Educational Effectiveness
The Development of the Field

Bert P. M. Creemers
Faculty of Behavioral and Social Sciences
University of Groningen

Keynote address presented at the first International Conference on School Effectiveness and School Improvement in China, Shenyang, September 2005

Abstract
After a short introduction about the history of educational effectiveness research, the paper focuses on some essential elements in theory and research of the educational effectiveness movement, namely the criteria for effectiveness related to the basic concept of effectiveness, and the levels and factors which can contribute to effectiveness. Finally, related to the current situation in research, we will advocate the development and testing of a more dynamic model for educational effectiveness.

Introduction: a short history of educational effectiveness research
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 programs 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 programs in other countries (MacDonald, 1991; Schon, 1971).

In addition to methodological critiques of the Coleman report, studies were published that tried to prove that some schools did much better than could be expected on student achievement tests than others did (using a research design of positive versus negative outliers).

At almost the same point in time research was published in both the United States and the United Kingdom that got much attention in both the scholarly and popular press. Ron Edmonds (1979), a schoolboard superintendent, particularly addressed educational practitioners and Brookover et al. (1979) the educational community. These studies led to a movement in school effectiveness research and in school improvement projects based on the findings of school effectiveness research in the United States. Quite a lot of research took place into the correlates of school effectiveness, involving correlational studies focusing on the relationships between the effects of education, i.e. the outcomes of schooling, and the characteristics of schools and classrooms.

School change projects were based on these correlates discovered in the effective school research. The most famous set of correlates formed the so-called five factor model propagated by Edmonds. Most of the correlational studies and outlier studies were heavily criticized (Ralph & Fennessey, 1983) and this led to a reorientation in research and theory development after 1985.

In the United Kingdom school effectiveness research started with the Rutter study. This study found that certain factors were not associated with overall effectiveness, among them class size, school size, the age and the size of school buildings. The important within-school factors determining high levels of effectiveness were the balance of the intellectually able and less able children in school, the reward system, the school environment, the opportunities for children to take responsibility, the use of homework, the possession of academic goals, the teacher as a positive role model, good management of the classroom and strong leadership combined with
democratic decision-making. In British school effectiveness research, in addition to academic outcomes other measures like levels of rates of attendance, rates of delinquency, and levels of behavior problems were incorporated. The suggestion was that effective schools were consistently effective across a wide range of types of student outcomes (Reynolds, 1976; Rutter et al., 1979). The Reynolds studies that were ongoing in the 1970’s and 1980’s utilized detailed observations of schools in the collection of a large range of material upon pupil attitudes to school, teachers’ perceptions of pupils, within-school organizational factors and school resource levels, and revealed a number of factors within the school that were associated with more effective regimes. These included a high proportion of pupils in authority positions, low levels of institutional control, positive academic expectations, low levels of coercive management, high levels of pupil involvement, small overall size, more favorable teacher/pupil ratios and more tolerant attitudes to the enforcing of certain rules regarding ‘dress, manners and morals’.

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 (See for an overview: Teddlie & Reynolds, 2000; Townsend, Clarke, & Ainscow, 1999).

The start of the school effectiveness research and school improvement took place initially in the United States and the United Kingdom. In other countries by the early eighties preliminary studies and summaries of research were being carried out. In the Netherlands for example the research was summarized in relation to an outline for the structure of the secondary education (Creemers, 1983; Creemers & Schaveling, 1985). In fact in these countries school effectiveness research was rooted in research on teacher effectiveness, teacher behavior, and other classroom studies (Veenman et al., 1992). Again, these studies were strongly influenced by American studies (for example Brophy & Good, 1986; Doyle, 1986; Emmer, 1987; Evertson & Green, 1986; Flanders, 1970; Gage, 1972; Rosenshine, 1979) and replicated and expanded in other countries like the United Kingdom (Bennett, 1988), Sweden (Lundgren, 1972), Germany (Bromme, 1981) and Australia (Biddle, 1967; Fraser, 1986). The advantage of the later start for school effectiveness research in other countries was that they could take advantage of both the results of the earlier studies and the criticisms of these studies, and therefore could make a better start by avoiding methodological flaws in past research.

The result of the early effectiveness studies was that a number of factors were related to effectiveness, such as the five factors of Edmonds (1979):
- strong educational leadership,
- high expectations of student achievement,
- an emphasis on basic skills,
- safe and orderly climate,
- frequent evaluation of pupil progress

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; Levine & 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 to 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 criticized 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, behavior, 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 conceptualized and measured in terms of new kinds of citizens’ sensibilities, moral and cultural practices 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 labor (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.

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.

Researchers in the field of educational effectiveness face the problem of deciding on the criteria for effectiveness. In a discussion about effectiveness quite often not only the terms 'effectiveness' or 'improvement' are used, but participants also talk about the 'quality' of education, of a class, of a school. The term 'quality' is rather vague because it can include almost anything, like effectiveness, efficiency, and statements about the content, processes, and inputs of education. That is the reason why the term 'effectiveness' is preferred, even when it has a rather negative connotation in some (practice- and policy-oriented) audiences. Related to the criteria of
effectiveness, the problem arises of what induces effectiveness. Because researchers are involved in what constitutes and what leads to school effectiveness, there is a good chance that goals and means will be exchanged. For example, educational leadership seems to be a goal instead of a means to achieve more effective education; changes in educational leadership can contribute to the improvement of schools towards more effectiveness.

Researchers are interested in the development and the testing of a theory about what induces effectiveness: what factors, what characteristics of education lead to effectiveness, or in other words, what makes the differences in effectiveness between schools. Most parties are looking for the components of education, for the behaviors of principals and teachers, that can differ between schools, classrooms et cetera. This may lead to the conclusion that when we see more of a certain kind of behavior or more of a certain factor in schools, in classrooms and in the educational system, then that particular system, that school, or that classroom has a higher quality. In several research traditions leadership, educational climate and the structure of teaching are measured and from these inferences are made about the effectiveness of the school. Factors such as structuring and climate and so on are not all about high effectiveness in themselves, but the importance of these factors and variables is based on the fact that they lead to (high) effectiveness.

The criteria for effectiveness have to be defined in another way. They must be something else than the characteristics and features of education in the educational system, in schools, and in classrooms. School effectiveness research has to define criteria for effectiveness which are something other than the features and characteristics of the education itself. These criteria are related to the effects of education, to the ends, the objectives, to the goals of schools and of education. However, this does not end the discussion about criteria. On the contrary, then we need a discussion about the ends, the effects of education and we have to determine how they can be measured. For example, in early research on school effectiveness, the criteria for effectiveness were the academic outcomes of students, but they could be defined in different ways as the outcomes of tests, as differences between the start and the end of education, as referrals to special education and as transition from one school to another.

The concept of educational effectiveness

Central to all the literature on effective schools is the concept of ‘effectiveness’. We define effectiveness in two dimensions (see figure 1). In figure 1, regressions for four schools are estimated separately. In this way maximum use is made of between en within school differences. The ‘quality’ dimension is modeled as the average score of each school on output (corrected for input) and is represented by the intercept (each school has a different intercept). The ‘equity’
dimension encompasses the compensatory power or selective quality of schools. Some schools can better compensate for input characteristics than others. This dimension is represented by the slopes of the within school regression of output on input. Pupils in schools A and B on average score better than pupils in schools C and D. This estimation of average scores gives insight into the quality differences between schools.

The slopes of these schools tell us something about the equity dimension. Schools A and C differentiate strongly between children with low and high input characteristics (e.g. socioeconomic background) and can be called ‘elitist’. Schools B and D differentiate less between low and high inputs. These schools are better capable of compensating for input and thus can be labeled as ‘egalitarian’ schools.

Educational effectiveness emphasizes the quality dimension, it is always about achievements of students. But the equity dimension raises attention for the fact that there are differential effects and some teachers and schools have better results with specific groups of students (Cambell et al., 2004).

![Graph showing regression lines of four different schools](image)

**Figure 1**
Regression lines of four different schools

**Different kinds of criteria for educational effectiveness**

In the introductory section we made a distinction between ‘means’ and ‘ends’ in education to avoid the confusion of having the characteristics of the educational process as the final criteria for effectiveness. But also with respect to the ends of education, different kinds of criteria are possible. In the current situation in the Netherlands, with declining amounts of students for schools, a very acceptable objective for schools is to attract more students. In fact, quite a lot of...
activities in schools like public relations activities and parent participation are developed in order to get more students. This is understandable because the allocation of resources depends on the number of students. The activities to attract students and to hold them can be seen as efforts to increase the financial and personal input for schools. This is also the case with the hiring of teachers, to provide them with in-service training, a good salary, and support so that they will stay. In fact the well-being of teachers is an acceptable objective for a school as an organization (it contributes to stability in the organization, which leads to an increase in output). When we look at education as a system we can make a distinction between the input, the process, the product, and the context of education. Within these components of the system all kinds of elements/variables can be discerned and criteria can be formulated with respect to different elements of the components of the system and probably the relations between the elements.

On the input side are the number of students; the resources like funding; the material like books, computers, audio equipment; the teacher quality (years of experience, academic level); the quality of the building and so on. With respect to the processes in the school different kinds of criteria can be formulated like the number of clubs, extracurricular activities, hours spent in school, pedagogical climate, quality of teaching, and the functioning of the school as an organization (e.g. decision making processes). With respect to the output, all kinds of criteria are possible too, from academic outcomes (and maybe a small set of academic outcomes) to the well-being of students and all kinds of pedagogical goals, and also to criteria that refer more to the teachers, their well-being and willingness to stay in school. Also in the context of schools and classrooms different elements can be discerned and criteria can be formulated. For a specific school, for example, the amount and quality of support and at the national level (in comparison with the educational context in other countries) the autonomy of schools, national guidelines and so on.

Based on the relationship between components of the system and the elements within the components, questions arise about the interdependency between components, elements and especially with the criteria.

The overview of possible criteria given above makes clear that virtually everything can be used as criteria for quality. Sometimes that is what happens with the use of educational indicators, when the indicator that is just meant to provide a description of the system is used as criterion for quality. It is possible to formulate criteria for quality for each of the elements of the system, but there is something distinctive about the quality of that element of the system, about inputs, and about process. But as a system schools and education as a whole should be judged and evaluated based on the outputs of the system, and the criteria for specific components and elements should be related to the output. This means that ultimately the criteria for the effectiveness of schools are
based upon their outcomes. Like other organizations, schools have to deliver something. Schools can deliver things like cooperation with parents, with other schools, and generating good and confident teachers. But in the end the criteria for effectiveness are formulated in terms of student outcomes. These outcomes determine the effectiveness of schools and of education in general. Even the outcomes such as the characteristics of teachers are in the end means to enhance the outcomes of the students. In itself it is possible to speak about the effectiveness criteria for each of the components but it is more important, in addition to the effectiveness criteria for each of the components and the elements within the components, to investigate the relationship between the components (in good synchronization) and the relationship to the outcomes of the students.

**Student outcomes**

When we agree that student outcomes are the criteria for effectiveness in education, the question remains what kind of outcomes, objectives and goals, can be achieved by schools. Especially in a time when other parts of the society can not fulfil their own tasks, there is a danger that the school is overstretched by formulating and stressing more and more objectives. Probably nobody disagrees with the statement that schools are for the learning of students, especially cognitive learning. That means that it can be expected from education that students learn reading, mathematics and language. The first question connected to this is how much knowledge and what kind of knowledge is important? Especially in times of economic recession there is always a tendency to go back to the basics, to ask for basic knowledge and skills and school effectiveness research is connected with this emphasis on basic skills. But when schools are busy pursuing those kinds of objectives, most of the time the criticisms concern the teaching of basic knowledge, rote learning, facts and figures, and there are ideas that children should learn more than these things. That leads to objectives in the areas of transfer, evaluation, metacognition, learning how to learn, et cetera. As long as we do not deny the importance of basic knowledge and skills, more than basic knowledge can be added to the objectives and the goals of education in schools. But based on research on metacognition (Prawat, 1989) and also according to Bloom’s taxonomy of educational objectives (Bloom et al., 1956), it is evident that for higher order learning, knowledge and skills basic learning and basic knowledge are required. Schools should ensure that a knowledge base is available and that on top of that students can acquire and develop other knowledge and skills.

---

1 This is also caused by the fact that the emphasis on basic skills was viewed as an important factor for effectiveness related to time on task.
So far we have restricted ourselves, in what we described before, to the cognitive domain. Next to the cognitive domain, schools can contribute to affective and aesthetic outcomes as well. Especially in a time when other organizations in society do not fulfil their tasks properly or sufficiently, schools are required to pursue more than cognitive outcomes. The school is a place where students and teachers meet, live with each other and exchange experiences, so it is more than just a place for cognitive development. However it is a question how much schools can teach and students can learn objectives in the area of the affective, motor and social domain. It depends on the time available and the possibility of teaching and learning these domains in schools. Schools can contribute a little bit to outcomes in these domains, although the impact of schools seems to be quite small, for example with respect to art education (Haanstra, 1994). When we take into account the fact that schools need quite a lot of time to fulfil the tasks in the cognitive domain properly, then probably all other goals are welcome side effects, non-intended outcomes of education. Apparently the influence of other social organizations is far more important than schools can offer in this respect. The chance that schools can achieve anything in these areas in greater when they are connected with cognitive development, like the cognitive component in social behavior or in affective or aesthetic education. As research in the past has shown, there is no discrepancy between achieving objectives in the cognitive domain and in other areas and also in the well-being of students. Students can feel quite happy in schools where cognitive objectives are pursued and also they can acquire cognitive knowledge that is important in other domains as well.

In a more instrumental way, there is another issue important in relation to outcomes. In the past, the outcomes of schools and education were measured quite inappropriately by percentages of students passing exams, dropping out or being referred to special education. Later measures for effectiveness and quality used were tests in a restricted range of skills, because at the time that there was discussion of criteria for effectiveness generally a very restricted set of tests was used to measure effectiveness: those of basic skills in language, reading and mathematics. It is apparent from what we described above that more, and different, kinds of measurements should be used to measure the quality and effectiveness of education.

Based on the results of factors that can explain variance in student outcomes, it became clear that we have to take into account the background of students. That means that we cannot just look at the outcomes at the end of education, but we have to take into account the entrance level of students as well. So the measurement of the effectiveness of education in different areas has to include measurements of input behavior to make sure we discover the added value of education.
From correlates, characteristics and factors to framework and theories

The five factors model (Edmonds, 1979) has had a major impact on educational practice in the USA and resulted in interventions such as leadership training and assessment of students’ progress. It was also the start for research to validate the five factors and / or to find other factors (Rutter et al., 1979; Reynolds, 1982).

The studies after 1985 were rooted in the tradition ‘founded’ by Rutter, Brookover and Edmonds, but as a response to the criticism on the research they were designed and carried out in a methodologically appropriate way and made use of new techniques for statistical analysis of the data such as hierarchical linear modeling (HLM).

Major studies were carried out by Mortimore et al. (1988) in the United Kingdom and by Teddlie and Stringfield (1993) in the United States. In the same period, in other countries like the Netherlands, Hong Kong and Norway, school effectiveness studies also began (for an overview see Reynolds et al., 1994).

Mortimore’s research was based upon 50 randomly selected London primary schools. Over a period of four years, the academic and social progress of 2,000 children were traced. Mortimore et al. identified a number of schools which were effective in both academic and social areas. These schools possessed the following characteristics:

1. *Purposeful leadership of the staff by the head*
   This occurred where the head understood the school’s needs, was actively involved in it, but was good at sharing power with the staff. He or she did not exert total control over teachers but consulted them, especially in decision-making such as spending plans and curriculum guidelines.

2. *Involvement of the deputy head*
   Where the deputy was usually involved in policy decisions, pupil progress increased.

3. *Involvement of teachers*
   In successful schools, the teachers were involved in curriculum planning and played a major role in developing their own curriculum guidelines. As with the deputy head, teacher involvement in decisions concerning which classes they were to teach was important. Similarly, consultation with teachers about decisions on spending was important.

4. *Consistency among teachers*
   Continuity of staffing had positive effects but pupils also performed better when the approach to teaching was consistent.
5. *A structured day*

Children perform better when their school day was structured in some way. In effective schools, pupils’ work was organized by the teacher, who ensured there was plenty for them to do yet allowed them some freedom within the structure. Negative effects were noted when children were given unlimited responsibility for a long list of tasks.

6. *Intellectually challenging teaching*

Not surprisingly, pupil progress was greater where teachers were stimulating and enthusiastic. The incidence of ‘higher order’ questions and statements was seen to be vital – that is where teachers frequently made children use powers of problem solving.

7. *A work-centered environment*

This was characterized by a high level of pupil industry, with children enjoying their work and being eager to start new tasks. The noise level was low, and movement around the class was usually work-related and not excessive.

8. *A limited focus within sessions*

Children progressed when teachers devoted their energies to one particular subject area and sometimes two. Pupil progress was marred when three or more subjects were running in the classroom.

9. *Maximum communication between teachers and pupils*

Children performed better the more communication they had with their teacher about the content of their work. Most teachers devoted most of their time to individuals, so each child could expect only a small number of contacts a day. Teachers who used opportunities to talk to the whole class by, for example, reading a story or asking a question, were more effective.

10. *Thorough record-keeping*

The value of monitoring pupil progress was important in the head’s role, but it was also an important aspect of teachers’ planning and assessment.

11. *Parental involvement*

Schools with an informal open-door policy which encouraged parents to get involved in reading at home, helping in the classroom and on educational visits, tended to be more effective.

12. *A positive climate*

An effective school has a positive ethos. Overall, the atmosphere was more pleasant in the effective schools.

(Mortimore et al., 1988)
The Louisiana School Effectiveness Study (Teddlie & Stringfield, 1993) was a program of in fact four studies, starting in 1980 with the first pilot study and finishing in 1992 with the fourth study. The study addressed both the meso level and the micro level and used a combination of quantitative and qualitative techniques. In the third study the data were analyzed at both school and classroom level. Differences between effective and ineffective schools were found, the effective schools being different with respect to high time on task, the presentation of new material, encouragement of independent practice, high expectations, positive reinforcement, a small number of interruptions, discipline, a (friendly) ambience, student work displayed, and the appearance of the classroom. Especially the qualitative case studies provided insight into the characteristics of effective and ineffective schools. This can be seen in the contrast between the so-called Kennedy and Coolidge Elementary schools (the former being effective and the latter being ineffective) (See Table 1).

In addition to these major studies quite a lot of research took place into the specific characteristics of effective schools. A review of North American studies is provided by Levine and Lezotte (1990). Their summary confirms the so-called five-factor model, although additional effectiveness characteristics are included in the review in the instructional area. According to the review, the characteristics of unusually effective schools are mentioned in Table 2.

At first sight there is a communality in the findings of the British and the American studies, and with the studies in the past and the major studies of the late eighties. There are also important differences, for example with respect to frequent monitoring which seems to be effective according to American studies but in the Mortimore et al. study (1988) is a characteristic of ineffective schools. Also, the nature of effective leadership seems to be somewhat different in the United States and in the United Kingdom. The United States research emphasizes strong educational leadership by the principal as effective but in the United Kingdom, according to the Mortimore and Reynolds studies, effectiveness is related to leadership by the head and the deputy head, and also the involvement of teachers in decision-making.

In the Netherlands after 1985 quite a lot of research was developed using almost the same designs, as well as the translation of instruments, used in British and American studies. Surprisingly enough, in the Netherlands quite often results found in American and British studies could not be replicated. Scheerens (1992) summarizes 16 Dutch school effectiveness studies. According to his analysis the connections between outcomes and characteristics found in other studies are recognized only quite weakly (see Table 3). Note especially that characteristics such
as ‘frequent evaluation’ and ‘educational leadership’ that turned out to be effective in other studies do not appear to be that effective in the Netherlands, although this conclusion in the Netherlands with respect to educational leadership could be different if large-scale studies of secondary schools were included. The organization of primary schools with little educational responsibility for principals (and more administrative tasks) is different from the organization in schools in North America and the United Kingdom. Secondary schools, however, are closer to the models in the United Kingdom and the United States.
Table 1 Contrasts between Kennedy and Coolidge Elementary Schools

<table>
<thead>
<tr>
<th>John F. Kennedy Elementary</th>
<th>Calvin Coolidge Elementary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principal</strong></td>
<td></td>
</tr>
<tr>
<td>1. Stable, appropriate leadership</td>
<td>1. Unstable, generally inappropriate leadership</td>
</tr>
<tr>
<td>2. Appropriate, informal organizational culture</td>
<td>2. Inappropriate, informal organizational structure</td>
</tr>
<tr>
<td>3. Shared academic leadership with faculty</td>
<td>3. Nonshared academic leadership</td>
</tr>
<tr>
<td>4. Resistant to external change</td>
<td>4. Accepting of external change</td>
</tr>
<tr>
<td>5. Close relationship among administrators</td>
<td>5. Strained relationship among administrators</td>
</tr>
<tr>
<td>6. Good use of academic support staff</td>
<td>6. Unimaginative use of academic support staff</td>
</tr>
<tr>
<td><strong>Faculty</strong></td>
<td></td>
</tr>
<tr>
<td>7. Faculty is warm, friendly</td>
<td>7. Faculty is cold, guarded</td>
</tr>
<tr>
<td>8. Strong faculty cohesiveness</td>
<td>8. Lack of faculty cohesiveness</td>
</tr>
<tr>
<td>9. No obvious personality conflicts among faculty</td>
<td>9. Open bickering among faculty</td>
</tr>
<tr>
<td>10. Integration of support staff into faculty</td>
<td>10. Inappropriate and uneven integration of support staff into faculty</td>
</tr>
<tr>
<td>11. Cooperative efforts to enhance teaching</td>
<td>11. Top-down effects to enhance teaching</td>
</tr>
<tr>
<td>12. High faculty stability</td>
<td>12. Moderate to low faculty stability</td>
</tr>
<tr>
<td>13. High time-on-task and positive classroom climate</td>
<td>13. Low time-on-task and evidence of negative classroom climate</td>
</tr>
<tr>
<td>14. Fairly uniform teaching behaviors across classes</td>
<td>14. Large variances in teaching behaviors across classes</td>
</tr>
<tr>
<td>15. Assistance freely given to new faculty members</td>
<td>15. Little assistance given to new faculty members</td>
</tr>
<tr>
<td><strong>Students</strong></td>
<td></td>
</tr>
<tr>
<td>16. Excellent discipline and understanding of rules</td>
<td>16. Poor discipline and understanding of rules</td>
</tr>
<tr>
<td>17. Students involved in running of school</td>
<td>17. Little or no student involvement in running of school</td>
</tr>
<tr>
<td>18. Little use of corporal punishments</td>
<td>18. Excessive use of corporal punishment</td>
</tr>
<tr>
<td>20. Consistently high student achievement</td>
<td>20. Consistently low student achievement</td>
</tr>
</tbody>
</table>

(Teddlie and Stringfield, 1993; 132)
Table 2 Characteristics of Unusually Effective Schools

<table>
<thead>
<tr>
<th>Productive School Climate and Culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Orderly environment</td>
</tr>
<tr>
<td>- Faculty commitment to a shared and articulated mission focused on achievement</td>
</tr>
<tr>
<td>- Problem solving orientation</td>
</tr>
<tr>
<td>- Faculty cohesion, collaboration, consensus, communications, and collegiality</td>
</tr>
<tr>
<td>- Faculty input into decision-making</td>
</tr>
<tr>
<td>- Schoolwide emphasis on recognizing positive performance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Focus on Student Acquisition of Central Learning Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Maximum availability and use of time for learning</td>
</tr>
<tr>
<td>- Emphasis on mastery of central learning skills</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Appropriate Monitoring of Student Progress</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Practice-Oriented Staff Development at the School Site</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Outstanding Leadership</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Vigorous selection and replacement of teachers</td>
</tr>
<tr>
<td>- ‘Maverick’ orientation and buffering</td>
</tr>
<tr>
<td>- Frequent, personal monitoring of school activities, and sense-making</td>
</tr>
<tr>
<td>- High expenditure of time and energy for school improvement actions</td>
</tr>
<tr>
<td>- Support for teachers</td>
</tr>
<tr>
<td>- Acquisition of resources</td>
</tr>
<tr>
<td>- Superior instructional leadership</td>
</tr>
<tr>
<td>- Availability and effective utilization of instructional support personnel</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Salient Parent Involvement</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Effective Instructional Arrangements and Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Successful grouping and related organizational arrangements</td>
</tr>
<tr>
<td>- Appropriate pacing and alignments</td>
</tr>
<tr>
<td>- Active/enriched learning</td>
</tr>
<tr>
<td>- Effective teaching practices</td>
</tr>
<tr>
<td>- Emphasis on higher order learning in assessing instructional outcomes</td>
</tr>
<tr>
<td>- Coordination in curriculum and instruction</td>
</tr>
<tr>
<td>- Easy availability of abundant, appropriate instructional materials</td>
</tr>
<tr>
<td>- Classroom adaptation</td>
</tr>
<tr>
<td>- Stealing time for reading, language and math</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High Operationalized Expectations and Requirements for Students</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Other Possible Correlates</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Student sense of efficacy/futility</td>
</tr>
<tr>
<td>- Multicultural instruction and sensitivity</td>
</tr>
<tr>
<td>- Personal development of students</td>
</tr>
<tr>
<td>- Rigorous and equitable student promotion policies and practices</td>
</tr>
</tbody>
</table>

(Levine and Lezotte, 1990; 10)
Table 3  Synthesis of Dutch School Effectiveness Research

<table>
<thead>
<tr>
<th>School Characteristic</th>
<th>Number of studies with positive effect</th>
<th>Total number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Teacher’s experience</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>- Changes in staff</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>- Private/state education</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>- Positive expectations of pupil achievement</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>- Frequent evaluation</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>- Achievement-oriented policy</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>- Teacher cooperation / consensus</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>- Child-centered approach</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>- Opportunity to learn</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>- Structured teaching</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>- City/rural school</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>- Orderly climate</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

(Scheerens, 1992; 65)

Several Dutch studies (Vermeulen, 1987; Van de Grift, 1990; Van de Grift & Houtveen, 1991) have even found a negative relationship between student outcomes and educational leadership as operationalized and measured in the American research. In a major evaluation study, namely the evaluation of the Educational Priority Program in the Netherlands, the knowledge base of school effectiveness was used to collect information about school and classroom characteristics at the primary school level. Table 4 provides an overview of the results of the statistically significant characteristics of effectiveness/ineffectiveness. The Table provides the results for grades 6 and 8 for three types of schools: 1. a select sample of Dutch elementary schools, 2. schools with additional funding on the basis of their numbers of disadvantaged students which are cooperating in Educational Priority Areas policies and 3. schools with additional funding on the basis of the number of disadvantaged students, not cooperating in Educational Priority Areas policies. The results are provided for both language and mathematics achievement.

An outlier analysis of consistently high, medium and low effectiveness schools (Van der Werf, 1995) and an in-depth study of schools with high proportions of minority students (Van der Werf, Nitert & Reezigt, 1994) confirm the results that the explanation of effectiveness could be found particularly at the classroom level and in the instructional process. At the school level again the negative relationship of educational leadership and outcomes was evident. Characteristics at the school level that show positive relationships with achievement are coherence, emphasis on basic skills, clear objectives, curriculum independent tests, and a homework policy. One should take
into account the fact that the organization of especially primary schools in the Netherlands is different from the organization in the United States and the United Kingdom. It is more based on the professional community and leadership among teachers, backed up by informal influence on the part of the principal. That becomes evident from the fact that some of the positive correlates of effectiveness in Dutch studies are seen as parts of educational leadership in other countries: the effective elements seem to be the same, but they are somewhere else in the organization. This result points to the fact that in replication studies in other countries the contextual differences should be taken into account.

Table 4  Characteristics that are related to differences between elementary schools in language and mathematics, after correction for the proportion of disadvantaged students in the school, student intelligence and prior achievement of students

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>+/-</th>
<th>lang.</th>
<th>math.</th>
</tr>
</thead>
<tbody>
<tr>
<td>denomination (Catholic)</td>
<td>+</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>number of students</td>
<td>+</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>% minority students</td>
<td>-</td>
<td>6,8</td>
<td>6,8</td>
</tr>
<tr>
<td>remedial teaching</td>
<td>+</td>
<td>6,8</td>
<td></td>
</tr>
<tr>
<td>school evaluation system</td>
<td>+</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>use of national test</td>
<td>+</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>time for instruction</td>
<td>+</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>homework</td>
<td>+</td>
<td></td>
<td>6,8</td>
</tr>
<tr>
<td>whole class instruction</td>
<td>+</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>feedback</td>
<td>+</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>evaluation of student progress</td>
<td>+</td>
<td>6,8</td>
<td></td>
</tr>
</tbody>
</table>

(Van der Werf & Guldemond, 1993)

+ positive effect on achievement
- negative effect on achievement
6 means effect in grade 6
8 means effect in grade 8

The research project carried out in the second half of the eighties had better research designs than the earlier studies and made use of more sophisticated techniques for data analysis. In some countries like the United States and the United Kingdom, the techniques for data collection also included observation, whereas in other countries, like the Netherlands, quite a lot of surveys were used to collect data. In the case of the survey research, we have to be cautious about the results. For example, the quite disappointing research results in the evaluation of the Dutch Education Priority Program could be caused by the fact that teachers knew how to fill in surveys about teaching. That could explain the fact that no differences between effective and ineffective schools were found.
In some studies, for example in the Louisiana studies, a distinction was already made between school and classroom levels whereas in other studies, especially the reviews of the literature, there is a mix of classroom and school factors. Although research was improved, the result was still a long list of correlates for effectiveness that needed regrouping and rethinking in order to generate better understanding. As Stringfield, Teddlie, Wimpelberg, and Kirby concluded in their presentation of the follow-up study in Louisiana: ‘We need to be developing sophisticated, practical, contextually, sensitive models of effective schooling’ (Stringfield et al., 1992).

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:

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

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 \( I \) 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 given to 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 that have been developed 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 organizational 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 (De 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).
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) 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 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 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. 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. Fourth, 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.

Creemers’ model is in line with at least two of the starting points upon which the dynamic model is based. It 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. 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 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 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 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 (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 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 of the factor upon the subjects addressed by the 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. Although this dimension could be considered as part of the quality dimension, special emphasis to differentiation is given due to the fact that research on teacher effectiveness (TER) has shown that adaptation to specific needs of each subject or group of subjects increases the successful implementation of a factor and ultimately maximizes 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, policymakers 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/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 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 second part of this section refers to the measurement of eight effectiveness factors concerned with teacher behavior in 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).

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 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 modeling) are closely related with the new goals of education.

Orientation

Orientation refers to teacher behavior in providing the objectives for which a specific task or lesson or series of lessons take(s) place and/or challenging students to identify the reason 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 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 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. 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. 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 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/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 & Stevens (1986) point out that achievement is maximized 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 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 the 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 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 the 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 assumed that the feedback that effective teachers give to student answers varies according to their needs.

**Teaching Modeling**

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 modeling 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 modeling 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-modeling 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 modeling. 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., cause more confusion about the problem). Finally, differentiation can be seen in terms of adopting teaching modeling 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 emphasizes 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 enables 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 task is 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 behavior 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 behavior of the stakeholders, whereas culture is seen as measuring the values and norms of the organization (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 behavior. The other three elements refer to the attempt of teacher to create a businesslike and supportive environment for learning (Walberg, 1984) 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) 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 in 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 the teacher is able to establish on task behaviour through the interactions she/he promotes since Creemers’ model emphasizes 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 behavior 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 of the lesson at which the problem arises. Quality is seen in relation to the impact that the teacher behavior has on solving the problem that arise as measured through students’ behavior. For example, a teacher may not use any strategy at all to deal with a classroom disorder problem or use a strategy that solves the problem only temporarily 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 organize and manage the classroom environment as an efficient learning environment and thereby to maximize 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 for 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, 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 to evaluation might reduce the actual time spent on teaching and learning whereas teachers who don’t collect any information are not able to adapt 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 multiple use 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 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. 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 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 will describe in more detail the factors associated with teacher behavior 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. Also, 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 regard 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 behavior 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 behavior 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 modeling, 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 generalizability 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. Generalizability theory asks how accurately observed scores permit us to generalize about persons’ behavior in a defined universe (Shavelson et al., 1989) and it is therefore essential to examine the generalizability 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 generalizability study and on the measures of the reliability of the scales of the questionnaire.

Two issues concerning the analysis of the data from this study need attention. Since the dynamic model belongs to the category of integrated models, it is important to use multi-level modeling 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 if 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 concerns the fact that the model assumes that factors operating at different levels can be interrelated. To examine this assumption, different statistical methods for analyzing data on teacher behavior 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 modeling 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 at all use orientation activities or very few of them may behave in a differentiated way. If these results are shown up in 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 to the ecological validity of the experiment as well as to 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 organizational 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 we have 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 a and b).

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 weakness is not helpful if you want to find out how you can develop yourself professionally in a better way. However, due to the dynamic nature of the model, different priorities for professional development can be identified for each teacher. 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 if he should attempt 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


Kyriakides, L. (2005a). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement 16*


