

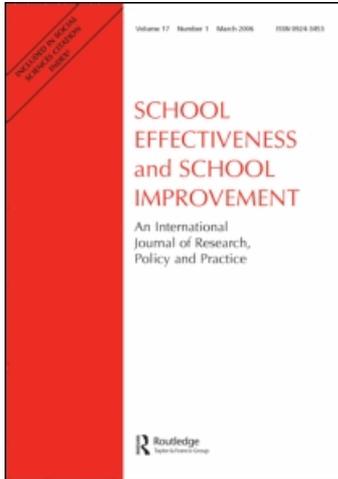
This article was downloaded by: [HEAL-Link Consortium]

On: 12 June 2010

Access details: Access Details: [subscription number 786636649]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



School Effectiveness and School Improvement

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t714592801>

Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model

B. P. M. Creemers^a; L. Kyriakides^b

^a Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands ^b Department of Education, University of Cyprus,

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

To link to this Article: DOI: 10.1080/09243450600697242

URL: <http://dx.doi.org/10.1080/09243450600697242>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Critical Analysis of the Current Approaches to Modelling Educational Effectiveness: The importance of establishing a dynamic model

B. P. M. Creemers^a and L. Kyriakides^{b*}

^a*Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands,* and ^b*Department of Education, University of Cyprus*

(Received 7 July 2005; accepted 17 January 2006)

Researchers in the area of educational effectiveness should attempt to develop a new theoretical framework. A critical analysis of the current models of educational effectiveness research is provided and reveals that a dynamic model of effectiveness must: (a) be multilevel in nature, (b) be based on the assumption that the relation of some effectiveness factors with achievement may be curvilinear, (c) illustrate the dimensions upon which the measurement of each effectiveness factor should be based, and (d) define relations among the effectiveness factors. In principle, each factor that refers to the classroom, school, and system, can be measured by taking into account five dimensions: frequency, focus, stage, quality, and differentiation. Examples of measuring effectiveness factors operating at different levels using these 5 dimensions are given. More attention in describing in detail factors associated with teacher behaviour in the classroom is given, since this is seen as the starting point for the development and the testing of the dynamic model. Finally, suggestions for the next steps in the development of other parts of the model are provided.

Introduction

In the last 25 years, Educational Effectiveness Research (EER) has improved considerably by the criticism on research design, the sampling, and statistical techniques. Methodological advances, particularly the availability of particular software for the analysis of multilevel data, have enabled more efficient estimates of

*Corresponding author. Department of Education, University of Cyprus, PO Box 20537, CY 1678 Nicosia, Cyprus. Email: kyriakid@ucy.ac.cy

teacher and school differences in student achievement to be obtained (Goldstein, 2003). It is also supported that longitudinal data are needed in order to enhance the validity of causal inferences in non-experimental research by providing a basis for assessing the direction of causation between two variables and by enabling some control over selection effects (Cook & Campbell, 1979). Moreover, incorporating multilevel effects into statistical analyses helps the researcher avoid a variety of errors of statistical inference including heightened probabilities of Type I errors, aggregation bias, and undetected heterogeneity of regression (Aitkin & Longford, 1986; DeLeeuw & Kreft, 1986; Goldstein, 1987; Raudenbush & Bryk, 1986, 1988). Finally, there is substantial agreement as to appropriate methods of estimating school differences or effects and the kinds of data required for valid comparisons to be made (Hopkins, Reynolds, & Gray, 1999). As far as the theoretical component of the field is concerned, progress was made by a more precise definition of the concepts used and the relations between the concepts (e.g., Levine & Lezotte, 1990; Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Scheerens, 1992). However, there is a shortage of rational models from which researchers can build theory. The problem is aggravated by infrequent use of whatever models exist (Scheerens & Bosker, 1997). As a consequence, most of the studies on educational effectiveness are atheoretical and are concerned with the establishment of statistical relationships between variables rather than with the generation and testing of theories which could explain those relationships (Creemers, 2002).

Another significant weakness of studies on educational effectiveness arises from the fact that almost all of them are exclusively focused on language or mathematics. Researchers have not been able to monitor pupils' progress in the full range of the school curriculum and did not examine educational effectiveness in relation to the new goals of education such as the development of metacognitive skills (Campbell, Kyriakides, Muijs, & Robinson, 2003). Thus, EER threw itself under the suspicion of being solely interested in the cognitive domain and restricting itself further by focusing on basic knowledge and skills. As a consequence, EER has been criticised by opponents for a narrow scope, reducing school learning to discrete, assessable, and comparable fragments of academic knowledge (Slee & Weiner, 1998, p. 2). For example, Lingard, Ladwig, and Luke (1998) state that educational effectiveness starts from an impoverished idea of what counts as achievement since it seems to assume that outcomes of schooling can be measured in conventional terms of skills, behaviour, knowledge, and competences. The arguments used by the critiques of EER can be countered by referring to numerous studies that used multiple measures of schooling outcomes (e.g., Bosker, 1990; Knuver & Brandsma, 1993; Kyriakides, 2005; Opdenakker & Van Damme, 2000). It becomes evident from these studies that it is possible to measure a broad range of outcomes in a valid and reliable way using traditional methods of assessment. In this context, this paper argues that there is a need to develop a new theoretical framework of EER which takes into account the new goals of education and emphasises teaching for understanding and development of metacognitive skills rather than teaching based on transmission of knowledge (Pines & West, 1986; Prawat, 1989). At the same

time, dynamic relations between the multiple factors associated with effectiveness should be identified. Thus, the next section of this paper provides a critical analysis of the current models of EER and identifies what the essential characteristics of the proposed dynamic model should be. Then, a dynamic model of EER focused on teachers' contribution to student learning is presented. Finally, suggestions for the next steps that have to be undertaken in order to develop the other parts of the model are provided.

A Critical Analysis of the Models of EER Based on Empirical Evidence

In the literature of educational effectiveness modelling, three basic approaches have been used. First, the economic approach is focused on estimating the relationship between the "supply of selected purchased schooling inputs and educational outcomes controlling for the influence of various background features" (Monk, 1992, p. 308). Such research is focused on producing a function which could explain each pupil outcome at a given time. The function may be linear, consisting of main effects and interaction terms or non-linear (Brown & Saks, 1986). The emerging "education production" models (e.g., Brown & Saks, 1986; Elberts & Stone, 1988) are based on the assumption that increased inputs will lead to increments in outcomes. However, the research done using these models revealed that the relation between input and outcomes is more complex than was assumed. For example, studies from Hanushek and Hedges (e.g., Hanushek, 1986, 1989; Hedges, Laine, & Greenwald, 1994) show that reducing student-teacher ratio and/or increasing the amount of funding education per student does not necessarily result in higher student outcomes.

The second approach to educational effectiveness modelling is similar to the economic approach but is focused on a different choice of antecedent conditions, since it is mainly focused on variables at student level which are assumed to predict student outcomes. Some attention is also paid processes from two different perspectives concerning learning and school as organisations. Within this approach, educational psychologists focused on student background factors such as "learning aptitudes", "personality", and "motivation", and on variables measuring the learning processes which take place in classrooms. On the other hand, the sociological perspective is focused on factors that define the educational background of students such as SES, gender, social-capital, and peer group. This perspective does not only examine student outcomes but also the extent to which schools manage to reduce the variance in student outcomes compared to prior achievement. Thus, two dimensions of measuring school effectiveness emerged from this perspective concerning the quality and equity. Moreover, the sociological perspective raises attention for process variables emerged from organisational theories such as the school climate, culture and structure, and for contextual variables.

Finally, the models of the third approach emerged by researchers' attempt to integrate the findings of School Effectiveness Research (SER), Teacher Effectiveness Research (TER), and the early input-output studies. Thus, the models of this

approach (e.g., Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992) have a multilevel structure, where schools are nested in contexts, classrooms are nested in schools, and students are nested in classrooms or teachers. Although these models make use of both organisational theories and theories of learning and refer to multiple factors at different levels, each of them is either focused on the classroom or the school level. Depending on this, more emphasis is given either to theories of learning (e.g., Creemers, 1994) or to organisational theories (e.g., Scheerens, 1992).

Four studies examined the validity of Creemers' model (De Jong, Westerhof, & Kruiter, 2004; Kyriakides, 2005; Kyriakides, Campbell, & Gagatsis, 2000; Kyriakides & Tsangaridou, 2004), which is one of the most influential theoretical constructs in the field (Teddle & Reynolds, 2000), and revealed that the influences on student achievement are multilevel. This finding is in line with the findings of most studies on educational effectiveness conducted in various countries (Teddle & Reynolds, 2000) and provides support to the argument that models of EER should be multilevel in nature. The analyses of the results of these studies also revealed that, next to the multilevel nature of effectiveness, the relationship between factors at different levels might be more complex than assumed in the integrated models. This is especially true for interaction effects among factors operating at classroom and student level which reveal the importance of investigating differentiated effectiveness (Campbell, Kyriakides, Muijs, & Robinson, 2004). Therefore, the proposed dynamic model of EER should demonstrate the complexity of improving educational effectiveness. In order to achieve this purpose, the following three major criticisms of current models of EER are taken into account.

First, meta-analyses of the effects of some effectiveness factors upon student achievement revealed that, although they have been perceived as factors affecting teacher or school effectiveness, the research evidence is problematic. For example, teacher subject knowledge is widely perceived as a factor affecting teacher effectiveness (Scriven, 1994), but teachers' subject knowledge, regardless of how it is measured, has rarely correlated strongly with student achievement (Borich, 1992; Darling-Hammond, 2000). The explanation may be, as Monk (1994) reported, that the relationship is curvilinear: A minimal level of knowledge is necessary for teachers to be effective, but beyond a certain point a negative relation occurs. Similar findings have been reported for the association of self-efficacy beliefs with teacher effectiveness (Schunk 1991; Stevenson, Chen, & Lee, 1993) and for the impact of classroom emotional climate and teacher management upon effectiveness. A negative emotional climate usually shows negative correlations, but a neutral climate is at least as supportive as a warm climate. Beyond an optimal level, teacher direction, drill, or recitation becomes dysfunctional (Soar & Soar, 1979). Rosenshine (1971) suggests inverted-U curvilinear relationships with student learning for verbal praise, difficulty level of instruction, teacher questions, and amount of student talk. The possibility of interaction with student individual differences is also supported. Therefore, the dynamic model of EER is based on the assumption that the relation of some effectiveness factors with achievement may be curvilinear. This assumption has implications for both the design and the analysis of effectiveness studies.

Second, there is a need to carefully examine the relationships between the various effectiveness factors. Walberg's (1984) model is one of the most significant educational productivity models which attempts to illustrate such relationships. Aptitude, instruction, and the psychological environment are seen as major direct causes of learning. They also influence one another and are in turn influenced by feedback on the amount of learning that takes place. Walberg's model was tested as a structural equation model on science achievement, indicating more complex, indirect relationships (Reynolds & Walberg, 1990). This implies that there is a need to develop a dynamic model of effectiveness revealing the relationships between the effectiveness factors which operate at the same level. Such approach to modelling educational effectiveness might reveal optimal combinations of factors that make teachers and schools effective.

Finally, the current models of EER do not explicitly refer to the measurement of each effectiveness factor. On the contrary, it is often assumed that these factors represent unidimensional constructs. For example, the comprehensive model of educational effectiveness states that there should be control at school level, meaning that goal attainment and the school climate should be evaluated (Creemers, 1994). In line with this assumption, studies investigating the validity of the model revealed that schools with an assessment policy focused on the formative purposes of assessment are more effective (e.g., Kyriakides, 2005; Kyriakides et al., 2000). However, the examination of assessment policy at school level can be examined not only in terms of its focus on the formative purpose but also in terms of many other aspects of the functioning of assessment such as the procedures used to design assessment instruments, the forms of record keeping, and the policy on reporting results to parents and pupils. This implies that the dynamic model of EER should not only illustrate the various effectiveness factors but also identify the dimensions upon which each factor can be measured. Considering effectiveness factors as multidimensional constructs provides a better picture of what makes teachers and schools effective and helps us develop specific strategies for improving educational practice.

A Proposed Dynamic Model of EER

The critical review of EER reveals the starting points for developing a dynamic model of EER. It has been argued that the dynamic model should take into account the new goals of education and related to this their implications for teaching and learning. This means that the outcome measures should be defined in a more broad way rather than be restricted to the achievement of basic skills. It also implies that new theories of teaching and learning can be used in order to specify variables associated with the quality of teaching. Moreover, the models of EER should be established in a way that helps policy-makers and practitioners to improve educational practice by taking rational decisions concerning the optimal fit of the factors within the model and the present situation in the schools or educational systems. Finally, the model should not only be parsimonious but also be able to describe the complex nature of educational effectiveness. This implies that the model could be based on specific theory but at the

same time some of the factors included in the major constructs of the model are interrelated within and/or between levels.

Based on the rationale of the dynamic model presented above, the essential characteristics of the model are as follows. First, the model refers to multiple factors of effectiveness which operate at different levels. Second, it is expected that some factors which operate at the same level are related to each other. It is therefore important to specify groupings of factors. Third, although there are different effectiveness factors and groupings of factors, it is assumed that each factor can be defined and measured using similar dimensions. This is a way to consider each factor as a multidimensional construct and at the same time to be in line with the parsimonious nature of the model. Finally, the model is designed in a way that takes into account the possibility that the relationship between the factors and the outcomes may not be linear. This refers to the possibility of searching for optimal values of the various dimensions of the factors and optimal combinations between factors.

The comprehensive model of educational effectiveness (Creemers, 1994) is in line with at least two of the starting points upon which the dynamic model is based. Creemers (1994) developed Carroll's (1963) model of learning by identifying three components within quality of instruction: curricular materials, grouping procedures, and teacher behaviour. However, there is an essential difference between Carroll's and Creemers' model. Carroll's model explains why students perform differently in handling a task, whereas Creemers' model ultimately explains why educational systems perform differently. Specifically, the comprehensive model of educational effectiveness is based on the assumption that the influences on student achievement are multilevel and, thereby, it refers to factors at different levels (i.e., student, classroom, school, system) which are related with student achievement. Direct and indirect relations between the levels and the outcomes are also identified. The proposed dynamic model also assumes that these relations may not be necessarily linear and that factors which operate at the same level may also be related to each other. Finally, in principle, each factor which refers to the classroom, school, and system can be measured by taking into account the following five dimensions: frequency, focus, stage, quality, and differentiation. Frequency is a quantitative way to measure the functioning of each effectiveness factor, whereas the other four dimensions examine the qualitative characteristics of the functioning of each effectiveness factor at the system, school, or classroom level. The importance of taking each dimension into account is discussed below, but it is, also, acknowledged that studies investigating the validity of the proposed measurement framework of effectiveness factors are needed.

Dimensions of Measuring Effectiveness Factors

First, the *frequency* dimension refers to the quantity that an activity associated with an effectiveness factor is present in a system, school, or classroom. This is probably the easiest way to measure the effect of a factor on student achievement and almost all studies used this dimension to define effectiveness factors. However, this dimension

may not always be related in a linear way with student outcomes. For example, personal monitoring at school level can be measured by taking into account how often the principles use a monitoring system to supervise their teachers. EER could attempt to identify whether this dimension of measuring personal monitoring is related not only directly to student outcomes but also indirectly through teacher behaviour in the classroom. Further, it is questionable that there is a linear relation between frequency of personal monitoring and both type of outcomes. On the contrary, it can be assumed that after an optimal value of using a monitoring system, this factor may not have an additional effect on outcomes but even may lead to a negative effect in teacher behaviour and ultimately in student outcomes.

Second, the factors are measured by taking into account the *focus* of the activities which reveals the function of the factor at classroom, school, and system level. Two aspects of focus of each factor are measured. The first one refers to the specificity of the activities, which can range from specific to general. For example, in the case of school policy on parental involvement, the policy could either be more specific in terms of concrete activities that are expected to take place (e.g., it refers to specific hours that parents can visit the school) or more general (e.g., it informs parents that they are welcome to the school but without giving them specific information about what, how, and when). The second aspect of this dimension addresses the purpose for which an activity takes place. An activity may be expected to achieve a single or multiple purposes. In the case of school policy on parental involvement, the activities might be restricted to a single purpose (e.g., parents visit schools to get information about student progress). On the other hand, the activities might be addressed to more than one purpose (e.g., parents visit the school to exchange information about children's progress and to assist teachers in and outside the classroom). It is expected that the measurement of the focus of an activity, either in terms of its specificity or in terms of the number of purposes that is expected to be achieved, may be related in a curvilinear way with student outcomes. For example, the guidelines on parental involvement, which are very general, may not be helpful either for parents or teachers in establishing good relations which can result in supporting student learning. On the other hand, a school policy which is very specific in defining activities may restrict the productive involvement of teachers and parents in creating their own ways for implementing the school policy. Similarly, if all the activities are expected to achieve a single purpose, then the chances to achieve this 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. On the other hand, if all the activities are expected to achieve multiple purposes, there is a danger that specific purposes are not addressed in such a way that they can be implemented successfully. This example also points to the possibility that an interaction between the two aspects of this dimension may exist.

Third, the activities associated with a factor can be measured by taking into account the *stage* at which they take place. It is expected 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. For example, school policy on opportunity to learn, which refers to

policy on cancellation of lessons and absenteeism, is expected to be implemented throughout the year and not only through specific regulations announced at a specific point of time (e.g., only at the beginning of the school year). It is also expected that the continuity will be achieved when the school is flexible in redefining its own policy and adapting the activities related to the factor by taking into account the results of its own self-evaluation mechanism. Measuring the stage dimension gives information about the continuity of the existence of a factor but the activities associated with the factor may not necessarily be the same. Finally, the importance of using the stage dimension to measure each effectiveness factor arises from the fact that it has been shown that the impact of a factor on student achievement partly depends on the extent to which activities associated with this factor are provided throughout the school career of the student (e.g., Creemers, 1994; Slater & Teddlie, 1992). 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.

Fourth, the dimension *quality* can be discerned in two different ways. The first one refers to the properties of the specific factor itself, as these are discussed in the literature. For instance, school policy on assessment can be measured by looking at the mechanisms which have been developed in order to establish instruments which meet psychometric standards (e.g., valid, reliable, representative to the content taught). At the same time, this policy makes clear and guarantees that teachers are expected to make use of the information gathered from assessment in order to meet their student's needs and this gives more emphasis to the formative function of assessment (Black & Wiliam, 1998; Harlen & James, 1997; Kyriakides et al., 2000).

Finally, the dimension *differentiation* refers to the extent to which activities associated with a factor are implemented in the same way for all the subjects involved with it. The importance of treating differentiation as a separate dimension of measuring effectiveness factors arises from the fact that students of any age and in any culture will differ from one another in various intellectual and psychomotor skills, in both generalised and specialised prior knowledge, in interests and motives, in their socioeconomical background, and in personal styles of thoughts and work during learning (Dowson & McInerney, 2003). Researchers in the area of educational effectiveness have shown that these differences are related to differences in students' learning progress (e.g., Creemers, 1994; Kyriakides, 2005; Slavin, 1987; Teddlie & Reynolds, 2000). These relations imply individual predispositions that somehow condition student readiness to profit from the particular instructional environments provided. Despite the fact that educational practice remained basically fixed and non-adaptive in most countries, research into differentiated effectiveness seems to reveal that teachers and schools may be differentially effective in promoting the learning of different groups of students (Campbell et al., 2004).

It is also important to note that, although this dimension could be considered as part of the quality dimension, special emphasis to differentiation is given due to the fact that TER has shown that adaptation to specific needs of each subject or group of subjects increases the successful implementation of a factor and ultimately maximises its effect on student learning outcomes. For example, instructional leadership is not

equally important for all the teachers of a school. Principles are, therefore, expected to adapt their leadership to the specific needs of the teachers by taking into account the extent to which they are ready to implement a task. Similarly, policy-makers are expected to adapt their general policy to the specific needs of groups of schools. The differentiation dimension does not necessarily imply that the subjects are not expected to achieve the same purposes. On the contrary, adapting the policy to the special needs of each group of schools, teachers, or students may ensure that all of them will become able to achieve the same purposes.

Above we have described in a more general way the five dimensions which can be used to measure each effectiveness factor. The examples which were given referred to factors at school and system levels. This was deliberately done in order to acknowledge the importance of establishing a comprehensive dynamic model which refers to effectiveness factors at all levels. However, in order to explain better how these five dimensions can be used to establish such a model, the second part of this section refers to the measurement of eight effectiveness factors concerned with teacher behaviour in the classroom. The choice made for the classroom level is based on the fact that studies on EER show that this level is more significant than the school and the system level (e.g., Kyriakides et al., 2000; Yair, 1997) and defining factors at the classroom level is seen as a prerequisite for defining the school and the system level (Creemers, 1994).

Specification for the Teacher Behaviour at the Classroom Level

Based on the main findings of TER (e.g., Brophy & Good, 1986; Campbell et al., 2004; Creemers, 1994; Kyriakides, Campbell, & Christofidou, 2002; Muijs & Reynolds, 2001; Rosenshine, 1983), our dynamic model refers to eight effectiveness factors which describe the teacher's instructional role and were found to be consistently related with student outcomes. Two of them (i.e., orientation, teaching modelling) are in line with the principles of teaching for understanding and promote the achievement of the new goals of education such as the development of metacognitive skills.

Orientation. Orientation refers to teacher behaviour in providing the objectives for which a specific task or lesson or series of lessons take(s) place and/or challenging students to identify the reason for which an activity takes place in the lesson. It is expected that the engagement of students with orientation tasks might encourage them to actively participate in the classroom since the tasks that take place are meaningful for them. As a consequence, the dimension frequency is measured by taking into account the number of orientation tasks that take place in a typical lesson, as well as how long each orientation task takes place. These two indicators may help us identify the importance that the teacher attached to this factor. As far as the focus dimension is concerned, it is possible that an orientation task may refer to a part of a lesson, to the whole lesson, or even to a series of lessons (e.g., a lesson unit). This classification refers to the specificity of the orientation task. The second aspect of

focus which refers to the purpose of the activity can be measured by examining the extent to which an activity is restricted to finding one single reason for doing a task or finding the multiple reasons for doing a task. The measurement of this dimension reveals the extent to which teachers help their students understand the importance of finding the meanings of each task they are expected to be involved in. The third dimension of measuring orientation refers to the stage at which an activity takes place. It is expected that orientation tasks will take place in different parts of a lesson or series of lessons (e.g., introduction, core, ending of the lesson). Further, it is assumed that effective teachers are able to take other perspectives into account during these orientation tasks. For example, students may come with suggestions for the reasons of doing a specific task, which an effective teacher should take into account. Fourth, the measurement of quality refers to the properties of the orientation task and especially whether it is clear for the students. It also refers to the impact that the task has on student engagement in the learning process. For example, teachers may present the reasons of doing a task simply because they have to do it and because it is part of their teaching routine without having much effect on student participation, whereas others may encourage students to identify the purposes that can be achieved by doing a task and therefore to increase their motivation towards a specific task, lesson, or series of lessons. Finally, differentiation is measured in a similar way for each of the eight factors. In the case of orientation, teachers are expected to provide different types of orientation tasks to students according to their learning needs.

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

First, the dimension frequency is measured in a similar way as in the case of orientations. The two indicators that can be used are the number of tasks that take place in a typical lesson, as well as how long each task takes place (e.g., the percentage of teaching time spent on structuring). Second, the focus dimension is measured by taking into account that a structuring task may either refer to a part of a lesson, to the whole lesson, or even to a series of lessons (e.g., a lesson unit). As far as the second aspect of focus is concerned, a structuring task may refer to the achievement of a single objective or to the relation of the elements of the lesson in relation to multiple objectives. It is expected that the structuring tasks which have an impact on student behaviour are those which refer to the achievement of multiple objectives, since the

tasks which refer to a single objective may increase the fragmentation of the learning process. The third dimension of measuring structuring which refers to the stage at which an activity takes place is also measured in the same way as orientation. Structuring tasks may take place in different parts of a lesson or series of lessons (e.g., introduction, core, ending of the lesson). Fourth, the dimension of quality is measured by examining whether structuring tasks are not only clear for the students but also help them understand the structure of the lesson. For this reason, we do not measure clarity as a property of structuring nor as an independent factor of teacher effectiveness, but clarity is seen as a condition for helping students to understand the structure and the content of a lesson or series of lessons. Quality is also measured by investigating the extent to which teachers organise their lessons or series of lessons in a way to move from easier tasks to those more complicated. Finally, in the case of structuring, differentiation is measured by investigating the extent to which teachers provide different types of structuring tasks to students according to their learning needs.

Questioning techniques. Muijs and Reynolds (2000) indicate that the focus of TER on the teacher actively presenting materials should not be seen as an indication that traditional lecturing and drill approach is an effective teaching approach. Effective teachers ask a lot of questions and attempt to involve students in class discussion. Although the data on cognitive level of question yield inconsistent results (Redfield & Rousseau, 1981), optimal question difficulty is expected to vary with context. There should also be a mix of product questions (i.e., expecting a single response from students) and process questions (i.e., expecting students to provide explanations) but effective teachers ask more process questions (Askew & William, 1995; Everston, Anderson, Anderson, & Brophy, 1980). Therefore, the frequency dimension has to be measured through different aspects. The total number of questions and the ratio between process and product questions are two major indicators of this dimension. Another dimension has to do with length of pause following questions, which is expected to vary according to the difficulty level of questions. Brophy and Good (1986) point out that a question calling for application of abstract principles should require a longer pause than a factual question. Focus is measured by looking at the type of question and especially its relation with the tasks that take place during a lesson (i.e., specificity), as well as with the objectives that are expected to be met. As far as the measurement of stage is concerned, it is taken into account that teachers may raise questions at different parts of the lesson and for different reasons. For example, teachers may ask questions in the introduction of the lesson in order to link the new lesson with previous lessons and/or during the core of the lesson in order to discover problem(s) that students have with the content of the lesson or need(s) for further clarifications. Questions may also be raised at the end of the lesson as part of the attempt of the teacher to assess students for formative reasons.

Quality is measured by taking into account the clarity of a question and especially the extent to which students understand what they are expected to find out. Another property that also can be measured is the appropriateness of the difficulty level of the

question, since it is possible that students may understand the question and still do not answer because it is too difficult for them. Finally, quality is measured by investigating the way the teacher deals with student responses to his or her questions. Correct responses should be acknowledged for other students' learning, while responses that are partly correct require affirmation of the correct part, and rephrasing of the question (Brophy & Good, 1986; Rosenshine & Stevens, 1986). Following incorrect answers, teachers should begin by indicating that the response is not correct but avoid personal criticism and show why the correct answer is correct (Rosenshine, 1971). Finally, differentiation is measured by looking at the extent to which teachers direct questions to specific students or take answers from specific students. It is also assumed that the feedback that effective teachers give to student answers varies according to their needs.

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 taken more attention during the last decade due to the emphasis given in policy on the achievement of new goals of education. Thus, TER has shown that effective teachers are expected to help pupils to use strategies and/or develop their own strategies which can help them solve different types of problems (Kyriakides et al., 2002). As a result of this, it is more likely that students will develop skills that help them organise their own learning (e.g., self-regulation, active learning). Thus, the frequency dimension of teaching modelling can be measured by looking at the number of teaching modelling tasks that take place in a lesson and the teaching time devoted to them. As far as the focus is concerned, teaching modelling tasks can be examined in relation to the extent to which they refer to strategies which can be used to solve problems under various conditions (e.g., problems of different subjects). This measure refers to the specificity aspect of this dimension. Moreover, focus can be seen in relation to the extent to which teachers provide opportunities to students to use or develop more than one strategy to solve specific problems or types of problems. Third, the stage dimension is concerned with the sequence under which a teaching modelling task is used in the classroom. It is possible that initially students are faced with a problem and then are expected to use or develop a particular strategy to solve it. On the other hand, teachers may teach a strategy or different strategies to students and then students are asked to use these strategies in order to solve a problem.

Fourth, the measure of the quality deals with the properties of teaching-modelling tasks and especially with the role that the teacher is expected to play in order to help students use a strategy to solve their problems. Teachers may either present a strategy with clarity or they may invite students to explain how they solve a problem and use that information for promoting the idea of modelling. The latter may encourage students not only to use but also to develop their own strategies for solving problems. Quality is also measured by looking at the impact that an activity has on student behaviour. Students may either become able to use a strategy in an effective way (i.e., finding the solution of the problem) or the use of the strategy may become an obstacle

in dealing with a problem (e.g., causes more confusion about the problem). Finally, differentiation can be seen in terms of adopting teaching modelling to specific needs of a group of students. These might result in more emphasis on applying a single strategy for a group of students to solve problems or more emphasis on using multiple strategies or even develop new strategies for other groups of students.

Application. Effective teachers also use seatwork or small group tasks since they provide needed practice and application opportunities (Borich, 1992) and can be linked to a direct teaching model (Rosenshine, 1983) which emphasises immediate exercise of topics taught during the lesson. Thus, the frequency can be measured by looking at the total time devoted to application tasks (e.g., percentage of teaching time). Focus can be measured by looking at the specificity of the tasks that students are expected to perform. We can examine, therefore, the extent to which the application tasks refer to some parts of the lesson, to the whole lesson, or even to a series of lessons. This way of measurement is also related to the second aspect of focus since it enables us to examine the number of purposes that application tasks cover. Stage is measured by looking at the phase of the lesson at which each application task takes place. The quality of application tasks is measured by investigating whether students are simply asked to repeat what they have already covered with their teacher or whether the application task is more complex than the content covered in the lesson. It is also examined whether the application tasks are used as starting points for the next step of teaching and learning. Finally, differentiation refers to the extent to which teachers give more opportunities for application to students who need them. It also refers to teacher behaviour in monitoring and supervising and giving corrective feedback during application activities. Brophy and Good (1986) argue that once the students are released to work independently, effective teachers circulate to monitor progress and provide help and feedback.

The classroom as a learning environment: The contribution of the teacher. Muijs and Reynolds (2000) point out that classroom climate is a factor that TER has found to be significant. The climate is usually seen as associated with the behaviour of the stakeholders, whereas culture is seen as measuring the values and norms of the organisation (Heck & Marcoulides, 1996; Hoy, 1990). It is supported that a healthy organisation deals effectively with outside forces while directing its energies towards its goals. Classroom climate research is described as the stepchild of psychological and classroom research (Creemers & Reezigt, 1996). The classroom effects research tradition initially focused on climate factors defined as managerial techniques (e.g., Doyle, 1986). Management is necessary to create conditions for learning and instruction, but management itself is not sufficient for student results (Creemers, 1994). On the other hand, the psychological tradition of classroom environment research paid a lot of attention to instruments for the measuring of students' perceptions of climate. Many studies report on their psychometric characteristics (Fraser, 1991), but climate factors (such as the way a teacher behaves towards the

students) and effectiveness factors (e.g., quality of teaching) were studied as isolated constructs (Johnson & Johnson, 1993; Wubbels, Brekelmans, & Hooymayers, 1991). In this context, EER has to take the first steps to integrate elements of different research traditions. Thus, in this stage of the study we concentrate on measuring teacher contribution in creating a learning environment in his or her classroom and five elements of 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. The first two elements are important components of measuring classroom climate as classroom environment research has shown (Cazden, 1986; Den Brok, Brekelmans, & Wubbels, 2004; Fraser, 1991), but in this study we concentrate on the type of interactions that exist in a classroom rather than on how students perceive teacher interpersonal behaviour. The other three elements refer to the attempt of teachers to create a businesslike and supportive environment for learning (Walberg, 1986) and classroom effectiveness research has shown their importance in promoting student learning (Brophy & Good, 1986; Hextall & Mahony, 1998; Scheerens & Bosker, 1997). The ways used to measure these five elements are very similar and are presented below.

Interactions are measured by taking into account the role of the teacher in establishing interaction between students and between students and himself or herself. The dimension frequency refers to the number and type of interactions which take place. Specifically, interactions are classified into different types based on their focus (i.e., specificity and the purpose(s) it serves). For example, interactions are classified according to the purpose(s) they are expected to serve (e.g., managerial reasons, learning, social encounter). As far as the stage is concerned, interactions are seen in relation to the phase of the lesson at which they take place. Quality is only measured by looking at the immediate impact that teacher initiatives have on establishing relevant interactions. We are mainly interested to see the extent to which a teacher is able to establish on task behaviour through the interactions she or he promotes since Creemers' model emphasises the importance of keeping students on task (Creemers, 1994). Finally, differentiation is measured by looking at the different teaching strategies the teacher is able to use in order to keep different groups of students involved in the classroom interactions which promote student learning.

As far as the other three elements of classroom as a learning environment are concerned, they are measured by taking into account the teacher behaviour in establishing rules, persuading students to respect and use the rules, and maintaining them in order to create a learning environment in their classroom. The first element refers to more general problems that can arise when students do not believe that they are treated fairly and are respected as individual persons by their teacher, whereas the other two deal with specific situations in the classroom which might create difficulties in promoting learning (i.e., competition between students, classroom disorder). Thus, frequency is measured by looking at the number of problems that arise in the classroom (e.g., classroom disorder: fight between two students) and the various ways that teachers use to deal with them. Focus is measured by looking at the specificity of the problem that is observed (e.g., incidental or a continuous one that takes the

classroom back to problems that were not solved successfully), as well as to the reaction of the teacher in terms of the purpose(s) that he or she attempts to achieve (e.g., solving only the specific problem or creating an atmosphere that avoids the further existence of similar problems). For example, in the case of investigating the way teachers deal with negative effects of competition, the teacher can either deal with the specific problem that arises or put the problem in a more general perspective in order to help students see the positive aspects of competition and avoid the negative ones. Stage can be measured by looking at the phase of the lesson at which the problem arises. Quality is seen in relation to the impact that the teacher behaviour has on solving the problems that arise as measured through students' behaviour. For example, a teacher may not use any strategy at all to deal with a classroom disorder problem, or use a strategy but the problem is only temporarily solved, or use a strategy that has a long-lasting effect. Finally, differentiation is measured by looking at the extent to which teachers use different strategies to deal with problems which are caused by different groups of students. For example, individual student(s) might cause a problem in order to get attention from classroom mates and/or the teacher. It is probably a better strategy not to pay attention when the problem is small, since any reaction from the teacher may promote the continuation of causing problems.

Management of time. Creemers' model considers opportunity to learn and time on task as two of the most significant factors of effectiveness that operate at different levels. Opportunity to learn is also related to student engagement and time on task (Emmer & Everston, 1981). Therefore, effective teachers are expected to organise and manage the classroom environment as an efficient learning environment and thereby to maximise engagement rates (Creemers & Reezigt, 1996). In this study, management of time is considered as one of the most important indicators of teacher ability to manage the classroom in an effective way. Thus, frequency is measured by taking into account how much time is used for teaching per lesson and how much time is covered within the time framework. Focus dimension is not measured separately since the main interest of this factor is whether students are on task or off task. Stage is measured by taking into account time attribution to different phases of the lesson. As far as the quality dimension is concerned, this is measured through the data collected in relation to the factor concerning the role of teacher in creating a learning environment in his or her classroom. Finally, differentiation is measured by looking at the allocation of time for different groups of students.

Teacher evaluation. Evaluation is seen as an integral part of teaching (Stenmark, 1992) and especially formative evaluation is one of the most important factors associated with effectiveness at all levels, especially at the classroom level (e.g., De Jong et al., 2004; Kyriakides, 2005; Shepard, 1989). Information gathered from assessment can be used in order to enable teachers to identify their students' needs, as well as to evaluate their own practice. In this study, frequency is measured in terms of the number of evaluative tasks and the time that they take place. It is expected that there is a curvilinear relation between the frequency of teacher evaluation and student

outcomes, since an overemphasis on evaluation might reduce the actual time spent on teaching and learning, whereas teachers who do not collect any information are not able to adapt their teaching to student needs. Focus is measured by looking at the ability of the teacher to use different ways of measuring student skills rather than using only one technique (e.g., written tests). It is also important to examine whether the teacher makes more than one use out of the information that she or he collects (e.g., identifying needs of students, conducting self-evaluation, adapting his or her long-term planning, using evaluation tasks as a starting point for teaching). Stage is measured in terms of the period at which the evaluation tasks take place (e.g., at the beginning, during, and at the end of a lesson or unit of lessons) and the time lag between collecting information, recording the results, reporting the results to students and parents, and using them. Quality is measured by looking at the properties of the evaluation instruments used by the teacher such as the validity, the reliability, the practicality, and the extent to which the instruments cover the teaching content in a representative way. This dimension is also measured by investigating the type of feedback that the teacher gives to the students and the way students use the teacher feedback. Specifically, effective teachers provide constructive feedback which has positive implications on teaching and learning (Muijs & Reynolds, 2001). Finally, differentiation is examined in relation to the extent to which teachers use different techniques for measuring student needs and/or different ways to provide feedback to groups of students by taking into account their needs.

Developing the Dynamic Model at the School and System Level

In this paper, we have outlined a dynamic model that takes into account the new goals of education and the importance of illustrating the multilevel and complex nature of effectiveness. Examples of measuring effectiveness factors operating at different levels using five dimensions concerning the frequency, focus, stage, quality, and differentiation of activities associated with each factor were given. This helps us to illustrate the dynamic nature of the integrated model which has to be developed and tested in order to explain in a better way variances in student achievement at the different levels. However, we pay more attention to describing in detail factors associated with teacher behaviour in the classroom since this was seen as the starting point for the development and the testing of the dynamic model. Thus, suggestions for the next steps in the development of other parts of the model are provided.

It has also been argued that the model could be developed by identifying factors at the school and system level which EER reveals that are associated with student achievement gains. However, the selection of these factors cannot only be based on a combination of the existing integrated models and especially those which arise from organisational theories (Scheerens, 1992). Factors at the school and system level that are related to classroom factors should mainly be examined. For instance, it is important to take into account school policy on student assessment, since it is expected to have an effect on one of the classroom-level factors of the model which refers to teacher's ability to evaluate his or her student. Moreover, the possibility that

these factors have curvilinear relations with student achievement should be considered. Finally, it is important not to treat the school and the system level factors as unidimensional constructs but, as has been shown above, the five dimensions of the model can be applied in order to better define each of them.

Given that the model refers to five dimensions that can be used to measure each of the eight TER factors and assumes that curvilinear relations between some of these factors and student achievement may exist, it is important to test the validity of this model. Results of validation studies may contribute to the establishment of a less complex model which is able to illustrate the different strategies that can be used for improving educational practice. In this context, a longitudinal study is currently undertaken in order to test the validity of the proposed measurement framework of effectiveness factors and identify the type(s) of relations that each effectiveness factor and its dimensions have with student learning outcomes (Kyriakides & Creemers, 2006). Currently, the emphasis is on developing and testing the model, but it is expected that the dynamic model of EER will reveal the complexity of educational effectiveness and may help us establish strong links between EER and improvement practices.

References

- Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society (Series A)*, 149(1), 1–43.
- Askew, M., & William, D. (1995). *Recent research in mathematics education 5–16*. London: Office for Standards in Education.
- Black, P., & William, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: King's College London School of Education.
- Borich, G. D. (1992). *Effective teaching methods* (2nd ed.). New York: Macmillan.
- Bosker, R. J. (1990). Theory development in school effectiveness research: In search for stability of effects. In P. van de Eedem, J. Hox, & J. Hauer (Eds.), *Theory and model in multilevel research: Convergence or divergence?* (pp. 77–98). Amsterdam: SISWO.
- 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.
- Brown, B. W., & Saks, D. H. (1986). Measuring the effects of instructional time on student learning: Evidence from the beginning teacher evaluation study. *American Journal of Education*, 94, 480–500.
- Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education*, 29(3), 347–362.
- Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2004). *Assessing teacher effectiveness: A differentiated model*. London: Routledge Falmer.
- Carroll, J. B. (1963). A model of school learning. *Teacher College Record*, 64, 723–733.
- Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 432–463). New York: Macmillan.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. New York: Rand McNally.
- Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.
- Creemers, B. P. M. (2002). From school effectiveness and school improvement to effective school improvement: Background, theoretical analysis, and outline of the empirical study. *Educational Research and Evaluation*, 8, 343–362.

- Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement*, 7, 197–228.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1). Retrieved March 16, 2005, from <http://epaa.asu.edu/epaa/v8n1/>
- De Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in Mathematics in the first year of junior general education in the Netherlands. *School Effectiveness and School Improvement*, 15, 3–31.
- DeLeeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11, 57–85.
- Den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement*, 15, 407–442.
- Dowson, M., & McInerney, D. M. (2003). What do students say about motivational goals?: Towards a more complex and dynamic perspective on student motivation. *Contemporary Educational Psychology*, 28(1), 91–113.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 392–431). New York: Macmillan.
- Elberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principles make a difference? *Economics Education Review*, 7, 291–299.
- Emmer, E. T., & Everston, C. M. (1981). Synthesis of research on classroom management. *Educational Leadership*, 38(4), 342–347.
- Everston, C. M., Anderson, C., Anderson, L., & Brophy, J. (1980). Relationships between classroom behaviour and student outcomes in junior high math and English classes. *American Educational Research Journal*, 17, 43–60.
- Fraser, B. J. (1991). Two decades of classroom environment research. In B. J. Fraser & H. J. Walberg (Eds.), *Educational environments: Evaluation, antecedents and consequences* (pp. 3–29). Oxford, UK: Pergamon.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Oxford University Press.
- Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Edward Arnold.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141–1177.
- Hanushek, E. A. (1989). The impact of differential expenditures on student performance. *Educational Research*, 66(3), 397–409.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3), 365–379.
- Heck, R. A., & Marcoulides, G. A. (1996). School culture and performance: Testing the invariance of an organizational model. *School Effectiveness and School Improvement*, 7, 76–106.
- Hedges, L. V., Laine, R., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(3), 5–14.
- Hextall, L., & Mahony, P. (1998). *Effective teachers effective schools*. London: Biddles.
- Hopkins, D., Reynolds, D., & Gray, J. (1999). Moving on and moving up: Confronting the complexities of improvement. *Educational Research and Evaluation*, 5, 22–40.
- Hoy, W. K. (1990). Organisational climate and culture: A conceptual analysis of the school workplace. *Journal of Educational and Psychological Consultation*, 1(2), 149–168.
- Johnson, D. W., & Johnson, R. T. (1993). Cooperative learning and classroom and school climate. In B. J. Fraser & H. J. Walberg (Eds.), *Educational environments: Evaluation, amendments and consequences* (pp. 55–75). Oxford, UK: Pergamon.
- Knuver, A. W. M., & Brandsma, H. P. (1993). Cognitive and affective outcomes in school effectiveness research. *School Effectiveness and School Improvement*, 13, 187–200.

- Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement*, 16, 103–152.
- Kyriakides, L., Campbell, R. J., & Christofidou, E. (2002). Generating criteria for measuring teacher effectiveness through a self-evaluation approach: A complementary way of measuring teacher effectiveness. *School Effectiveness and School Improvement*, 13, 291–325.
- Kyriakides, L., Campbell, R. J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of Creemers comprehensive model of educational effectiveness. *School Effectiveness and School Improvement*, 11, 501–529.
- Kyriakides, L., & Creemers, B. P. M. (2006, April). *Testing the dynamic model of educational effectiveness: Teacher effects on cognitive and affective outcomes*. Paper presented at the 87th annual meeting of the American Educational Research Association, San Francisco.
- Kyriakides, L., & Tsangaridou, N. (2004, April). *School effectiveness and teacher effectiveness in physical education*. Paper presented at the 85th annual meeting of the American Educational Research Association, Chicago.
- Levine, D. U., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, WI: National Center for Effective Schools Research and Development.
- Lingard, B., Ladwig, J., & Luke, A. (1998). School effects in post-modern conditions. In R. Slee & G. Weiner (with S. Tomlinson) (Eds.), *School effectiveness for whom, challenges to the school effectiveness and school improvement movements* (pp. 84–100). London: Falmer Press.
- Monk, D. H. (1992). *Microeconomics of school productions*. Paper for the Economics of Education Section of the International Encyclopedia of Education.
- Monk, D. H. (1994). Subject matter preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters*. Wells, UK: Open Books.
- Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in Mathematics: Some preliminary findings from the evaluation of the Mathematics Enhancement Programme (primary). *School Effectiveness and School Improvement*, 11, 273–303.
- Muijs, D., & Reynolds, D. (2001). *Effective teaching: Evidence and practice*. London: Sage.
- Opdenakker, M. C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement*, 11, 65–196.
- Pines, A. L., & West, L. H. T. (1986). Conceptual understanding and Science learning: An interpretation of research within a source-of-knowledge framework. *Science Education*, 70(5), 583–604.
- Prawat, R. S. (1989). Teaching for understanding: 3 key attributes. *Teaching and Teacher Education*, 5(4), 315–328.
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1–17.
- Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, 15, 423–476.
- Redfield, D., & Rousseau, E. (1981). A meta-analysis of experimental research on teacher questioning behaviour. *Review of Educational Research*, 51, 237–245.
- Reynolds, A. J., & Walberg, H. J. (1990). *A structural model of educational productivity*. DeKalb, IL: Northern Illinois University.
- Rosenshine, B. (1971). *Teaching behaviors and student achievement*. London: National Foundation for Educational Research.
- Rosenshine, B. (1983). Teaching functions in instructional programs. *Elementary School Journal*, 89, 421–439.
- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 376–391). New York: Macmillan.
- Scheerens, J. (1992). *Effective schooling: Research, theory and practice*. London: Cassell.

- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.
- Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26(3), 207–231.
- Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education*, 8, 151–184.
- Shepard, L. A. (1989). Why we need better assessment. *Educational Leadership*, 46(2), 4–8.
- Slater, R. O., & Teddlie, C. (1992). Towards a theory of school effectiveness and leadership. *School Effectiveness and School Improvement*, 3, 242–257.
- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293–336.
- Slee, R., & Weiner, G. (with Tomlinson, S.) (Eds.). (1998). *School effectiveness for whom? Challenges to the school effectiveness and school improvement movements*. London: Falmer Press.
- Soar, R. S., & Soar, R. M. (1979). Emotional climate and management. In P. Peterson & H. Walberg (Eds.), *Research on teaching concepts: Findings and implications*. Berkeley, CA: McCutchan.
- Stenmark, J. K. (1992). *Mathematics assessment: Myths, models, good questions and practical suggestions*. Reston, VA: NCTM.
- Stevenson, H. W., Chen, C., & Lee, S. Y. (1993). Mathematics achievement of Chinese, Japanese and American children: Ten years later. *Science*, 259, 53–58.
- Stringfield, S. C., & Slavin, R. E. (1992). A hierarchical longitudinal model for elementary school effects. In B. P. M. Creemers & G. J. Reezigt (Eds.), *Evaluation of educational effectiveness* (pp. 35–69). Groningen, The Netherlands: ICO.
- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.
- Walberg, H. J. (1984). Improving the productivity of American schools. *Educational Leadership*, 41, 19–27.
- Walberg, H. J. (1986). Syntheses of Research on Teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 214–229). New York: Macmillan.
- Wubbels, T., Brekelmans, M., & Hooymayers, H. (1991). Interpersonal teacher behaviour in the classroom. In B. J. Fraser & H. J. Walberg (Eds.), *Educational environments: Evaluation, antecedents and consequences* (pp. 141–161). Oxford, UK: Pergamon.
- Yair, G. (1997). When classrooms matter: Implications of between-classroom variability for educational policy in Israel. *Assessment in Education*, 4(2), 225–248.