1986; Den Brok, Brekelmans, and Wubbels 2004; Harjunen 2012). However, according to the dynamic model, the types of interactions that exist in a classroom need to be examined rather than how students perceive their teacher’s interpersonal behaviour. Specifically, the dynamic model is concerned with the immediate impact teacher initiatives have on establishing relevant interactions and it investigates the extent to which teachers are able to establish on-task behaviour through promotion of such interactions. The other three elements refer to teachers’ attempts to create an efficient and supportive environment for learning in the classroom (Walberg 1986). These aspects of the classroom as a learning environment are measured by taking into account the teacher’s behaviour in establishing rules, persuading students to respect and use the rules and ensuring they are adhered to in order to create and sustain a learning environment in the classroom.

Management of time

According to the dynamic model, effective teachers are able to organise and manage the classroom as an efficient learning environment and thereby maximise engagement rates (Creemers and Reezigt 1996). Therefore, management of time is considered an important indicator of teacher ability to manage the classroom effectively.

Assessment

Assessment is seen as an integral part of teaching (Stenmark 1992) and, in particular formative assessment has been shown to be one of the most important factors associated with effectiveness at all levels, especially the classroom level (e.g. De Jong, Westerhof, and Kruiter 2004; Kyriakides 2005; Shepard 1989). Therefore, information gathered from assessment is expected to be used by teachers to identify their students’ needs, as well as to evaluate their own practice. Thus, 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.

Measurement dimensions

The dynamic model is based on the assumption that, although there are different effectiveness factors, each factor can be defined and measured in terms of five dimensions: *frequency, focus, stage, quality and differentiation*. *Frequency* is a quantitative means of measuring the functioning of each effectiveness factor, and

most effectiveness studies to date have only focused on this dimension. 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 these four dimensions follows. Two aspects of the *Focus* dimension are taken into account: the first refers to the specificity of the activities associated with the functioning of the factor; the second refers to the number of purposes for which an activity takes place. *Stage* reflects the need for factors to take place over a long period of time to ensure that they have a continuous direct or indirect effect on student learning; the stage refers to when factors take place. *Quality* refers to properties of the specific factor itself, as 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 students, teachers and schools involved with it. It is expected that adaptation to the specific needs of each subject or group of subjects is likely to increase the successful implementation of a factor and will ultimately maximise its effect on student learning outcomes (Kyriakides 2007).

Methods

The international study that informs this paper tests the validity of the dynamic model of educational effectiveness using data collected in six European 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 ($n = 326$) and data on quality of teaching were obtained by all grade 4 students ($n = 9967$). At the end of the school year, students were asked to complete a questionnaire concerned with the behaviour of their teacher in the classroom according to the eight factors of the dynamic model. The questionnaire was based on an original instrument that had been developed to test the dynamic model in the earlier studies mentioned above (i.e. Creemers and Kyriakides 2010; Kyriakides and Creemers 2008) and which examined the eight factors and their dimensions. Specifically, students were asked to indicate the extent to which their teacher behaved 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 asked students to indicate whether the teacher would explain at the beginning of a new lesson how the lesson would relate to previous ones; another item asked whether the teacher would spend some time at the end of each lesson reviewing the main ideas covered in the lesson. Similarly, the following item provides an example of how the differentiation dimension of the application factor was measured: 'The mathematics teacher assigns to some pupils different exercises than to the rest of the pupils'.

The original questionnaire instrument was discussed by the members of each country's research team. The members of the country teams were experts in the field of EER and they considered the applicability and relevance of each questionnaire item to their own country's teaching context and whether or not the items were appropriate for grade 4 students in their country. Following consultations among the research team members, a substantial number of items were dropped from the original questionnaire. Thus, while the items of the revised instrument measured all eight factors, the questionnaire did not cover all five measurement dimensions of each factor. As a result, the items of each factor were classified into two overarching categories, which were broadly concerned with the quantitative and the qualitative characteristics of the functioning of each factor. The quantitative category referred to

the frequency and stage dimensions, which were treated as indicators of the importance attached to each factor by the teacher; the qualitative category referred to the other three dimensions (i.e. focus, quality and differentiation). An English version of the revised questionnaire was developed, which covered all eight factors and the two broader categories measuring the quantitative and qualitative characteristics of each factor. This was then translated and back-translated into four other languages (i.e. Dutch, German, Greek and Slovenian).

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

Results

Using a unified approach to test validation (AERA, APA and NCME 1999; Kane 2001), this study provides construct-related evidence obtained from the student questionnaire to measure quality of teaching. The factor structure of the questionnaire was identified through SEM analyses using EQS software (Bentler 1995). Each model was estimated by using normal theory maximum likelihood methods (ML). Three separate fit indices were used to evaluate the extent to which the data fitted the tested models: 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 facilitate the interpretation of the models. The main results of the SEM analysis for testing the construct validity of the student questionnaire are presented in the first part of this section. In order to test the assumption of the dynamic model that teacher factors are inter-related, both across- and within-country SEM analyses were conducted. The results of these two types of analysis are presented in the second part of this section.

The construct validity of the student questionnaire

Confirmatory factor analysis (CFA), using EQS (Byrne 1994), was conducted for each teacher factor of the dynamic model to test whether the data fitted a hypothesised measurement model, that is, the assumptions of the dynamic model regarding the two broader measurement dimensions of each teacher factor. Two sets of CFA were conducted: across countries (i.e. using the full data-set) and within countries (i.e. separate analysis for each country). The results of the across-country CFA confirmed the construct validity of the questionnaire. Although the scaled chi-square was statistically significant, the values of RMSEA were smaller than 0.05 and the values of CFI were greater than 0.95, thus meeting the criteria for an acceptable level of fit. Moreover, the standardised factor loadings were all positive and moderately high, ranging from 0.48 to 0.84, with most of them higher than 0.65.

However, the dynamic model takes into account only the frequency dimension in measuring the management of time. CFA was not used to test the validity of the questionnaire measuring this factor, as there were only three items measuring the frequency dimension and the one-factor model was just identified (i.e. degrees of freedom = 0). Therefore, in the case of the time management factor, exploratory

factor analysis was conducted and provided satisfactory results. Specifically, the first eigenvalue was equal to 1.40 and explained almost 50% of the total variance, whereas the second eigenvalue was less than 1 (i.e. 0.81). These results showed that these three items could be treated as belonging to one factor, especially since they had relatively high loadings (i.e. >0.67).

The within-country CFA analyses were less straightforward. Nine of the 49 questionnaire items had to be removed in order to keep items with relatively high factor loadings. Specifically, four items measuring the differentiation dimension of the eight factors had to be removed, and five of the negative items had to be removed. Finally, the items concerned with the classroom as a learning environment were found to belong to two different one-factor models which measured the type of interactions in the classroom and teacher's ability to deal with student misbehaviour. (For more information about the CFA models that emerged from across- and within-country analyses, go to www.ucy.ac.cy/esf).

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 analysis of data was to examine how these effectiveness factors were related to each other. Our assumption was that the factors concerned with: (a) management of time, (b) teacher ability to deal with student misbehaviour and (c) the quantitative dimension of the questioning factor (measuring the extent to which teachers raise appropriate questions and avoid loss of teaching time) belonged to one second-order factor, whereas the other factors could be grouped together as another second-order factor. This assumption was initially tested by conducting across-country SEM analysis. A model containing the two second-order factors was developed, based on the data from all the countries, and then replicated by conducting relevant within-country analyses (see Figure 1). The fit statistics (scaled $\chi^2(325) = 3604$, $p < 0.001$; RMSEA = 0.032; CFI = 0.929) were acceptable. The figure shows that most of the standardised path coefficients relating the first-order factors to the second-order factors were higher than 0.70. The first second-order factor consisted of the three factors measuring time management, teacher ability to deal with student misbehaviour and the quantitative characteristics of the questioning factor, and could be treated as the factor measuring the ability of teachers to maximise the use of teaching time (i.e. quantity of teaching). All the other factors were found to load on the other second-order factor, which could be treated as an indicator of the qualitative use of teaching time. Figure 1 also shows that the correlation coefficient between these two overarching factors was small, implying that teachers who maximise the use of teaching time do not necessarily use the teaching time effectively.

Kline (1998, 212) argues that even when the theory is precise about the number of factors of a first- or second-order model, the researcher should determine whether the fit of a simpler model is comparable. Following this practice, two alternative models were tested to compare their fit with the data to that of the proposed model. In the first alternative model (Model 2), all the items that were used for the SEM analysis were considered as belonging to a single first-order factor. The aim of this model was to see if the questionnaire items referred to a social desirability factor and, therefore, the questionnaire might not produce valid data. In the second alternative model (Model 3), the 19 items measuring the quality of teaching factors of the

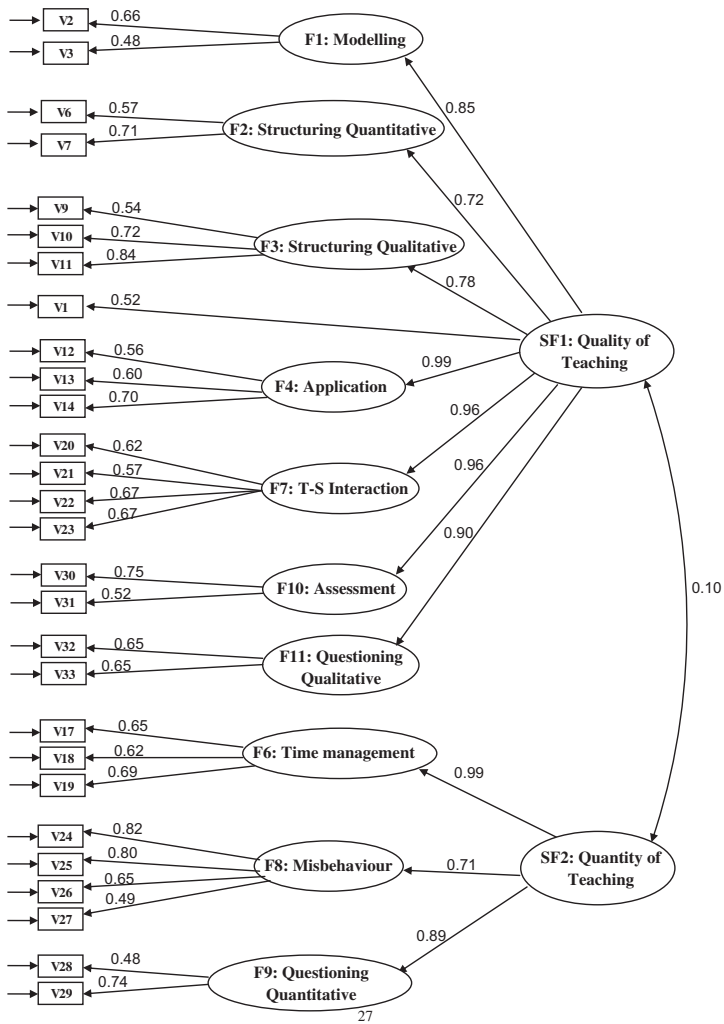


Figure 1. The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates.

Figure 1 presents the results from the across-country SEM analysis and shows the second-order factor model that fits the data of the student questionnaire best. Below you can see explanations for the first- and second-order factors that are included in the diagram.

First-order factors

- F1: Modelling
- F2: Structuring – Quantitative characteristics
- F3: Structuring – Qualitative characteristics
- F4: Application
- F6: Time management
- F7: Classroom as a learning environment – Qualitative characteristics: Teacher – Student interaction

F8: Classroom as a learning environment – Quantitative characteristics: Dealing with student misbehaviour

F9: Questioning – Quantitative characteristics: Raising non-appropriate questions

F10: Assessment

F11: Questioning – Qualitative characteristics

V1: Orientation

Second order factors

SF1: Quality of teaching

SF2: Quantity of teaching (Time management, Misbehaviour and Questioning quantitative: Raising non-appropriate questions)

dynamic model were considered as belonging to a single first-order factor, whereas the items measuring the three quantities of teaching factors of the dynamic model were considered to belong to another first-order factor. The thinking behind Model 3 was that if it was found to fit the data better, doubts might be cast as to whether individual scores for each teacher factor included in the dynamic model could be produced. The fit indices of each of the three models are shown in Table 1. It is clear that Model 1 best fits the data and is the only model where the fit indices can

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	χ^2	df	χ^2/df	<i>p</i>	CFI	RMSEA
A) Whole sample (<i>N</i> = 9967)						
Model 1	3604	325	11.1	0.001	0.929	0.032
Model 2	16,507	350	47.1	0.001	0.648	0.068
Model 3	6502	349	18.3	0.001	0.866	0.042
B) Belgium (<i>N</i> = 1908)						
Model 1	731	297	2.4	0.01	0.929	0.028
Model 2	2668	324	8.3	0.001	0.616	0.061
Model 3	1395	323	4.3	0.001	0.824	0.042
C) Cyprus (<i>N</i> = 1881)						
Model 1	825	317	2.6	0.01	0.943	0.029
Model 2	3441	350	9.8	0.001	0.652	0.069
Model 3	1584	349	4.3	0.001	0.861	0.043
D) Greece (<i>N</i> = 905)						
Model 1	560	312	1.8	0.01	0.944	0.030
Model 2	2386	350	6.8	0.001	0.542	0.080
Model 3	1285	349	3.7	0.001	0.789	0.054
E) Ireland (<i>N</i> = 2140)						
Model 1	915	327	2.8	0.01	0.929	0.029
Model 2	2416	350	6.9	0.001	0.752	0.053
Model 3	1416	349	4.1	0.001	0.872	0.038
F) Slovenia (<i>N</i> = 2049)						
Model 1	1158	281	4.1	0.01	0.926	0.039
Model 2	4573	324	14.1	0.001	0.640	0.080
Model 3	2196	323	6.8	0.001	0.841	0.053
G) Germany (<i>N</i> = 1072)						
Model 1	547	219	2.5	0.01	0.959	0.037
Model 2	3472	275	12.6	0.001	0.599	0.104
Model 3	1434	274	5.2	0.001	0.855	0.063

be considered satisfactory. Finally, six separate within-country SEM analyses were conducted. The results, also provided in Table 1, show that the second-order factor model (i.e. the theoretical model) best fits the data from each country separately, while neither of the two alternative models meets the requirements. Moreover, the within-country analysis revealed that most of the correlations between the two second-order factors are small, which suggests that teachers who are effective at maximising the use of teaching time may not be as effective at using the teaching time appropriately.

Discussion

The findings of this study present an opportunity to draw several implications for research on quality of teaching and teacher education. Firstly, the results of the SEM analyses provided support for the construct validity of the student questionnaire measuring teacher behaviour in the classroom. Student responses were also found to be generalisable. It can therefore be claimed that primary students in grade 4 are capable of providing valid data on the classroom behaviour of their teachers, based on the teacher factors included in the dynamic model. The fact that valid data about the teacher factors were obtained from grade 4 students could be attributed to the fact that the eight teacher factors in the model referred to observable behaviour and to teaching actions that were identifiable by young students. In other words, the questionnaire items did not refer to inferences about the quality of students' teachers in an abstract way, but students were expected to report on whether concrete actions took place in their classroom. For example, students were asked to indicate whether their teacher provided feedback when an answer was given and to indicate whether the lessons started and/or finished on time. Furthermore, students were asked to report on their teacher's behaviour at the end of the school year, which gave them ample opportunity to observe the classroom behaviour of their teachers over a relatively long period of time. Thus, due to the specificity of the questionnaire items and the fact that students had a lot of experience of how their teachers behaved in the classroom, the data they provided are, in all likelihood, reliable and valid. In addition, the questionnaire items were not concerned with student perception of teacher knowledge level or their teacher's personality traits that would require students to have some special knowledge or evaluation skills. As outlined in the first part of the paper, the dynamic model is only concerned with observable behaviour of teachers rather than any other variables that may explain their behaviour. This focus of the dynamic model on teachers' observable behaviour is based on the findings of many studies and meta-analyses that show that teacher behaviour in the classroom is more closely associated with student achievement than any other teacher characteristics (e.g. Kyriakides and Christoforou 2011; Seidel and Shavelson 2007). Therefore, teachers and other school stakeholders could use this questionnaire to collect data about quality of teacher behaviour in classrooms and develop school improvement projects to address factors found to be associated with student learning outcomes. At the same time, schools can draw on the theoretical framework presented here to develop their own policies on quality of teaching, especially since this study reveals that a significant percentage of teachers in each participating country did not perform at a high level on all factors, which suggests that there is ample scope for improvement (see Kyriakides, Creemers, and Panayiotou 2012).

Secondly, collecting information from all of the students in a class about the behaviour of their teacher allows researchers to test the generalisability of the data and identify the extent to which the object of measurement is the teacher. In this paper, we argue that at the very first stage of analysis of student ratings, researchers need to investigate the generalisability of such ratings, especially when single scores per teacher for each teaching skill are generated. In this respect, we advocate the use of generalisability theory in analysing student ratings. Our research of the literature indicated, however, that very few studies on student ratings have made use of generalisability theory. It is also important to note that when other sources of data are used to measure quality of teaching, it may not be possible to check data quality, since only one person rates each teacher. For example, by collecting data on teacher factors from an external observer or even from the teacher himself/herself (i.e. using self-ratings), the generalisability of the data cannot be easily determined and the identification of possible biases (either in favour of or against specific teachers) may not be possible. Nevertheless, it should be acknowledged that if we had additional resources to collect data from both students and external observers, we could have generated more precise, reliable and valid data on quality of teaching. Further international research could make use of such observation instruments as those that have been developed to collect data about the teacher factors in one of the participating countries. Data from these national studies provided support for the validity of the observation instruments and demonstrated the importance of using multiple sources to collect data on teaching quality (see Kyriakides and Creemers 2008).

Thirdly, the use of student ratings helped to classify the teacher factors of the dynamic model into two categories, that is, the quantity factors concerned with the teacher's ability to maximise the use of available teaching time, and the quality factors referring to the use of teaching time in an effective way. The results also suggest that teachers who maximise the use of teaching time may not necessarily be able to use the time effectively. By taking into account findings of EER and findings of studies demonstrating that teacher factors are related to student achievement (e.g. Creemers and Kyriakides 2008; Scheerens and Bosker 1997; Teddlie and Reynolds 2000), may be plausibly suggested that effective teachers are not only expected to manage their teaching time in an efficient way and deal with student misbehaviour in such a way as to ensure that students are kept on task, but should also be able to use teaching time effectively by providing specific opportunities and activities that promote learning, such as structuring, orientation, learning strategies and application tasks. One of the central findings of this study is that there is a weak correlation between the two overarching teacher factors, which suggests some teachers may be more effective in one overarching factor and less effective in the other. This finding of weak correlation between the two overarching factors has strong implications for teacher professional development. More specifically, it highlights the need for training courses to be concerned with both quantity and quality of teaching in order to help teachers improve their teaching skills (Antoniou and Kyriakides 2011; Creemers, Kyriakides, and Antoniou 2013). Teacher educators may support teachers by using the student questionnaire to identify teachers' professional needs and offer area-specific training courses that may be tailored to the professional needs of each teacher.

In addition, teacher educators can use the theoretical framework supporting the model on quality of teaching to focus their training courses on the proposed teacher factors and the five different dimensions. The fact that the teacher factors were

found to be interrelated implies that teacher professional development courses should not address each teaching skill in an isolated way, as has been proposed by the competency-based approach to teacher professional development (Brooks 2002; Last and Chown 1996; Robson 1998; Whitty and Willmott 1991). Rather, teacher professional development should adopt a more holistic training approach to assist teachers in ways that reflect their practical needs in the classroom, both in terms of quantity and quality of teaching. The argument for adopting a more holistic training approach is also supported by some recent national studies that demonstrated the added value of using the proposed approach to teacher professional development rather than the earlier competency-based approach (see Creemers, Kyriakides, and Antoniou 2013).

Fourthly, this international study was not designed to produce data for each measurement dimension of the teacher factors. This is partly due to the fact that a substantial number of items had to be removed from the original instrument in order to accommodate a broad range of students coming from different countries and ensure that grade 4 students can provide valid answers. In addition, student responses to most of the items concerned with the differentiation dimension were not comparable in all countries and thereby all of them had to be removed from the analysis. This finding in itself is an indication that the concept of differentiation is not interpreted in the same way by young students from different countries. For example, some students may consider it to be unfair when the teacher responds differently to different groups of students in specific teaching situations (e.g. giving different assessment tasks in a test or giving different types of feedback to students with different learning needs). Further research employing mixed-method approaches is required to find out how students understand concepts of equity, fairness and differences in teacher treatment, as well as the kinds of difficulties young students face in answering items measuring the differentiation dimension of teacher factors (Teddlie and Sammons 2010). Such research may help to develop the instrument further and would facilitate investigation of the extent to which the five dimensions of each factor can be measured. In addition, it is likely that students are not in a position to evaluate factors and dimensions of a more complex nature, thus the use of complementary data sources, such as external observation, should be considered.

Fifthly, the dynamic model of educational effectiveness adopts an integrated approach to defining quality of teaching. Some of the teacher factors are associated with the direct and active teaching approach (e.g. structuring, application), while others are in line with the constructivist approach to learning (e.g. orientation, modelling). Data emerging from student ratings suggest that factors associated with different teaching approaches tend to be closely related to each other, which implies that teachers who perform better than others on factors associated with the direct and active teaching approach tend also to perform better than others on factors associated with the new constructivist learning approach. The fact that the factors associated with different teaching approaches tend to be closely related to each other is in line with the results of recent meta-analyses of studies on effective teaching (Seidel and Shavelson 2007; Kyriakides and Christoforou 2011), which suggest that when it comes to effective teaching and the factors contributing to its effectiveness, imposing unnecessary dichotomies between different teaching approaches may be counterproductive. Instead, by being agnostic to the teaching approach pursued in instruction, and by considering what exactly the teacher and the students *do* during the lesson and *how they interact* – regardless of whether their actions and interac-

tions resonate more with one approach or another – may be more productive (Grossman and McDonald 2008).

Sixthly, the findings of this study can inform pre-service and in-service teacher education programmes. In particular, teacher educators, as well as those involved in professional development efforts, could enrich their programmes by engaging pre-service and in-service teachers in discussions regarding the importance of the teacher factors included in the theoretical framework of this study. More critically, however, they could also give prospective and practising teachers the opportunity to rehearse and practise these factors in their teaching. For pre-service teachers, such opportunities could be afforded in microteaching environments where, while working in a relatively ‘safe’ environment with fellow students and without having the pressure of the actual teaching conditions, novice teachers could experiment with incorporating different factors in their practice and receive specific and detailed feedback on their performance (Antoniou and Kyriakides 2013). For in-service teachers, such opportunities might arise when teachers are encouraged to plan lessons that are underpinned by considerations of such factors, enact these lesson plans, reflect on them and receive feedback on how they could further improve their practice by involving their students in measuring quality of teaching (Creemers, Kyriakides, and Antoniou 2013).

Finally, it is argued that one of the main implications of this study for teacher education has to do with the importance of using an instrument that was developed for collecting feedback from students. Feedback is seen as a powerful learning tool and teachers are aware of the fact that their students’ skills improve by being given feedback. Yet, there is no solid feedback instrument for teachers. How can teachers, especially in-service teachers, who are usually the only adult in the class, receive feedback on their own functioning? Indeed, in a daily classroom situation, students very rarely give feedback to their teachers about their teaching skills. Even if a student is dissatisfied with the quality of teaching, he/she may be reluctant to share his/her feelings with the teacher. Because of the importance of feedback, the need for a well-developed, evidence-based instrument is obvious. Such a questionnaire would enable students to give feedback in a more formal and less personal way, overcoming their reluctance for giving feedback to someone in ‘authority’. Obviously, teachers may learn from this feedback, and may be able to adapt their behaviour in class accordingly, making their teaching more effective. Since both teachers and students can benefit from such an instructive interaction, this study may contribute to teacher education by establishing the first steps in the development of a feedback instrument for teachers.

Acknowledgements

The research presented in this paper is part of a three-year project (2009–2012) entitled ‘Establishing a knowledge-base for quality in education: Testing a dynamic theory of educational effectiveness’, funded by the European Science Foundation (08-ECRP-012) and the Cyprus Research Promotion Foundation (Project Protocol Number: ΔΙΕΘΝΗ/ΕΣΦ/0308/01).

Notes on contributors

Leonidas Kyriakides is a professor of Educational Research and Evaluation at the University of Cyprus. His main research interests are in the area of educational effectiveness and especially in modelling effectiveness and using research for improving quality in education.

Bert P.M. Creemers is a professor emeritus in Educational Sciences at the University of Groningen, The Netherlands. His main interest is educational quality in terms of students learning outcomes at classroom, school and system level.

Anastasia Panayiotou is a PhD student at the Department of Education of the University of Cyprus. She has participated in several international projects and her main research interests are in the area of educational effectiveness research and teacher professional development.

Gudrun Vanlaar is a doctoral researcher investigating which educational practices are the most effective for high-risk students. She is interested in multilevel modelling and causal inference techniques.

Michael Pfeifer works as an assistant professor at the Institute for School Development Research (IFS) at the University of Dortmund, Germany in the areas of school pedagogy, school effectiveness research, research methods and international educational research.

Gašper Cankar is a researcher at National Examinations Centre in Slovenia, where his interests include measurement properties of external examinations, use of student background characteristics, value added measures and item response theory.

Léan McMahon's academic background is in Social Science. She has extensive experience of educational research. She has conducted studies on pupil achievement, school effectiveness, and educational provision at primary and post-primary levels.

References

- AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education). 1999. *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Antoniou, P., and L. Kyriakides. 2011. "The Impact of a Dynamic Approach to Professional Development on Teacher Instruction and Student Learning: Results from an Experimental Study." *School Effectiveness and School Improvement* 22 (3): 291–311.
- Antoniou, P., and L. Kyriakides. 2013. "A Dynamic Integrated Approach to Teacher Professional Development: Impact and Sustainability of the Effects on Improving Teacher Behaviour and Student Outcomes." *Teaching and Teacher Education* 29: 1–12.
- Aparicio, J. J., and M. R. Moneo. 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., and D. William. 1995. *Recent Research in Mathematics Education 5–16*. London: Office for Standards in Education, 53.
- Bentler, P. M. 1990. "Comparative Fit Indexes in Structural Models." *Psychological Bulletin* 107: 238–246.
- Bentler, P. M. 1995. *EQS: Structural Equations Program Manual*. Encino, CA: Multivariate Software.
- Borich, G. D. 1992. *Effective Teaching Methods*. 2nd ed. New York: Macmillan.
- Brooks, R. 2002. "The Individual and the Institutional: Balancing Professional Development Needs within Further Education." In *Professional Development and Institutional Needs*, edited by G. Trorey and C. Cullingford, 35–50. Hampshire: Ashgate.
- Brophy, J., and T. L. Good. 1986. "Teacher Behavior and Student Achievement." In *Handbook of Research on Teaching*, 3rd ed., edited by M. C. Wittrock, 328–375. New York: MacMillan.
- Brown, M. W., and G. Mels. 1990. *RAMONA PC: User Manual*. Pretoria: University of South Africa.
- Byrne, B. M. 1994. *Structural Equation Modeling with EQS and EQS/Windows: Basic Concepts, Applications, and Programming*. Thousand Oaks, CA: Sage.

- Carle, A. C. 2009. "Evaluating College Students' Evaluations of a Professor's Teaching Effectiveness across Time and Instruction Mode (Online vs. Face-to-face) Using a Multilevel Growth Modeling Approach." *Computers & Education* 53: 429–435.
- Cazden, C. B. 1986. "Classroom Discourse." In *Handbook of Research on Teaching*, edited by M. C. Wittrock, 432–463. New York: MacMillan.
- Creemers, B. P. M., and L. Kyriakides. 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., and L. Kyriakides. 2008. *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge.
- Creemers, B. P. M., and L. Kyriakides. 2010. "School Factors Explaining Achievement on Cognitive and Affective Outcomes: Establishing a Dynamic Model of Educational Effectiveness." *Scandinavian Journal of Educational Research* 54 (3): 263–294.
- Creemers, B. P. M., L. Kyriakides, and P. Antoniou. 2013. *Teacher Professional Development for Improving Quality of Teaching*. Dordrecht: Springer.
- Creemers, B. P. M., L. Kyriakides, and P. Sammons. 2010. *Methodological Advances in Educational Effectiveness Research*. London: Taylor & Francis.
- Creemers, B. P. M., and G. J. Reezigt. 1996. "School Level Conditions Affecting the Effectiveness of Instruction." *School Effectiveness and School Improvement* 7 (3): 197–228.
- 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/>.
- 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., K. J. Westerhof, and J. H. Kruiter. 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., M. Brekelmans, and T. Wubbels. 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.
- Evertson, C. M., C. Anderson, L. Anderson, and J. Brophy. 1980. "Relationships between Classroom Behaviors and Student Outcomes in Junior High Mathematics and English Classes." *American Educational Research Journal* 17: 43–60.
- Gijbels, D., G. Van de Watering, F. Dochy, and P. Van den Bossche. 2006. "New Learning Environments and Constructivism: The Students' Perspective." *Instructional Science* 34 (3): 213–226.
- Grieve, A. M. 2010. "Exploring the Characteristics of 'Teachers for Excellence': Teachers' Own Perceptions." *European Journal of Teacher Education* 33 (3): 265–277.
- Grossman, P., and M. McDonald. 2008. "Back to the Future: Directions for Research in Teaching and Teacher Education." *American Educational Research Journal* 45 (1): 184–205.
- 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* 35 (2): 139–161.
- Janosz, M., I. Archambault, and L. Kyriakides. 2011. "The Cross-cultural Validity of the Dynamic Model of Educational Effectiveness: A Canadian Study." Paper presented at the 24th International Congress for School Effectiveness and Improvement (ICSEI) 2011, Limassol, Cyprus, January 2011.
- Kane, M. T. 2001. "Current Concerns in Validity Theory." *Journal of Educational Measurement* 38 (4): 319–342.
- Kline, R. H. 1998. *Principles and Practice of Structural Equation Modelling*. London: Gilford Press.

- Kyriakides, L. 2005. "Extending the Comprehensive Model of Educational Effectiveness by an Empirical Investigation." *School Effectiveness and School Improvement* 16 (2): 103–152.
- Kyriakides, L. 2007. "Generic and Differentiated Models of Educational Effectiveness: Implications for the Improvement of Educational Practice." In *International Handbook of School Effectiveness and Improvement*, edited by T. Townsend, 41–56. Dordrecht: Springer.
- Kyriakides, L., R. J. Campbell, and E. Christofidou. 2002. "Generating Criteria for Measuring Teacher Effectiveness through a Self-evaluation Approach: A Complementary Way of Measuring Teacher Effectiveness." *School Effectiveness and School Improvement* 13 (3): 291–325.
- Kyriakides, L., and C. Christoforou. 2011. "A Synthesis of Studies Searching for Teacher Factors: Implications for Educational Effectiveness Theory." Paper presented at the American Educational Research Association (AERA) 2011 Conference, New Orleans, LA, April 2011.
- Kyriakides, L., and B. P. M. Creemers. 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., B. P. M. Creemers, and P. Antoniou. 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., B. P. M. Creemers, and A. Panayiotou. 2012. *Report of the Data Analysis of the Teacher Questionnaire Used to Measure School 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: University of Cyprus.
- Lasagabaster, D., and J. M. Sierra. 2011. "Classroom Observation: Desirable Conditions Established by Teachers." *European Journal of Teacher Education* 34 (4): 449–463.
- Last, J., and A. Chown. 1996. "Competence-based Approaches and Initial Teacher Training for FE." In *The Professional FE Teacher, Staff Development and Training in the Corporate College*, edited by J. Robson, 20–32. Aldershot: Avebury.
- Marsh, H. W. 1987. "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research." *International Journal of Educational Research* 11: 253–388.
- Marsh, H. W. 2007. "Do University Teachers Become More Effective with Experience? A Multilevel Growth Model of Students' Evaluations of Teaching over 13 Years." *Journal of Educational Psychology* 99: 775–790.
- Muijs, D., and D. Reynolds. 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.
- Paris, S. G., and A. H. Paris. 2001. "Classroom Applications of Research on Self-regulated Learning." *Educational Psychologist* 36 (2): 89–101.
- Robson, J. 1998. "A Profession in Crisis: Status, Culture and Identity in the Further Education College." *Journal of Vocational Education and Training* 50 (4): 585–607.
- Rosenshine, B., and R. Stevens. 1986. "Teaching Functions." In *Handbook of Research on Reaching*, 3rd ed., edited by M. C. Wittrock, 376–391. New York: Macmillan.
- Scheerens, J., and R. J. Bosker. 1997. *The Foundations of Educational Effectiveness*. Oxford: Pergamon.
- Seidel, T., and R. J. Shavelson. 2007. "Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-analysis Results." *Review of Educational Research* 77 (4): 454–499.
- Shepard, L. A. 1989. "Why We Need Better Assessment." *Educational Leadership* 46 (2): 4–8.
- Stenmark, J. K. 1992. *Mathematics Assessment: Myths, Models, Good Questions and Practical Suggestions*. Reston, VA: NCTM.
- Teddle, C., and D. Reynolds. 2000. *The International Handbook of School Effectiveness Research*. London: Falmer Press.

- Teddlie, C., and P. Sammons 2010. "Applications of Mixed Methods to the Field of Educational Effectiveness Research." In *Methodological Advances in Educational Effectiveness Research*, edited by B. P. M Creemers, L. Kyriakides, and P. Sammons, 115–152. London: Routledge, Taylor & Francis.
- Walberg, H. J. 1986. "Syntheses of Research on Teaching." In *Handbook of Research on Teaching*, 3rd ed., edited by M. C. Wittrock, 214–229. New York: Macmillan.
- Whitty, G., and E. Willmott. 1991. "Competence-based Teacher Education: Approaches and Issues." *Cambridge Journal of Education* 21 (3): 309–318.