A CRITICAL ANALYSIS OF THE CURRENT APPROACHES TO MODELLING EDUCATIONAL EFFECTIVENESS: THE IMPORTANCE OF ESTABLISHING A DYNAMIC MODEL

B.P.M. CREEMERS¹ & L. KYRIAKIDES²
Faculty of Behavioural and Social Sciences, University of Groningen¹
Department of Education, University of Cyprus ²

ABSTRACT

This paper argues that researchers in the area of educational effectiveness should attempt to develop a new theoretical framework. Thus, 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 which 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 five 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. Thus, different methodological approaches that can be used to test this part of the model and suggestions for the next steps in the development of other parts of the model are made. Finally, suggestions for using the dynamic model in order to improve educational practice are provided.
HISTORY OF EDUCATIONAL EFFECTIVENESS THEORY AND RESEARCH:

SOME OBSERVATIONS

Stringfield (1994) defines educational effectiveness research as the process of differentiating existing ideas and methods along dimensions deemed to be of value. Educational effectiveness research (EER) does not attempt to invent new ideas or programmes but to concentrate on understanding the lessons to be drawn from existing practices. In this way, EER attempts to establish and test theories which explain why and how some schools and teachers are more effective than others. Most of the research is naturalistic in nature and gives us information about plausibility of theoretical notions but some experimental studies were also conducted in order to help us identify cause and effect relations. Neither of these approaches provides sufficient evidence to validate theories of effectiveness since experimental studies are weak in terms of their ecological validity (Miller, 1984; Robson, 1993) and inconsistent data emerged from these studies (Kyriakides, Campbell & Christofidou, 2002; Medley, 1979). Although some longitudinal studies have also been conducted, researchers should consider the possibility of using this research paradigm to test theories since this is methodologically a more appropriate way to do it (Magnusson, Bergman, Rudinger & Torestad, 1991).

The origins of educational effectiveness stem from reactions to the work on equality of opportunity undertaken by James Coleman and his collaborators (Coleman et al., 1966) and Christopher Jencks (Jencks et al., 1972). These two studies coming from two different disciplinary backgrounds (i.e., sociological and psychological) came almost to a similar conclusion in relation to the amount of variance that can be explained by educational factors. After taking into consideration student background characteristics, such as ability and family background not much variance in student achievement was left. This pessimistic feeling was also fed by the failure of large-scale educational compensatory programs such as the “Headstart” in the U.S.A. and comparable programmes in other countries (MacDonald, 1991; Schon, 1971).
Thus, two studies undertaken, independently, by Edmonds (1979) and Rutter et al. (1979) during the 1970s were concerned with examining evidence and making an argument about the potential power of schooling to make a difference to students' life chances. This was an optimistic point of view because many studies published in that period showed that teachers, schools, and maybe even education in general, did not make much difference at all. The early existence of independent research projects in two countries asking similar questions and drawing, to a certain extent, on similar methodologies demonstrated the potential for establishing a scientific domain dealing with effectiveness in education. Thus, the publications by Brookover et al. (1979) and Rutter et al. (1979) were followed by numerous studies in different countries into school effectiveness and school improvement efforts, which were aimed at putting the results of research into practice (Teddlie & Reynolds, 2000; Townsend, Clarke, & Ainscow, 1999).

In the last 25 years the research into educational effectiveness 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 teacher and school differences in student achievement to be obtained (Goldstein, 2003). There is also substantial agreement as to appropriate methods of estimating school differences/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., Mortimore et al., 1988; Scheerens, 1992; Levin & Lezotte, 1990). 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 and contribute in the establishment of strategies for improving educational effectiveness (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 meta-cognitive skills (Campbell et al., 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 departs 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. They see these narrow conceptions as historical artefacts of an industrial era, versions of schoolings and constructions of the modernist human subject. Their modernist conception of education suggests that effectiveness can be conceptualised and measured in terms of new kinds of citizens’ sensibilities, moral and cultural practises and indeed kinds of discourses and cultural productions that are generative and redistributive on new conditions, rather than simply reproductive of existing divisions of wealth discourse, gender and labour (Bage, 1997). 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, 2005a; Opdenakker & Van Damme, 2000). These studies also reveal that schools which are among the most effective in cognitive outcomes were also among the most effective in other domains (Kyriakides, 2005a). Therefore, the criteria of measuring effectiveness should arise from the goals of education as defined within a particular society and political context. At the same time 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 attempts to raise the
importance of establishing a dynamic model of EER which takes into account the new goals of education.

Finally, an important constraint of the existing approaches of EER is the fact that the whole process does not contribute significantly to the improvement of school effectiveness. Although in 1989 an *International Congress for School Effectiveness and Improvement* (ICSEI) was established together with the journal *School Effectiveness and School Improvement* and also projects attempting to establish links between effectiveness and improvement research were undertaken (e.g., the Effective School Improvement project), there are still problems in the relation between effectiveness and improvement (Creemers, 2002). The question persists how to apply the effectiveness knowledge base in practice, in other words how to get valid and useful information about school improvement out of educational effectiveness. It can be argued that there are still tensions between educational effectiveness, theory and research on one hand and school improvement on the other. Probably this will remain to be the case, but the tensions between the two have also led to further clarification about what is at stake. The development of a knowledge base about educational effectiveness certainly needs to be expanded, but it has to be said that school improvement is more than just application of the available knowledge base. It needs intermediate goals and careful research and evaluation about how the ultimate goals, such as student performance and the characteristics at school and classroom level (the so-called effective characteristics), are related to the objectives of the improvement policies.

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 at the same time defines the dynamic relations between the multiple factors found to be associated with effectiveness. Such approach to educational effectiveness modelling might help EER to establish stronger links with educational improvement practice since the proposed dynamic model of EER will reveal the complexity of the process of improving effectiveness. 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 conducting research in order to develop and test the validity of the model and for using the dynamic model in order to establish links between EER and improvement practices are drawn.

THE THEORETICAL MODELS OF EDUCATIONAL EFFECTIVENESS RESEARCH

In the literature of educational effectiveness modeling 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). Resource input variables such as student/teacher ratio, teacher salary and overall measures of per student expenditure were of primary interest in the earlier studies. Such research is focused on producing a function which could explain each pupil outcome at a given time, and which according to Hanushek (1979) has the following form:

\[ A_{it} = f(B_{it}, P_{it}, S_{it}, I_i) \]

where \( A_{it} \) is the outcome of the \( i \)th student at time \( t \), \( B_{it} \) is the vector of family background influences of the \( i \)th student cumulative at time \( t \), \( P_{it} \) is the vector of influence of peers of the \( i \)th student cumulative at time \( t \), \( S_{it} \) is the vector of school inputs of the \( i \)th student cumulative at time \( t \), and \( I_i \) is the vector of innate abilities of the \( i \)th student.

The function may be linear, consisting of main effects and interaction terms or non-linear (Brown & Saks, 1986). Therefore, the emerging “education production” models (e.g., Elberts & Stone, 1988; Brown & Saks, 1986) are based on the assumption that increased inputs will lead to increments in outcomes and their main characteristics are concerned with: a) the selection of resource inputs as the major type of selection of antecedent condition, b) the measurement of direct effects, and c) the use of data at only one level of aggregation {i.e., either at micro (e.g., student) level or aggregated (e.g., school) level}. It is, however,
important to note that 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. Therefore, an evidence-based policy making approach (Fitz-Gibbon, 1996) is needed in order to help policy-makers have a rational basis for taking decisions on how to improve education and at the same time keeping them accountable to the public.

The second approach to educational effectiveness modeling 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 on processes from two different perspectives concerning learning and school as organizations. 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 different 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 organizational theories such as the school climate, culture and structure and for contextual variables.

One of the most influential models emerged from this approach is the Carroll’s Model (Carroll, 1963) which states that the degree of mastery is a function of the ratio of amount of time students actually spend on learning tasks to the total amount of time they need. Carroll (1963) argued that time actually spent on learning is defined as equal to the smallest of three variables: a) opportunity (time allowed for learning), b) perseverance (the
amount of time students are willing to engage actively in learning), and c) aptitude (the amount of time needed to learn under optimal instructional conditions). Numerous studies and meta-analyses have confirmed the validity of the Carroll’s model. It was also the basis for Bloom’s concept of mastery learning (Bloom, 1968) and is also related to “direct instruction” as described by Rosenshine (1983). However, as Carroll (1989) pointed out 25 years after the construction of his model, the one factor in his original model that needed further elaboration was ‘quality of instruction’.

Making use of the principles of mastery learning and direct instruction, Creemers (1994) developed Carroll’s 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 the 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. For this reason, Creemers’ model is based on the assumption that the influences on student achievement are multilevel. His model can therefore be considered as a model of the third approach to educational effectiveness modelling, namely the generalist-educationalist approach. The models of this approach emerged by researchers attempt to integrate the findings of school effectiveness research, teacher effectiveness research and the early input-output studies. Thus, the resulting models (e.g., Stringfield & Slavin, 1992; Scheerens, 1992; Creemers, 1994) 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).

A CRITICAL ANALYSIS OF THE MODELS OF EER BASED ON EMPIRICAL EVIDENCE
Four studies, which have been conducted in order to test the validity of Creemers’ model in Netherlands (Jong, Westerhof, & Kruiter, 2004) and Cyprus (Kyriakides, Campbell, & Gagatsis, 2000; Kyriakides, 2005a; Kyriakides & Tsangaridou, 2004) 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 (Teddlie & Reynolds, 2000) and provides support to the argument that models of EER should be multi-level in nature. The analysis of the results of these studies reveal that next to the multi-level 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 et al., 2004). Although such interactions between school and classroom effectiveness factors have rarely been identified, this result can be attributed to difficulties in measuring sufficient variation between schools through national studies. International comparative studies such as the TIMSS study has shown that despite the criticisms of national educational policy in UK about achievement pressure for schools and teachers, a significant gain in student achievement in mathematics was found during the last five years (Mullis, Martin, Gonzalez, & Chrostowski, 2004). Therefore, the proposed dynamic models of EER should not be only multi-level in nature but should also 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 effect 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, EER should not simply provide a list of factors associated in a linear way with effectiveness but give a more accurate picture of what makes schools and teachers effective. As a consequence, the dynamic model of EER should be based on the assumption that the relation of some effectiveness factors with achievement may be curvilinear. This assumption does not only have implications for the design and analysis of effectiveness studies but also reveals that EER should provide a more accurate picture of what makes schools and teachers effective. However, the criticism against EER that it provides simplistic strategies for improving educational effectiveness can also be addressed if the dynamic model takes into account the other two weaknesses of current models of EER mentioned below.

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. Specifically, Walberg (1984) formulated an encompassing model of educational productivity which is based on the main factors of the Carroll’s model and included an additional category of environmental variables. 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. The Walberg’s model was tested as a structural equation model on science achievement, indicating more complex, indirect relationships (Raynolds & Walberg, 1990). This study seems to provide support to our argument that there is a need to develop a dynamic model of effectiveness revealing the relationships between the
factors of effectiveness which operate at the same level. Such approach to modelling educational effectiveness might reveal optimal combinations of factors that make teachers and schools effective which could also contribute in establishing strategies of improving effectiveness.

It is finally important to indicate that the current models of EER do not explicitly refer to the measurement of each factor of effectiveness. 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 et al, 2000; Kyriakides, 2004). 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 researchers who attempt to establish a dynamic model of EER should not only refer to the various factors of effectiveness but also explain the various dimensions upon which each factor can be measured. Considering effectiveness factors as multidimensional constructs does not only provide a better picture of what makes teachers and schools effective but also help us develop more specific strategies for improving educational practice.

A PROPOSED DYNAMIC MODEL OF EER

The critical review of EER helps us identify 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 restricting 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. Decisions can also be made about the design of relevant interventions as well as comprehensive evaluations of the implementation and their effects. 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 the factors included in the major constructs of the model are expected to be interrelated within and between levels.

Based on the rationale of the dynamic model presented above, the essential characteristics of the model are as follows. First, the model is expected to belong to the integrated approach to educational effectiveness modelling since it refers to multiple factors of effectiveness which operate at different levels and it is therefore multi-level in nature. 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 be curvilinear rather than linear. This refers to the possibility of searching for optimal values of the various dimensions of the factors and optimal combinations between factors.

One of the integrated models mentioned above seems to be in line with at least two of the starting points upon which the dynamic model is based. Creemers’ model refers to factors at different levels (i.e., student, classroom, school, system) and at the same time it is based on the assumption that there are direct and indirect relations between the levels and the outcomes. The 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.

A) Dimensions of measuring effectiveness factors

First, the frequency 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 of the effectiveness 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 can lead to negative effect in teacher behaviour and ultimately in student outcomes. Finally, it has to be acknowledged that frequency is only one dimension of measuring personal monitoring at school level, as it is argued below.

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 can be 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., the policy 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 more than one purpose (e.g., parents visit the school to exchange information about children 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 achieve 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 at all 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 chance 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.

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 (i.e., valid, reliable, representative to the content taught, making use of different techniques). 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 needs and this gives more emphasis to the formative function of assessment (Black & Wiliam, 1998; Harlen & James, 1997; Kyriakides et al., 2000). This refers to the second aspect of measuring quality which has to do with the impact a factor has on the subjects which are addressed by this factor. In the case of school policy on assessment, the subjects are the teachers who are expected to implement the policy whereas when we measure the effect of the factor within the EER framework we measured the impact that the factor has on student learning outcomes.

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 (e.g., all the students, teachers, schools). It is expected that adaptation to specific needs of each subject or group of subjects will increase the successful implementation of a factor and ultimately maximize 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 adopt 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 adopt their general policy into 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, adopting the policy on the special needs of each group of schools / teachers / pupils 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 factors. The examples which are given refer 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 following section refers to the specific measurement of eight factors concerning teacher behavior in classroom which are expected to be related to student achievement gains. This description could be the basis for the development of a research project attempting to test the validity of the model at the classroom level. 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., Hextall & Mahony, 1998; Kyriakides et al., 2000; Yair, 1997) and defining factors at the classroom level can be seen as a prerequisite for defining the school and the system level (Creemers, 1994).

**B) 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 et al., 2002; Muijs & Reynolds, 2001; Rosenshine, 1983), our dynamic model refers to the following eight effectiveness factors which describe teacher’s instructional role: orientation, structuring, questioning, teaching modeling, applications, management of time, teacher role in making classroom a learning environment, and teacher evaluation.

**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 can look at the number of orientations 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 or 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 of orientation is expected to reveal the extent to which teachers help their students understand the importance of finding the meanings of each task they are expected to be involved. 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 expected that the teacher will be 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 is expected to take into account. The measurement of the dimension quality refers to the properties of the orientation task and especially whether it is clear for the students. It also refers to the impact that the task has on student engagement in learning process. For example, teachers may present the reasons of doing a task simply because they have to do it and is part of their teaching routine without having much effect on student participation whereas others may encourage students to identify the purposes that can be achieved by doing a task and therefore to increase their motivation towards a specific task/lesson/series of lessons. Finally, differentiation is measured in a similar way for each of the eight factors. In the case of orientation, it is expected that teachers provide different types of orientation tasks to students according to their learning needs.

Structuring
Rosenshine & 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 in a similar way as in the case of orientation since it is possible that a structuring task may either refer to a part of a lesson or 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 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 the impact that a task has on student learning. It is expected that structuring tasks are not only clear for the students but also help them understand the structure of the lesson. For this reason, we don’t 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/series of lessons. On the contrary, the aspect of quality which refers to the properties of a structuring task has to do with the extent to which teachers organise their lessons/series of lessons in a way to move from easier tasks to 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 & Reynolds (2000) indicate that the focus of TER on 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 are also expected to ask more process questions (Everston et al, 1980; Askew & William, 1995). 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 & 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 lesson as part of the attempt of 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 don’t answer because it is too difficult for them. The aspect of impact mainly refers to the student responses and the way teacher deals with those responses. 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 student or take answers from specific students. It is also expected that the feedback that 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. 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/develop more than one strategies to solve specific problems/types of problems. Third, the stage dimension is concerned with the sequence under which a teaching modelling is used in the classroom. It is possible that initially students are faced with a problem and then are expected to use/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 later 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 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 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, therefore, examine the extent to which the application tasks refer to some parts of the lesson or 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 enable us to examine the number of purposes that application tasks cover. Stage is measured by looking at the phase of the lesson that each application task takes place. As far as the measurement of the quality of application tasks is concerned, the appropriateness of each application task can be measured by looking at the extent to which students are simply asked to repeat what they have already covered with their teacher or the application task is more complex than the content covered in the lesson or even it is used as a starting point for the next step of teaching and learning. The extent to which application tasks are used as starting points of learning can also be seen as an indication of the impact that application tasks have on students. 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 & 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 & Reynolds (2000) point out that classroom climate is a factor that teacher effectiveness research 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 (Hoy, 1990; Heck & Marcoulides, 1996). It is supported that a healthy organization 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 et al, 1991). In this context, educational effectiveness research 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/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, & Wubels, 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 teacher 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/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) that 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 that 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 teacher is able to establish on task behaviour through the interactions she/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 is 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/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 and therefore help students see the positive aspects of competition and avoid the negative ones. Stage can be measured by looking at the phase at the lesson at which the problem arises. Quality is seen in relation to the impact that the teacher behaviour has on solving the problem 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 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, this is measured through the data collected in relation to the factor concerning the role of teacher in creating a learning environment in his/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 and especially at the classroom level (Jong et al., 2004; Kyriakides, 2005a; 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 to evaluation might reduce the actual time spent on teaching and learning whereas teachers who don’t collect any information are not able to
adopt their teaching to student needs. Focus is measured by looking at the ability of 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 uses out of the information that she/he collects (e.g., identify needs of students, conducting self-evaluation, adopting his/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 the teaching of a lesson/unit of lessons) and the time lack 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. As far as the impact of the evaluation activities is concerned, we examine the type of feedback that teacher gives to the students and the way students use the teacher feedback. It is expected that effective teachers provide constructive feedback which has positive implications to 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.

SUGGESTIONS FOR POSSIBLE USES OF THE DYNAMIC MODEL FOR IMPROVING EDUCATIONAL PRACTICE

In this paper we have outlined a dynamic model that takes into account the new goals of education and the importance of illustrating the multi-level and complex nature of effectiveness. Examples of measuring effectiveness factors operating at different levels using five dimensions concerning with 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 in describing in more 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, this section refers to different methodological approaches that can be used to test this part of the model and suggestions for the next steps in the development of other parts of the model are made. Finally, some suggestions for using the dynamic model in order to improve educational practice are provided.

A) Testing the dynamic model at the classroom level: Suggestions for Methodological Approaches

The studies which have been used in order to test the validity of Creemers’ model (Jong et al., 2004; Kyriakides, 2005a; Kyriakides et al., 2000; Kyriakides & Tsangaridou, 2004) reveal the importance of using multiple measures of effectiveness factors and of conducting longitudinal studies rather than case studies in order to be able to identify the relations which exist between the various measures of each factor and student achievement gains. In this context, a longitudinal study is currently undertaken in Cyprus in order to develop and test the dynamic model. As far as the measure of student outcomes is concerned, the study does not only attempt to investigate educational effectiveness in mathematics and language but measures concerning with the main aims of religious education are also taken into account. In this respect, next to student knowledge also student attitudes towards people who have different beliefs from themselves are measured (Knuver & Brandsma, 1993; Williams & Batten, 1981). Thus, the extent to which the dynamic model can be considered as a generic or a differentiated model can be tested (Campbell et al., 2004).

While there has been substantive development of teacher effectiveness research with regards to content, the issue of measurement has been neglected to a large degree (Kyriakides et al., 2002). In the literature there is a debate whether quality of teaching is best evaluated by independent observers or by students (Aleamoni, 1981; Fraser, 1995). Both methods have their advantages and disadvantages (Ellet, 1997; Rosenshine & Furst, 1973). Thus, the
explanatory variables of the study mentioned above, which refer to the eight effectiveness factors dealing with teacher behaviour in the classroom, are measured by both independent observers and students. Specifically, taking into account the way the five dimensions of each effectiveness factor are defined, one high-inference and two low-inference observation instruments were developed. One of the low-inference observation instruments is based on Flanders system of interaction analysis (Flanders, 1970). However, we developed a classification system of teacher behaviour which is based on the way each factor of the proposed dynamic model is measured. Moreover, the observer is expected to identify the specific students involved in classroom interaction. As a consequence, the use of this instrument enables us to generate data about teacher-student and student-student interaction.

The second low-inference observation instrument refers to the following five factors of the model: orientation, structuring, teaching modelling, questioning techniques, and application. This instrument is designed in a way that enable us to collect more information in relation to the quality dimension of these five factors. Thus, the two instruments can help us generate data for all eight factors and their dimensions. The high-inference observation instrument covers all eight factors of the model and a special scale is used in order to allow us to search for curvilinear relations between the factors and student outcomes. Specifically, the scale points used are as follows: “not at all”, “scarcely”, “satisfactorily”, and “more than enough”.

The use of different types of observation instruments allows us to cover all the factors and dimensions mentioned in the proposed model. Moreover, the internal validity of the study is examined by investigating the extent to which data emerged from different observation instruments are supporting each other.

The eight factors and their dimensions are also measured by administering a questionnaire to students. Specifically, students were asked to indicate the extent to which their teacher behaves in a certain way in their classroom (e.g., at the beginning of the lesson the teacher explains how the new lesson is related to previous ones). A Likert scale was used to collect data and it is expected that students within a class view the behaviour of their teacher similarly but differently from students in other classes. Thus, we can make use of the
generalisability theory (Cronbach et al., 1972) in order to identify whether students of different classrooms agree among themselves about the way their teacher behaves in their classroom. Generalisability theory asks how accurately observed scores permit us to generalize about persons’ behaviour in a defined universe (Shavelson et al., 1989) and it is therefore essential to examine the generalisability of the data collected through the student questionnaire. Moreover, the reliability for each of the scales of the questionnaire (factor) is computed by calculating multilevel \( \lambda \) (Snijders & Bosker, 1999) and Cronbach alpha for data aggregated at the class level. Thus, our decision to treat student responses to the questionnaire as indicators of the effectiveness of their teachers is based on the results of the generalisability study and on the measures of the reliability of the scales of the questionnaire.

Two issues concerning with the analysis of the data emerged from this study need attention. Since the dynamic model belongs to the category of integrated models, it is important to use multi-level modelling techniques, which are able to identify variables at student, teacher, school and system level that are associated with student achievement. However, an issue that has to be taken into account is that the dynamic model assumes that some variables are not related in a linear way with student achievement gains. For this reason, both the effect of the various explanatory variables \((X_i)\) and the effect of the second power of these variables (i.e., \(X_i^2\) values) upon student achievement have to be identified. This approach may allow us to find out whether some variables have inverted-U curvilinear relationships with student outcomes and thereby their optimal values {i.e., the values of \(X_i\) for which \(f(X_i)\) has a maximum value} can be defined. The second issue which needs further consideration is concerned with the fact that the model assumes that factors operating at different levels can be interrelated. To examine this assumption, different statistical methods for analysing data on teacher behaviour can be used. One possibility is to use multi-level path analytic methods (Heck & Thomas, 2000) which help us not only to examine relationships between factors operating at the same level but also relevant cross level relationships. Another approach is to use multivariate multilevel modelling techniques which allow us to have more than one dependent variable. For example, the testing of the relationship between two effectiveness factors (e.g., structuring and
orientation) can be conducted by treating one effectiveness factor (e.g., orientation) as an explanatory variable and both student achievement gains and the other effectiveness factor (e.g., structuring) as dependent variables.

Two other approaches can also be used to test the validity of the proposed model and have their own specific advantages. First, international longitudinal studies can tap the full range of variation in school and classroom quality, and therefore in potential school and classroom effects. It is also likely that the existing estimates of the size of educational influences (i.e., schools and classrooms/teachers together) upon student outcomes are potentially merely artefacts of the studies’ lack of school and classroom variation. Thus, international studies could help us identify the importance of the eight factors in explaining variation in student outcomes since in national studies the lack of a significant effect might be due to the difficulties that we have to identify variation in either the student outcomes and more likely in the explanatory variables. In addition, international studies on educational effectiveness research could examine the system-effect by measuring the effect of national educational policies (Kyriakides, 2005b).

Second, national studies on effectiveness may not help us establish enough variation in the factors included in the model. This does not necessarily imply that these factors do not contribute significantly in student learning. For example, most of the teachers in a country may not use at all orientation activities or very few of them may behave in a differentiated way. In case that these results emerge from a national study, it will not be possible to demonstrate any effect. Therefore, the use of experimental approaches might be considered but attention should be given on the ecological validity of the experiment as well as on the ethical issues associated with the experimentation (Miller, 1984; Robson, 1993).

B) Developing the dynamic model at the school and system level

In this paper we suggested 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 evaluation since it is expected to have an effect on teacher evaluation which is associated with student achievement gains. Moreover, the possibility that these factors have curvilinear relations with student achievement might be considered. It is therefore important not to treat them as unidimensional constructs but, as it has been shown above, the five dimensions of the model can be applied in order to better define each of them. Similar approaches as those discussed above for the testing of the model at the classroom level can be used in order to test further elaborations of the model (Creemers & Kyriakides, 2005).

C) Suggestions for possible uses of the dynamic model

In the current phase, the emphasis is on developing and testing the model rather than on investigating the impact that the use of the dynamic model may have on improving effectiveness. However, it is expected that the dynamic model of EER will help us establish links between EER and improvement practices. In order to support our argument two possible uses of the dynamic model at the classroom level presented above are discussed. First, since the proposed part of the model refers to the instructional role of teacher and especially to specific dimensions of eight significant aspects of teaching, it can be a useful tool for teacher self-evaluation, which is considered as the key to improvement (Macbeath, 1999). At the heart of self-evaluation is the establishment of a set of criteria measuring effectiveness (Kyriakides & Campbell, 2004). Teachers could, therefore, be encouraged to draw their own meanings of what makes a teacher effective by considering the knowledge-base of effective teaching practice provided by the model. Second, based on the various dimensions of each effectiveness factor presented in the model, different teaching profiles, which affect in different ways student achievement, can be produced. Teachers may, therefore, identify the extent to which their classroom behavior is similar to any of these profiles and whether specific changes to their practice are needed in order to adopt a more effective profile. For
example, a teacher may find out that his/her effectiveness is limited due to the fact that: a) s/he does not use enough teaching modeling activities that can help students use or develop strategies for solving problems and b) the great majority of the orientation tasks he/she offers are at the introduction of the lesson. The identification of more than one weaknesses is not helpful for identifying how you can develop professionally in a better way. However, due to the dynamic nature of the model, different priorities for professional development for each teacher can be identified. These will be based on the fact that the effects of the improvement of a factor on student outcomes depend on the stage at which each individual teacher is at the moment. Thus, one teacher who attempts to improve his/her orientation skills may result in improving student outcomes more than attempting to improve his/her skills in teaching modeling. A completely different interpretation can be drawn for another teacher by looking at the situation at which he/she is at the moment.

Using the proposed model, policy-makers could conduct large-scale evaluation studies. Since some of the effectiveness factors are expected to have a curvilinear relation with student achievement, the impact of an intervention program attempting to improve a specific aspect of teaching practice (e.g., questioning techniques, teacher evaluation) will depend on what the current situation is. Therefore, data collected through these studies may help policy-makers identify those dimensions that constitute the major weaknesses of the system and therefore design relevant intervention programs for improving its effectiveness. Research is, however, needed to investigate the impact that the use of the model may have on improving teaching practice at teacher-level through building self-evaluation mechanisms and at national level through establishing an “evidence-based” approach on introducing educational policy (Fitz-Gibbon, 1996).

References


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.


