



Using educational effectiveness research to promote quality of teaching: the contribution of the dynamic model

Leonidas Kyriakides¹ · Bert P. M. Creemers² · Anastasia Panayiotou¹

Accepted: 29 January 2018 / Published online: 6 February 2018
© FIZ Karlsruhe 2018

Abstract

The dynamic model of educational effectiveness refers to eight factors that describe teachers' instructional role. A multidimensional framework for measuring both quantitative and qualitative characteristics of teaching factors is also proposed. Empirical support for the validity of the model has been provided, which revealed that the teaching factors can be grouped into five stages situated in developmental order. In this study, for the first time, a qualitative methodology is used to provide an in-depth analysis of three video-lessons through the perspective of the five stages of effective teaching. Thus, we present how each stage is defined and use the cases of the three video-lessons to justify the rationale for these stages and help readers see how observational data are used to identify individual improvement priorities and provide differentiated feedback, even to teachers allocated to the same stage. Finally, based on the qualitative analysis of the three case-studies, strengths and limitations of the dynamic model to evaluate quality of teaching for formative reasons are identified.

Keywords Quality of teaching · Stages of effective teaching · Evaluation of teaching · Educational effectiveness research

1 The dynamic model: theoretical background

In this paper we analyse three video-lessons using a framework based on the dynamic model (Creemers and Kyriakides 2008) that refers to specific teaching factors and dimensions for the purpose of measuring their functioning. Previous studies investigating the validity of the model revealed that the teaching factors can be grouped into five stages of effective teaching. The aim of this paper is to analyze three lessons to detect strengths and weaknesses in using the proposed framework to measure quality of teaching in mathematics. In opposition to previous studies which were

mostly quantitative, this study uses for the first time a qualitative methodology to provide an in-depth analysis of the five stages of effective teaching. Thus, we attempt to use the analysis of the three video-lessons to justify the rationale for these stages. We also investigate whether teachers of the same stage need to focus on different aspects for improving their teaching practice. Thus, the first part of this section briefly presents the framework upon which the video analysis was based. We then provide a description of the teaching factors and their measurement dimensions.

1.1 The rationale for the dynamic model

During the last 35 years, researchers have turned to teacher behaviour as predictor of student achievement in order to build up a knowledge base on effective teaching. A series of process–product studies have thus taken place and led to the identification of a list of factors that link specific teaching behaviours and characteristics to student outcomes, leading Teacher Effectiveness Research (TER) to substantial development regarding its content (Muijs et al. 2014). As mentioned by Brophy and Good (1986), this list of teaching behaviours included quantity and pacing of instruction, classroom management, structuring and clarity of the lesson presentation, asking questions and providing feedback, and

✉ Leonidas Kyriakides
kyriakid@ucy.ac.cy
Bert P. M. Creemers
b.p.m.creemers@rug.nl
Anastasia Panayiotou
panayiotou.anastasia@ucy.ac.cy

¹ Department of Education, University of Cyprus, P. O. Box 20537, 1678 Nicosia, Cyprus

² Faculty of Behavioural and Social Sciences, Department of Pedagogy and Educational Science, University of Groningen, Groningen, The Netherlands

classroom climate. The dynamic model was developed taking into account the abovementioned results in the field of TER. Specifically, teaching factors that were shown to have an effect on student achievement, such as structuring of lessons, time management, questioning and application were included in the model. The dynamic model also took into account the criticism placed on the previous process–product studies conducted in the field of TER, regarding the exclusive emphasis given to cognitive student outcomes. The fact that most of the effectiveness studies are exclusively focused on language or mathematics, rather than on the aims of the whole school curriculum (cognitive, meta-cognitive and affective), reveals that the models of educational effectiveness should take into account the new goals of education. This means that the outcome measures should be defined in a broader way, rather than restricting these to the achievement of basic skills. It also implies that new theories of teaching and learning should be taken into account in order to specify variables associated with the quality of teaching. Thus, the dynamic model is based on traditional views on learning and instruction such as direct learning and teaching which emphasise not only the role of teacher as instructor responsible for providing knowledge and skills, but also the specific behaviours he/she should apply. Further, the model also takes into account new ideas on learning and instruction associated with constructivism, which give emphasis to independent learning and the construction of knowledge by the learner, including factors such as orientation and modelling (Brekelmans et al. 2000; Schoenfeld 1998).

Another essential difference of the dynamic model from previous models in the field of educational effectiveness research (EER) is that it explicitly refers to the measurement of each factor and assumes that these factors represent multi-dimensional constructs. In particular, the dynamic model proposes five measurement dimensions which are assumed to provide more information concerning not only the quantitative aspects of the factors (i.e., the frequency with which a factor appears), but also the qualitative aspects, which may thus contribute to the theoretical development of EER. These five dimensions, as well as their rationale, are presented in the second part of the next section.

Finally, the dynamic model assumes that the effectiveness factors are generic in nature, namely, that they are associated with student achievement in different learning domains. It is, however, acknowledged that their impact on different groups of students/teachers/schools may vary. Teachers are expected to adapt their teaching to correspond to the different needs of their students. Teacher's adaptive instructional behaviour makes them able to form their teaching in ways that provide equal opportunities to students with different background and personal characteristics. For example, studies investigating teacher differential effectiveness in relation to student socioeconomic status (SES) revealed that low-SES

students need more structure and more positive reinforcement (Campbell et al. 2004).

1.2 Teaching factors and their measurement dimensions

1.2.1 Teaching factors

Based on the main findings of TER, the dynamic model refers to eight factors that describe teachers' instructional role and have been consistently shown to be associated with student outcomes: *orientation, structuring, questioning, teaching-modelling, application, management of time, teacher role in making classroom a learning environment, and classroom assessment*. The model includes factors/teaching skills associated with direct teaching and mastery learning (Joyce et al. 2000), such as structuring and questioning, and with theories of teaching associated with constructivism (Brekelmans et al. 2000) including factors such as orientation and teaching modelling. Teachers' ability to promote collaboration among students is also taken into consideration. Therefore, an integrated approach to quality of teaching is adopted. A short description of each factor follows.

- a. *Orientation* This factor refers to teacher behaviour in providing the students with opportunities to identify the reason(s) for which an activity is presented or lesson or a series of lessons occur, and/or actively involving students in the identification of the reason(s) for which a lesson includes a specific task. Through this process it is expected that the activities that take place during lessons, become meaningful to students and consequently increase their motivation for participating actively in the classroom. This factor may thus have an impact on increasing student motivation and through that, on increasing student learning outcomes.
- b. *Structuring* Student learning is positively influenced when teachers actively present materials and structure them by: (a) beginning with overviews and/or review of objectives; (b) outlining the content to be covered and signalling transitions between lesson parts; (c) calling attention to main ideas; and (d) reviewing main ideas at the end (Rosenshine and Stevens 1986). Structuring activities aim at assisting students develop links between the different parts of lessons, instead of dealing with them in an isolated way (Creemers and Kyriakides 2015).
- c. *Questioning* This factor is concerned with teacher ability in: (a) raising different types of questions (i.e., process and product) at appropriate difficulty level; (b) giving time for students to respond; and (c) dealing with student responses. Raising numerous questions in a lesson increases the active involvement of students in class discussion and promotes interactions, both with the teacher

- and among students. Questioning can also be used in order to assess students' understanding and help them clarify and verbalize their thinking in order to develop a sense of mastery (Muijs et al. 2014).
- d. *Teaching-modelling* This factor is related to self-regulated learning (Muijs et al. 2014). Modelling is based on the assumption that effective teachers should encourage students to use or develop their own strategies in order to solve different types of problems.
 - e. *Application* Providing students with practice and application opportunities can enhance learning outcomes. Learning new information cannot be a constant process, since according to the Cognitive Load Theory the working memory can process only a limited amount of information at each given time (Kirschner 2002). Effective teachers may use seatwork or small-group tasks in order to provide necessary practice and application opportunities as starting points for the next step in teaching and learning.
 - f. *The classroom as a learning environment* This factor as described in the dynamic model consists of five components which were shown to be the most important aspects of the classroom climate through studies and meta-analyses: (a) teacher-student interaction, (b) student-student interaction, (c) students' treatment by the teacher, (d) competition between students, and (e) classroom disorder. The first two elements can be seen as important for measuring classroom climate, while the other three elements refer to teachers' efforts to create a well-organized and accommodating environment for learning in the classroom.
 - g. *Management of time* To address this factor the amount of time used per lesson for on-task behavior is investigated. Teachers are expected to do the following: (a) prioritize academic instruction and allocate available time to curriculum-related activities; and (b) maximize student engagement rates. Time management skills are not restricted solely to teachers' ability to avoid the loss of teaching time through minimizing external classroom disruptions, or through dealing effectively with organizational issues (e.g., moving between classes, organizing and distributing materials or giving instructions). Apart from the overall teaching time, management of time skills also include teacher actions that increase the learning time for each individual student (i.e., the on-task time).
 - h. *Assessment* Assessment is seen as an essential part of teaching (Stenmark 1992). Especially formative assessment has been shown to be one of the most important factors associated with effectiveness at all levels, especially at the classroom level (Christoforidou et al. 2014). Effective teachers are therefore expected to: (a) use appropriate techniques to collect data on student knowl-

edge and skills; (b) analyse data in order to identify student needs; (c) report assessment results to students and parents; and (d) evaluate their own practices.

1.2.2 Measurement dimensions

The model assumes that each factor can be defined and measured using the following five dimensions: *frequency*, *focus*, *stage*, *quality*, and *differentiation*. These dimensions help us better describe the functioning of a factor. Most effectiveness studies examined how frequently an activity related to a factor took place; therefore they took into account only the quantitative characteristics of a factor. Frequency is in line with this need as it comprises a quantitative way to measure the functioning of each factor. However, only examining the number of activities related with a factor is not sufficient to determine the quality of teaching offered, as the relation of some factors with student achievement may not be linear but curvilinear (Creemers and Kyriakides 2008). For example, providing students with opportunities to apply new knowledge was found to have a positive impact on their outcomes. However, spending too much teaching time on application activities may not allow sufficient time for teaching new content; which in turn may have a negative effect on student outcomes. Therefore, when measuring the functioning of a factor one should also take into consideration its qualitative characteristics. The other four dimensions included in the dynamic model examine the qualitative characteristics of the functioning of a factor. The dimensions are not only important from a measurement perspective, but also, and even more, from a theoretical point of view. The importance of taking each dimension into account is illustrated below by explaining how one of the factors included in the model (i.e., orientation) is defined. The frequency dimension of orientation is measured by taking into account the number of orientation activities that take place in a typical lesson, as well as how long each orientation activity takes place. These two indicators help us identify the importance that the teacher attached to this factor. In the case in which a factor did not appear in a lesson, the frequency dimension would be rated with a zero and consequently no other dimensions could be measured (i.e., quality, stage, focus and differentiation).

Two aspects of the *focus* dimension are measured. The first aspect addresses the purpose(s) for which an activity takes place (i.e., single or multiple purposes). The importance of measuring this aspect can be attributed to research findings showing that if all the activities are expected to achieve a single purpose, then the chances of achieving the purpose are high, but the effect of the factor might be small due to the fact that other purposes are not achieved and/or synergy may not exist since the activities are isolated (Schoenfeld 1998). On the other hand, if all the activities are expected to achieve multiple purposes, specific purposes

may not be addressed in a way that allows them to be implemented successfully (Pellegrino 2004). In other words, it is important to have both, activities that are focused on achieving one purpose and activities that aim at achieving multiple purposes. The appropriateness of an activity in terms of the number of purposes that the teacher aims to achieve depends on the content and goals of the lesson. Prior to conducting classroom observations, the observers hold a discussion with the teacher in order to identify the content/aims of the lesson and its relation to previous lessons (e.g., the amount of new information in a lesson). In the case of orientation, this aspect of focus is measured by examining the extent to which an activity is restricted to finding one single reason for doing a task or finding multiple reasons for doing a task. The second aspect of this dimension refers to the specificity of the activities, which can range from specific to general. The specificity of the orientation activities is measured by taking into account that an orientation activity may refer to a part of a lesson, the whole lesson or even a series of lessons (e.g., a lesson unit). A balance in terms of the specificity of the activities provided is observed in an effective classroom. For example, it is equally significant to discuss why learning to multiply is important (which is something more general), as well as why it is important to actively engage in an activity that is given to students in order to understand a single concept (more specific).

Activities associated with a factor can be measured by taking into account the *stage* at which they take place. It is supported that the factors need to take place over a long period of time to ensure that they have a continuous direct or indirect effect on student learning. It has been shown that the impact of a factor on student achievement partly depends on the extent to which activities associated with this factor are provided throughout the school career of the student. Although measuring the stage dimension gives information about the continuity of the existence of a factor, activities associated with the factor may not necessarily be the same. Therefore, using the stage dimension to measure the functioning of a factor can help us identify the extent to which there is constancy at each level and flexibility in using the factor during the period in which the investigation takes place. In the case of orientation, it is taken into account that orientation activities may take place in different parts of a lesson or series of lessons (e.g., introduction, core, ending of the lesson). The *quality* dimension refers to the properties of the specific factor itself, as these are discussed in the literature. In the case of orientation, we look at the extent to which orientation activities are clear for the students. This dimension also refers to the impact that the activity has on student engagement in the learning process. For example, teachers may present the reasons for doing an activity simply because they have to do it and is part of their teaching routine, without having much effect on student participation,

whereas others may encourage students to identify the purposes that can be achieved by doing an activity and thereby to increase their motivation towards a specific activity/lesson/series of lessons.

The dynamic model takes into account the findings of research into differential educational effectiveness (Campbell et al. 2004). As a consequence, *differentiation* is treated as a measurement dimension and is concerned with the extent to which activities associated with a factor are implemented in the same way for all the subjects involved with it. Adaptation to the specific needs of each group of students may increase the successful implementation of a factor and ultimately maximise its effect on student outcomes. In the case of orientation, differentiation is measured by looking at the extent to which teachers provide different types of orientation activities to students according to their learning needs, and especially by taking into account differences in the personal and background characteristics of students.

Therefore, the dynamic model attempts to describe the complex nature of effective teaching not only by pointing out the importance of specific factors and dimensions, but also by explaining how the functioning of each factor can be defined. To achieve this purpose, the model also assumes that the eight teaching factors and their dimensions may be inter-related, meaning that teachers that perform well in some factors or aspects of some factors (e.g., the quality of application) may also perform well in others (e.g., quality of structuring). In the next section, we refer to studies investigating the main assumptions of the model and especially whether grouping of factors could be identified.

2 Investigating the validity of the dynamic model and identifying stages of effective teaching

Some material supporting the validity of the dynamic model has been produced since 2006, when the model was developed. Specifically, longitudinal studies and one meta-analysis have been conducted to test the main assumptions of the model, as mentioned in the first section of this paper. Table 1 refers to these studies and the type of support that each assumption of the model has received. The following observations arise from this Table. First, it is clear that none of these studies has provided negative results in relation to any assumption of the dynamic model. Moreover, all studies have provided support for the multilevel nature of the model since factors operating at different levels were found to be associated with student achievement gains. These studies have also revealed that the teaching factors and their dimensions are associated with student achievement gains. Cognitive learning outcomes in different subjects (i.e., mathematics, language,

science and religious education) as well as non-cognitive outcomes, such as student attitudes towards mathematics, were used to measure the impact of factors. Thereby some support for the assumption that these factors are associated with student achievement gains in different learning outcomes has been provided. The generic nature of the factors is also supported by the fact that the effects of these factors on different student learning outcomes were similar (i.e., Cohen's *d* values were around 0.20).

Second, the meta-analysis provided support for the assumption that the teaching factors have an impact on student achievement, and revealed that the great majority of effectiveness studies conducted during the last three decades were concerned only with the impact of the quantitative characteristics of a given factor upon student achievement (i.e., the frequency dimension). For example, a study may have examined only whether the number of application activities offered to students had an effect on student outcomes. The empirical studies which have been conducted in order to test the validity of the dynamic model have also revealed that all five dimensions used to measure quantitative and qualitative characteristics of the functioning of factors should be used to explain variation in student achievement gains. Namely, testing the factorial structure of the observation data, these studies demonstrated that a five factor structure was better than one factor structure (see point 2 in Table 1). In addition, using all five dimensions to measure the functioning of the teaching factors was found to explain a higher percentage of variance in student outcomes rather than using a single dimension. Furthermore, some studies were not in a position to identify the impact of the frequency dimension of a specific factor on student achievement, but revealed that other dimensions of this factor were associated with student achievement.

Table 1 Empirical evidence supporting the main assumptions of the dynamic model emerging from empirical studies and meta-analyses

Assumptions of the dynamic model about the teacher level factors	Studies and Meta-analysis
1. Multilevel in nature: Factors associated with student achievement operate at different levels (student, classroom, school)	All
2. Separate scores for each dimension are needed to measure each teaching factor	1, 2, 3, 4, 6, 7, 8, 12
3. Teaching factors and their dimensions are associated with student learning outcomes (cognitive and non-cognitive)	1, 2, 3, 4, 5, 6, 7, 8, 12
4. Relations exist among factors operating at the same level: stages of effective teaching have been identified	1, 4, 5, 8, 9, 10, 11
Negative results in relation to any assumption	None

Studies:

1. A longitudinal study (Kyriakides et al. 2009) measuring teacher effectiveness in different subjects (mathematics, language, religious education).
2. A study investigating the impact of teaching factors on achievement in mathematics and language of Cypriot students at the end of pre-primary school (Kyriakides and Creemers 2009).
3. A European study testing the validity of the dynamic model by investigating the impact of teaching factors on student achievement in mathematics and science (Panayiotou et al. 2014).
4. A study in Canada searching for stages of effective teaching (Kyriakides et al. 2013).
5. An experimental study investigating the impact upon student achievement of a teacher professional development approach based on the Dynamic Approach (Antoniou and Kyriakides 2011).
6. A longitudinal study investigating the impact of teaching factors on mathematics achievement of primary students in Ghana (Azigwe et al. 2016).
7. A longitudinal study searching for the impact of teacher behavior on promoting students' cognitive and metacognitive skills (Creemers and Kyriakides 2015).
8. An experimental study searching for the impact and sustainability of the dynamic approach on improving teacher behaviour and student outcomes (Antoniou and Kyriakides 2013).
9. An experimental study searching for stages of teacher's skills in assessment (Christoforidou et al. 2014).
10. The effects of two intervention programs on teaching quality by considering the impact of teaching factors on student achievement in mathematics (Azkiyah et al. 2014).
11. Using the dynamic model to identify stages of teacher skills in assessment in different countries: a longitudinal study (Christoforidou and Xirafidou 2014).
12. A quantitative synthesis of 167 studies searching for the impact of generic teaching skills on student achievement (Kyriakides et al. 2013).

Third, with regard to the attempt of the model to search for relationships among factors operating at the same level, seven studies were conducted in different countries. These studies supported the assumption that the teaching factors of the dynamic model and their dimensions are inter-related, and revealed that they can be classified into stages of effective teaching, structured in a developmental order. In each of these studies, the Rasch model was used to analyse teacher performance in relation to the teaching skills included in the dynamic model and it was found that these skills were well targeted against the teachers' measures (see Creemers

and Kyriakides 2015). Moreover the reliability/separability of each scale was satisfactory (i.e., higher than 0.90). It was thus possible to identify stages of teaching skills in different countries. The fact that the Rasch model was found to fit the data in these seven studies can be attributed to the strong correlations found to exist among the factors and their dimensions. However, the functioning of each factor separately should be taken into account in defining effective teaching and using observational data for teacher improvement purposes. This study investigates this assumption further by searching for the extent to which teachers located at the same stage may need to set different improvement priorities. Table 2 presents the classification of factors and dimensions on the basis of their difficulty level, as these parameters emerged from the Rasch model, which showed that they are optimally clustered into five clusters. The parameter estimates presented in Table 2 emerged from study 1 (see Table 1). Similar estimates were also found in the other studies searching for stages of effective teaching. Teachers exercising more advanced types of teaching behaviour were found to have better student learning outcomes (see Creemers and Kyriakides 2015).

Table 2 reveals that the lowest classification of factors (i.e., stage 1) was found to be related to the basic elements of direct teaching. This stage refers to the quantitative characteristics of factors such as management of time, structuring of lessons, posing questions and assigning application activities to students. The second group of factors (i.e., stage 2) focuses not only on the quantitative aspects of the functioning of the teaching factors but also incorporates some qualitative features of the three factors associated with the direct teaching approach (i.e., structuring, application, and questioning). Specifically, it is assumed that teachers located at this stage are able to ensure not only the sufficient use of these factors but also their appropriate use, taking into consideration their qualitative characteristics. Skills included in these two lower stages are considered as easier to acquire than skills located at the upper more demanding stages. Moving to stage 3, teachers are not only expected to use skills effectively related to the direct teaching approach, but also to be able to establish and maintain a learning environment in the classroom that encourages different types of on-task interactions (i.e., student–student interactions and teacher–student interactions). Focus at this stage is also given to the orientation of students towards the learning goals and to their contribution in identifying the objectives of a lesson or a series of lessons. Stage 4 is mainly concerned with the differentiation dimension of factors associated with direct teaching so as to accommodate lessons to the specific needs of different groups of students, whilst stage 5 also provides emphasis to the differentiation of factors which are in line with the constructivist approach, such as orientation and modelling.

Table 2 Rasch parameter estimates of each teaching skill associated with the teaching factors of the dynamic model

Classroom level factors	Rasch parameter estimates
Stage 1	
Frequency management time	– 2.69
Stage management of time	– 2.62
Frequency structuring	– 2.58
Frequency application	– 2.45
Frequency assessment	– 2.40
Frequency questioning	– 2.38
Frequency teacher-student relation	– 2.16
Stage 2	
Stage structuring	– 1.56
Quality application	– 1.50
Stage questioning	– 1.48
Frequency student relations	– 1.42
Focus application	– 1.37
Stage application	– 1.33
Quality questioning	– 1.30
Stage 3	
Stage student relations	– 0.74
Stage teacher-student relation	– 0.71
Stage assessment	– 0.62
Frequency teaching modelling	– 0.60
Frequency orientation	– 0.50
Focus student relations	– 0.36
Quality: feedback	– 0.32
Focus questioning	– 0.31
Focus teacher-student relation	– 0.31
Quality structuring	– 0.29
Quality assessment	– 0.26
Stage 4	
Differentiation structuring	0.59
Differentiation time management	0.61
Differentiation questioning	0.71
Differentiation application	0.88
Focus assessment	0.94
Differentiation assessment	1.17
Stage teaching modelling	1.21
Stage orientation	1.29
Stage 5	
Quality teacher-student relation	2.32
Quality student relations	2.39
Dif teacher-student relation	2.50
Differentiation student relations	2.72
Focus orientation	2.89
Quality orientation	2.95
Differentiation orientation	3.00
Quality of teaching modelling including differentiation	3.04
Focus teaching modelling	3.05

1. The name of each teaching skill is indicated firstly by the Dimension and secondly by the Factor used to define the teaching skill
2. Rasch parameter estimates indicate the difficulty level of each aspect of teaching factor measured in logits

3 Methods

The aim of this paper is to use the dynamic model to analyze three mathematics video-lessons for the following purposes: (a) to identify strengths and weaknesses in using the dynamic model to measure quality of teaching in mathematics; and (b) to examine the added-value of using a qualitative research approach in identifying individual teacher professional development needs in teachers located at different and/or the same stage.

3.1 Participants

The three video-lessons analyzed were available for usage from the NCTE video study of Harvard University and concern 4th grade mathematics. Having three lessons from different teachers with different improvement needs may allow for an in-depth analysis of the teaching processes that take place irrespective of the stage at which they are located. Pseudonyms are used in describing the results of the lesson analyses to ensure confidentiality. More details on the lessons observed can be found in the introductory paper of this issue (see Charalambous and Praetorius this issue).

3.2 Observation instruments for measuring quality of teaching

For the analysis of the three video-lessons two low-inference (LIO1 and LIO2) and one high-inference observational instruments were used since each type of instrument has advantages as well as disadvantages. In particular, the low observation instruments demonstrate a higher level of reliability, however, in the attempt to develop specific scores for each factor, information on its qualitative characteristics may be lost. On the other hand, even though the high inference instrument provides a more holistic view of the lesson, reliability is more difficult to achieve. Using all three instruments together may provide more information on the lesson observed regarding the teaching factors of the dynamic model. In particular, these instruments were designed to collect data concerned with different aspects of the eight teaching factors of the dynamic model, and previous studies provided empirical support for their construct validity (see Table 1). Specifically, LIO1 and LIO2 are best used when combined together as they examine different aspects of the factors and together they are able to generate data for all teaching factors of the dynamic model (except student assessment) and their five dimensions. In practice, when conducting classroom observations using LIO1 and LIO2, two observers are needed (i.e., one observer codes the lesson using LIO1 and the other using LIO2). LIO1 provides information on the classroom learning environment

(including the teacher-student and student-student interactions) and the management of time factor. This instrument is based on Flanders' system of interaction analysis (Flanders 1970). However, a classification system of teacher behaviour was developed, based on the way these two factors of the dynamic model are measured. Specifically, this instrument is concerned with 17 types of interactions that may be observed in a lesson, such as the teacher commenting on students' answers, teacher giving instructions, or students collaborating. It also helps generate management of time and classroom learning environment factor scores. Observers using this instrument should record the type of behavior that is observed every 10 s. Therefore it is important to note that this observation instrument, as well as LIO2, are time based.

LIO2 refers to five factors of the model: (a) orientation, (b) structuring, (c) teaching modelling, (d) questioning, and (e) application. The LIO2 instrument was designed in a way that enables the collection of more information in relation to the quality dimension of these five factors. For example, in regard to the measurement of the quality of a modelling activity, observers have to indicate whether this activity is: (a) given by the teacher, (b) occurring through guided discovery or (c) is the product of student individual thought (i.e., discovery). From the beginning of the lesson each "activity" related to the five factors is documented and observers need to document its duration in minutes. An activity can be a set of teacher actions-statements that have a certain goal. For instance, in the case of orientation, a teacher may hold a discussion with his/her students (which may last for 3 min) in order to help them understand the importance of the aims of the lesson. For example, the teacher may call on students to express their views on why it is important to calculate the area of the circle. Information collected regarding these activities concerns the following aspects: (a) the sequence of an activity (used to generate a score on stage dimension), (b) its duration in minutes (in order to measure the frequency dimension), (c) whether activities are specific or more general (to generate a score on focus), (d) their quality and (e) whether they are implemented the same way for all students (i.e., differentiation). Also, if the teacher begins the lesson by reminding students what was done in the previous lessons of a unit then this first activity, which concerns the factor of structuring, will receive code 1 for the sequence in which it appeared in the lesson. With regard to focus, this activity will be coded as "related to previous lessons" and then in terms of quality, the observer, based on students' reactions and involvement during this activity, will state whether it was clear or not for them.

Finally, the high-inference observation instrument provides a more general overview of the lesson and covers the five dimensions of all eight factors of the model. Observers

are expected to complete a Likert scale to indicate how often each teaching behaviour was observed. For example, an item concerned with the frequency dimension of orientation is asking observers to indicate how much time the teacher spent to explain the objectives of the lesson. In order to measure the quality dimension of this factor, one of the items of the high-inference observation instrument is asking observers to indicate the extent to which the orientation activities that were organised during the lesson helped students understand the new content. The high-inference observation instrument is completed as soon as the lesson finishes, since observers are asked to document a broader view of the lesson based on the factors of the dynamic model. Prior to conducting the observations all observers are trained with the use of video-lessons as well as live lessons and observers with low inter-observer reliability are either retrained or not selected for further observations.

3.3 Data analysis

For the analysis of the three video-lessons four independent observers were used. The observers selected for this study had been trained and gained experience in several other studies for conducting classroom observations. After the four observers watched the whole video-lessons twice in order to code them using the three observation instruments, their scores were compared to examine whether there was consensus (i.e., the inter-rater reliability was examined). Differences in some scores, mainly concerned with the quality dimension, were identified but after a discussion, consensus was reached. In previous studies, initially the Rasch and then the Saltus model was used to analyze the data that emerged from the three instruments (see Kyriakides et al. 2009). The Saltus model allows the researcher to differentiate between major and less pervasive changes in moving from one stage to the other without sacrificing the idea of one common underlying continuum (Mislevy and Wilson 1996). Since in this paper only three teachers were observed we entered the quantitative data from the observations of the three video-lessons into a data base that was developed in a previous experimental study that examined the long term effect of a program based on the Dynamic Approach on the quality of teaching (Kyriakides et al. 2017), and the analysis was re-run. The sample from that study was used as it referred to teachers teaching the same subject (i.e., mathematics) to a similar age group of students. Although the data used derive from a different country, the content of the three video-lessons is also taught in the country in which the experimental study was conducted, to the same age group of students. It should also be taken into consideration that the factors examined are considered as generic. To determine the stage in which a teacher is located, the scores of an individual teacher cannot be seen only in isolation. On the contrary, by

utilizing the Rasch and Saltus models each teacher's observation scores are compared with those of other teachers to determine which is in a highest or lowest stage. In this way, we could identify in which stage each of the three teachers in the video-lessons is situated.

At the same time, by conducting a qualitative analysis of the three-video lessons, a more in-depth analysis of the skills of each teacher was conducted, which helped us to justify his/her allocation to a specific stage. Having videotaped data sources instead of live lessons provided us with the opportunity to go over certain points of the lesson that were seen as important for assessing the quality of the activities offered to students. The qualitative analysis was conducted prior to the quantitative using the Constant Comparative Method in order to generate relevant units of analysis without being influenced by the results of the quantitative analysis. To identify units of analysis we first of all used the framework of the dynamic model (i.e., the eight teaching factors and their dimensions). In addition, units of analysis related with the classroom context as well as on the way students responded to the activities of the lesson were used.

4 Results

4.1 Quantitative analysis: using the Rasch model to identify stages of teaching skills

For the analysis of the observational data, the extended logistic model of Rasch was initially used (Andrich 1988) in order to identify the extent to which the five dimensions of the teaching factors could be reducible to a common unidimensional scale. By using the Rasch model to analyse teacher performance in relation to the teaching skills included in the dynamic model, it was found that these skills were well targeted against the teachers' measures ($n = 103$) since teachers' scores ranged from -2.96 to 3.04 logits and the difficulties of the 44 teaching skills ranged from -2.69 to 3.05 logits. By referring to teaching skills we mean the teacher's knowledge and ability to initiate student learning, monitor it, and evaluate the learning outcomes. Specifically, we search for the extent to which teachers demonstrate their ability to use each of the eight teaching factors and dimensions included in the dynamic model when they teach mathematics. Moreover, the reliability of persons (i.e., teachers) and items (i.e., teaching factors) is calculated through the Rasch analysis, indicating how well the scale discriminates among teachers based on their estimated teaching skills and how well each of the teaching skills can be discriminated from one another on the basis of their difficulty. It was found that the separability of each scale was satisfactory (i.e., higher than 0.93). Having established the reliability of the scale, it was investigated whether the various teaching skills

could be systematically grouped into the five stages identified in the previous studies. Applying this method, it was found that the cumulative D for the five-cluster solution was 58%, whereas the sixth gap added only 4%. We then used the Saltus model to find out how deep is the divide separating the five stages of effective teaching emerged from cluster analysis. The results that emerged from the Saltus model revealed that the gap between two consecutive stages is in line with the ones that emerged in previous studies (i.e., the gap between stage 1 and stage 2, as well as stage 2 and stage 3 is much smaller than the gaps between stages 3 and 4 and between stages 4 and 5). By looking at the Rasch estimate of each teacher offering the video-lessons, we could then identify in which stage he/she was found to be situated. The first two teachers were found to be situated in stage 1 (i.e., Mr. Smith with a score of -2.40 logits and Ms. Young with a score of -2.32) whereas the third teacher (i.e., Ms. Jones) was allocated to stage 2.

4.2 Qualitative analysis of the three video-lessons

The aim of this study was not only to identify the stage at which each of the three teachers is located, but to move a step further and examine in more depth the possible differences in the quality of teaching, even in teachers that are allocated to the same stage. To achieve a more thorough understanding of the reasons for which each of the three teachers was allocated to each respective stage when conducting the above mentioned Rasch analysis, as well as elaborate on the individual needs of each teacher, a detailed analysis per lesson is provided below. One should however bear in mind, that we assigned teachers in the three video-lessons to stages having observed only a single lesson that was available per teacher. In previous studies we used at least three observations to draw conclusions about the professional needs of teachers (see Antoniou and Kyriakides 2013).

4.2.1 Mr. Smith's lesson

The main reasons for which Mr. Smith was allocated to stage one, can be traced to the low scores that emerged regarding the qualitative aspects of the factors associated with the direct teaching and mastery learning (i.e., structuring and questioning), as well as to the limitations observed in his ability to manage the classroom effectively, in terms of keeping all students on-task and maximising their engagement rates.

Teacher of video-lesson 1 was able to deal effectively with the overall time of the lesson since the activities planned were sufficient for the time of the lesson and time was left at the end for explaining homework, however, not all students were on-task in many points of the lesson. For

example, when Mr. Smith called some students to go to the board in order to practice in measuring angles, all the other students were only watching and therefore for them, the teaching time was not used equally effectively. Even though an activity in which only one student is practicing at the board may seem as an application activity, it provides the opportunity only to a limited number of students to actually apply new knowledge, as the others are merely observers. To avoid having students off-task, Mr. Smith could have asked all students to practice in measuring angles by giving them a leaflet with relevant application tasks.

Second, all application activities were conducted at only one point of the lesson and were product-oriented since students were repeatedly carrying out the same application activity; in this case measuring angles. No process-oriented application activities were observed during the lesson that requested students to apply the new knowledge to something more complex such as solving problems including angles (low scores in the stage and quality dimension). Similarly, the one orientation activity held during the lesson also lacked in terms of the quality dimension. In particular, even though the subject of the lesson was easily related with the real life of students the teacher did not manage to link the topic of the lesson effectively with real life situations, thus making the activities of the lesson meaningful for students. The only attempt of the teacher to explain why learning to measure angles accurately is useful (i.e., orientation activity) lacked in terms of the quality dimension, since it was not related to the age and context of the students. Namely, the teacher said that "If you decided to become an engineer one day and you get 66 degrees and the answer's 65 degrees, then it's not gonna be built well. If you decide you want to be an architect or civil engineer—should have the buildings be perpendicular. Shouldn't they form 90 degree angles? If you make your building 91 degrees everybody's gonna be walking slanted a little bit".

Third, even though Mr. Smith was able to communicate clearly with students and pose questions that were clear for the students in terms of their content, some issues may be raised in terms of the quality of the questions posed and the feedback provided. Namely, most questions were addressed to all the students in the classroom and multiple students provided their answers simultaneously. In this case, it is not possible for the teacher to examine whether all students actually know the correct answer or whether some are just repeating their classmates' answer. In many cases, the teacher was also providing the answers to the questions himself, and when students answered correctly he provided the explanation instead of asking the students to explain the way they had reached the specific answer. At the same time, in cases in which a student failed to give a correct answer, the feedback provided (if provided) was not constructive enough

to help the student understand the mistake made (i.e., the teacher said “No” or waved his head).

Finally, the teacher of lesson 1 did not attempt to promote any interactions among students either by inviting students to comment on their classmates’ answers to questions or by assigning them group application activities.

4.2.2 Ms. Young’s lesson

The teacher of the second video-lesson was also classified in the lower stage of teaching skills (i.e., stage 1). Even though some common needs for improvement were identified between the teachers of the first two lessons, some further issues regarding the quality of teaching of the second lesson arise.

First, it should be noted that during this lesson, some structuring activities took place at the beginning when the objectives of the lesson were written, mentioned and linked to previous lessons and at the end, when the teacher was calling attention to the main points of the lesson. This teacher therefore, had a higher score regarding the stage dimension than did Mr. Smith. However, it was not clear whether the objectives of the lesson were understood by the students, since they were only mentioned by the teacher without any student participation. Another aspect of the factor concerned with the structuring of lessons, refers to the progression of activities in a lesson in such a way that activities gradually increase in terms of their difficulty level. In this lesson, progression was observed from the beginning until the end when the activities became more intellectually demanding.

Second, with regard to questioning, some process questions were posed in the second lesson and the teacher attempted to use questioning techniques to aid learning (for example, when a student gave an answer, she asked another student to say it in another way and then another one to explain why the answers given by the previous two students were correct), however, similarly to lesson one, the feedback provided was not constructive enough to help students understand their mistakes (e.g., the teacher said: “You’re making the mistake, so fix it”) and the conclusions drawn from a student’s answer were derived in most cases from the teacher’s words instead of the teacher asking students to comment on their classmates’ answers. However, the teacher attempted to promote interaction between students by asking them to co-operate during an application activity. Yet, the specific activity did not require student collaboration in order to be achieved and therefore, as expected, students worked individually.

Finally, like Mr. Smith, one of the main reasons for which Ms. Young was allocated to stage 1 was the difficulty faced in keeping students on-task throughout the lesson and using the teaching time effectively for all students. Nevertheless, in this case, the teaching time was lost for some students,

not due to the nature of the activities, like lesson one, but due to the teacher’s difficulty in dealing with student misbehaviour effectively. In particular, during the lesson, students were talking to each other, standing, going to the camera and moving around the classroom, while generally an orderly classroom climate was not achieved. In some cases, students who interrupted the lesson due to their behaviour were asked to leave the classroom. Therefore, due to the teacher’s difficulty in dealing effectively with misbehaviour, the teaching time was lost for these students.

4.2.3 Ms. Jones’s lesson

In general, during the third lesson, structuring activities were observed at different stages of the lesson (i.e., beginning and end), as well as progression in the difficulty level of the activities in which students were involved, application opportunities at different points of the lesson and some established routines (e.g., how to show that students have finished the task they are working on).

In particular, at the beginning of the lesson Ms. Jones explained the objectives of the lesson and related them to what had been taught in the previous lesson, providing links between different parts of the unit. Similarly, at the end of the lesson she summed up the main points of the lesson, including students in the conversation. During the whole lesson the students were asked questions, yet in many cases either multiple students simultaneously shouted the answer or the teacher provided the answer. Moreover, even though the teacher established on-task interactions with students throughout the main part of the lesson, by asking questions and calling them to explain their answers, interactions amongst students were not equally observed and this is one of the reasons that Ms. Jones could not be allocated to stage 3. In particular, students were not invited to comment on their peers’ answers, or discuss with each other possible ways of solving a problem, even during group work. Thus, low scores for the factor concerned with interactions among students were given. Still, during application activities the teacher moved around the classroom providing feedback and when necessary she gave clarification to the whole class.

In addition, even though well-established routines were implemented and followed by all students so as to minimise the loss of teaching time (e.g., “put a thumbs-up on your desk so I can see who is ready to completely move on”), almost ten minutes of teaching time were spent at the beginning of the lesson, for preparation. Namely, students were asked to divide their paper into three sections and copy the title from the board. This activity, which was not directly related to the objectives of the lesson, could have been avoided, saving time for other activities. However, by using only one observation this cannot be considered as a sign of insufficient time management, especially since after

the first ten minutes of preparation the other time was used for on-task activities.

5 Discussion

In the first part of this section we discuss the further support provided to the framework for measuring quality of teaching based on the dynamic model through the in-depth analysis of the three video-lessons. In the second part, implications for its further development are drawn, based on the limitations observed during the qualitative analysis, and the contribution of the dynamic model in research on mathematics teaching is discussed.

5.1 Empirical support of the validity of the dynamic model emerged from the three case studies

First, in this paper, we demonstrate and justify the rationale behind the five stages of effective teaching. Table 2 shows that factors associated with the direct and active teaching approach were found to be situated in lower stages whereas those associated with the constructivist approach (i.e., orientation and modelling) belong to higher stages. In these three lessons, activities associated with all factors related with the direct and active teaching approach (i.e., structuring, application and questioning) were observed and therefore positive scores for the frequency dimension of all factors were generated. The fact that almost no orientation and modelling activities were observed reveals why in Table 2 the frequency dimension of factors associated with the constructivist approach is situated in higher stages. Second, by looking at the dimensions measuring qualitative characteristics of teaching factors, one can see that the qualitative dimensions of each factor are situated at higher stages than the quantitative characteristics (see Table 2). By comparing the three lessons, one can also see that all teachers offered activities associated with structuring, application and questioning and therefore positive scores for the frequency dimension of these factors were generated, but only relevant activities observed in lesson 3 were rated positively in regard to their qualitative characteristics (except differentiation). As a consequence, teachers of the first two lessons were considered to belong in stage 1 whereas the teacher of lesson 3 was found to be situated in stage 2. These findings seem to explain why the frequency dimension of each factor is situated at a lower stage than the other dimensions. Third, in these three lessons no differentiation activity associated with any factor was observed and this finding seems to reveal why previous studies found the differentiation dimension of most factors situated at stage 4 or 5. Fourth, by looking at these three lessons, one can also see that there is no lesson where teachers were rated positively for any skill situated above the stage in

which they were found to belong. For example, we did not observe any differentiation when students were dealing with application activities (stage 4) and at the same time only one of the teachers (i.e., Mr Smith) was found to provide students with an orientation activity, which, however, was not planned, lasted only one minute and was not contextually relevant for the students.

The in depth analysis of these three lessons not only reveals the strengths of this approach but also helps us see that some teachers may be situated at the same stage but may have different professional needs. By comparing the two teachers found to be situated in stage 1 one can see that they need to find out how to maximise the use of teaching time and for this reason relatively low scores on management of time and dealing with misbehaviour were generated. These two skills were both situated at stage 1 but we found out that no student misbehaviour was observed in lesson 1, whereas in lesson 2 disturbing incidents were not handled properly (see Sect. 4.2). On the other hand, the activities offered to students during lesson 1 were not inviting all students to participate (see Sect. 4.2.1). Although the quantitative data reveal that both teachers are situated at the same stage, this paper reveals that in asking teachers to design and implement their action plans one may expect them not to give the same emphasis to all factors and dimensions situated in their stage. At this point, it is important to acknowledge a limitation in measuring the skills of teachers in dealing with student misbehaviour. When no student misbehaviour incident takes place, the observer can say nothing about the ability of a teacher in regard to this factor since no data about the teacher's ability to deal with misbehaviour can be generated.

5.2 Searching for ways to expand the framework used to measure quality of teaching

In the next part of this section, we identify aspects of a lesson that are not examined and possibilities for integrating this approach, which is generic, with others that are domain-specific and are based on findings of research on teaching mathematics. First, our attempt to use the dynamic model to analyse the three video lessons revealed some difficulties in using the instruments developed to test the validity of the model. Although the low-inference observation instruments were in a position to generate more precise data about the skills of each teacher, this was not always the case especially when the quality dimension of an activity associated with a specific factor was rated. For example, in rating the quality dimension of a structuring activity, raters should evaluate the impact that the activity has on students. In this case, consensus among raters was not always reached. For example, when the teacher in lesson 3 reminded students that in earlier years they learnt that multiplication is repeated addition, all raters agreed about the score of the focus dimension of this

structuring activity, but this was not the case with the quality dimension. The quality dimension of structuring refers to the clarity of an activity to students but not all observers found the activity mentioned above clear.

Second, another difficulty we encountered in using the instruments to analyse the three video-lessons had to do with the fact that we could not conduct a short interview with the teacher before observing the lessons in order to find out what was taught in the previous lessons and what the aims of the lesson observed were. When this information is not available, the rater cannot easily identify which application activities are not a simple repetition of what was taught in the previous lesson. Before conducting any observation, a short interview with the teacher should therefore take place in order to help raters understand the context in which the lesson takes place.

Third, the dynamic model refers to generic factors measuring teacher behaviour in the classroom without considering that some teachers may have insufficient knowledge about the topic they teach. By observing these three lessons, we noticed that one of the teachers made some mathematical errors, but the framework and its instruments do not enable us to generate scores on teacher knowledge and identify relevant professional needs. Although most of the studies testing the validity of the framework were used to evaluate teachers in mathematics, one could argue that specific aspects of mathematics teaching are not addressed. It is partly for this reason that raters do not examine the mathematical errors that teachers may make during a lesson. Given the recent emphasis on domain-specific teaching practices (e.g., Chen et al. 2011), the framework presented here could also be expanded to include the effects of more subject-specific teaching factors on student learning. For example, one could examine the representations that are used to explain a specific construct in mathematics and/or the explanation that is provided to students in order to avoid misconceptions. It should however be acknowledged that since this framework is used for measuring quality of teaching in different subjects, it is not feasible to develop instruments that examine too many details about the subject taught.

Even though the dynamic model is focused only on generic teaching factors, its main contribution to research on teaching mathematics may lie in the fact that it proposes five measurement dimensions so as to provide a more complete understanding of the functioning of each factor. Unlike most studies conducted both in the field of EER, as well as in the field of mathematics teaching, studies based on the framework proposed by the dynamic model provide a clear distinction and a more accurate measurement not only of the quantitative aspects of the factors (i.e., the frequency), but also of the qualitative (i.e., stage, focus, quality and differentiation). Especially, collecting information on the qualitative aspects of the teaching factors may allow for the allocation

of teachers into stages of effective teaching and based on the needs identified for improving their teaching practice, provide them with individual improvement feedback.

Having said that, one could also argue that we need a more precise definition of all generic and domain-specific factors and a systematic comparison of these factors, which may reveal the extent to which there is an overlap between some generic and domain-specific factors. It could also be examined whether domain-specific factors can be included in the dynamic model and also if these factors can also be grouped into stages of effective teaching. Combining both generic and domain-specific factors may allow the development of a comprehensive framework for measuring quality of teaching.

References

- Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education*, 1(4), 363–378.
- Antoniou, P., & Kyriakides, L. (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., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behavior and student outcomes. *Teaching and Teacher Education*, 29(1), 1–12.
- Azigwe, J. B., Kyriakides, L., Panayiotou, A., & Creemers, B. P. M. (2016). The impact of effective teaching characteristics in promoting student achievement in Ghana. *International Journal of Educational Development*, 51, 51–61.
- Azkiyah, S. N., Doolaard, S., Creemers, B. P. M., & Van Der Werf, M. P. C. (2014). The effects of two intervention programs on teaching quality and student achievement. *Journal of Classroom Interaction*, 49(1), 4–11.
- Brekelmans, M., Slegers, P., & Fraser, B. (2000). Teaching for active learning. New learning (pp. 227–242). Netherlands: Springer.
- Brophy, J., & Good, T. L. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: MacMillan.
- Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2004). Effective teaching and values: Some implications for research and teacher appraisal. *Oxford Review of Education*, 30(4), 451–465.
- Chen, W., Hendricks, K., & Archibald, K. (2011). Assessing pre-service teachers' quality teaching practices. *Educational Research and Evaluation*, 17(1), 13–32.
- Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher skills in assessment. *Studies in Educational Evaluation*, 40, 1–11.
- Christoforidou, M., & Xirafidou, E. (2014). Using the dynamic model to identify stages of teacher skills in assessment. *Journal of Classroom Interaction*, 49(1), 12–25.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London and New York: Routledge.
- Creemers, B. P. M., & Kyriakides, L. (2015). Developing, testing and using theoretical models of educational effectiveness for

- promoting quality in education. *School Effectiveness and School Improvement*, 26(1), 102–119.
- Flanders, N. A. (1970). Analyzing teacher behavior (pp. 100–107). Reading: Addison-Wesley.
- Joyce, B., Weil, M., & Calhoun, E. (2000). *Models of teaching*. Boston: Allyn & Bacon.
- Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, 12(1), 1–10.
- Kyriakides, L., Archambault, I., & Janosz, M. (2013). Searching for stages of effective teaching: A study testing the validity of the dynamic model in Canada. *Journal of Classroom Interaction*, 48(2), 11–24.
- Kyriakides, L., Christoforidou, M., Panayiotou, A., & Creemers, B. P. M. (2017). The impact of a three-year teacher professional development course on improving quality of teaching: Strengths and limitations of the dynamic approach. *Teaching and Teacher Education*, 51(1), 51–61.
- Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152.
- 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.
- Mislevy, R. J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, 61, 41–71.
- Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art—teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231–256.
- Panayiotou, A., Kyriakides, L., Creemers, B. P. M., McMahon, L., Vanlaar, G., Pfeifer, M., Rekalidou, G., & Bren, M. (2014). Teacher behavior and student outcomes: Results of a European study. *Educational Assessment, Evaluation and Accountability*, 26(1), 73–93.
- Pellegrino, J. W. (2004). Complex learning environments: Connecting learning theory, instructional design, and technology. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instructional design* (pp. 25–49). Mahwah: Lawrence Erlbaum Associates.
- 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.
- Schoenfeld, A. H. (1998). Toward a theory of teaching-in-context. *Issues in education*, 4(1), 1–94.
- Stenmark, J. K. (1992). Mathematics assessment: Myths, models, good questions and practical suggestions. Reston, VA: NCTM.