

USING EDUCATIONAL EFFECTIVENESS RESEARCH TO DESIGN THEORY-DRIVEN EVALUATION AIMING TO IMPROVE QUALITY OF EDUCATION

Bert Creemers & Leonidas Kyriakides

INTRODUCTION: A BACKGROUND OF EER

Educational Effectiveness Research (EER) can be seen as an overarching theme that links together a conglomerate of research in different areas; including research on teacher behaviour and its impacts, curriculum, student grouping procedures, school organisation, and educational policy. The main research question underlying EER is the identification and investigation of which factors in the teaching, curriculum, and learning environments, (operating at different levels such as the classroom, the school, and above-school) can directly or indirectly explain measured differences (variations) in the outcomes of students. Further, such research frequently takes into account the influence of other important background characteristics, such as student ability, Socio-Economic Status (SES), and prior attainment. Thus, EER attempts to establish and test theories which explain why and how some schools and teachers are more effective than others in promoting better outcomes for students.

It is also important to note that the three terms—school effectiveness, teacher effectiveness and educational effectiveness – are used inconsistently in the literature and that these are themselves inter-related. In this book, school effectiveness is taken to mean the impact that school-wide factors, such as a school policy for teaching, school climate, and the ‘mission’ of a school, have on students’ cognitive and affective performance. On the other hand, teacher effectiveness is taken to mean the impact that classroom factors have on student performance,

and includes teacher behaviour, teacher expectations, classroom organisation, and use of classroom resources.

Teddlie (1994) argued that most teacher effectiveness studies have been concerned with only the processes that occur within classrooms to the exclusion of schoolwide factors, whereas most school effectiveness studies have involved phenomena that occur throughout the school with little emphasis on particular teaching behaviours within individual classrooms. Only a few EER studies have attempted to examine school and classroom effectiveness simultaneously (Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Teddlie & Stringfield, 1993). The attempts to deal with both teacher and school influences can be seen as a significant development in EER since joint studies on school and teacher effectiveness reveal that neither level can be adequately studied without considering the other (Reynolds et al., 2002). In this context, we are using the term educational effectiveness rather than teacher and/or school effectiveness to emphasise the importance of conducting joint school and teacher effectiveness research which can help us identify interactions between the school, classroom and student levels and their contributions in explaining variation in students' outcomes, both academic and non-cognitive. Finally, it is important to note that EER also refers to the functioning of the educational system as a whole and this research can, therefore be used to support the development and testing of different models of effectiveness (e.g., Creemers, 1994; Creemers & Kyriakides, 2008; Scheerens, 1992; Stringfield & Slavin, 1992). In turn, these models of effectiveness ultimately attempt to explain why educational systems and their sub-components perform differently, towards the aim of providing relevant evidence for policy makers.

The origins of EER largely stem from reactions to seminal work on equality of opportunity in education that was conducted in the USA and undertaken by Coleman, Campbell,

Hobson, McParttland, Mood, Weinfield and York (1966), and Jencks, Smith, Ackland, Bane, Cohen, Grintlis, Heynes and Michelson (1972). These two innovative studies from two different disciplinary backgrounds - sociology and psychology, respectively – drew very similar conclusions in relation to the amount of variance in student outcomes that can be explained by educational factors. Although, the studies did not suggest schooling was unimportant, the differences in student outcomes that were attributable to attending one school rather than another were modest. However, these studies were also criticised for failing to measure the educational variables that were of the most relevance (Madaus, Kellagham, Rakow, & King, 1979). Nevertheless, it is important to note that these two studies both claimed that after taking into consideration the influence of student background characteristics such as ability and family background (e.g. race and SES), only a small proportion of the variation in student achievement could be attributed to the school or educational factors. This pessimistic feeling of not knowing what, if anything, education could contribute to reducing inequality in educational outcomes and in society as a whole was also fed by the apparent failure of large-scale educational compensatory programmes, such as “Headstart” and “Follow Through” conducted in the USA, which were based on the idea that education in pre-school/schools would help compensate for initial differences between students. Similarly disappointing results have since also been reported for the effects of compensatory programs that have been conducted in other countries (Driessen & Mulder, 1999; MacDonald, 1991; Schon, 1971; Taggart & Sammons, 1999; Sammons, Power, Elliot, Campbell, Robertson, & Whitty, 2003).

The first two school effectiveness studies that were independently undertaken by Edmonds (1979) in the USA and Rutter, Maughan, Mortimore, Ouston and Smith (1979) in England during the 1970s were concerned with examining evidence and making an argument

about the potential power of schooling to make a difference in the life chances of students. This was an optimistic point of view because many studies published in that period had shown that teachers, schools, and maybe even education in general had failed to make much of a difference. The early existence of these two independent research projects in different countries that asked similar questions and drew to a certain extent on similar quantitative methodologies demonstrated the potential for establishing a scientific domain dealing with effectiveness in education (Kyriakides, 2006). Thus, the publications by Brookover, Beady, Flood, Schweitzer and Wisenbaker (1979) and Rutter et al. (1979) were followed by numerous studies in different countries on educational effectiveness and the development of international interest and collaboration through the creation of the International Congress for School Effectiveness and Improvement in 1990 (Teddle & Reynolds, 2000). Looking at the history of EER, we see four sequential phases in the field which address different types of research questions and promote the theoretical development of EER.

1. First Phase: A focus on Size of School Effects. Establishing the field by showing that ‘school matters’

During the early 1980s conducted studies attempted to show that there were differences in the impact that particular teachers and schools have on student outcomes. This research showed how important it is for students to have effective teachers and schools, and that school and teacher effects tend to be larger for disadvantaged groups (Scheerens & Bosker, 1997).

2. Second Phase: A focus on the Characteristics/correlates of effectiveness. Searching for factors associated with better student outcomes

In the late 1980s and early 1990s, researchers in the area of EER were mainly concerned with identifying factors which are associated with student outcomes. These studies resulted in a list of factors which were treated as characteristics of effective teachers and schools (Levine & Lezotte, 1990; Sammons, Hillman, & Mortimore, 1995; Scheerens & Bosker, 1997).

3. Third Phase: Modelling Educational Effectiveness. The development of theoretical models that explain why specific factors are important for explaining variation in student outcomes.

By the late 1990s and early 2000s several integrated models of educational effectiveness (Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992) had been developed.

These models sought to explain why factors which operate at different levels are associated with student outcomes and these guided not only the theoretical development of EER but also the design of empirical studies within this field (Kyriakides, Campbell, & Gagatsis, 2000; de Jong et al., 2004).

4. Fourth Phase: Focus on Complexity. A more detailed analysis of the complex nature of educational effectiveness which developed further links with the study of school improvement. This features a focus on change over time and addresses issues such as consistency, stability, differential effectiveness, and departmental effects.

A graduate movement from the third to fourth place was observed particularly after 2000. Researchers increasingly gave attention to the study of complexity in education and pointed to the fact that the theoretical models of the third phase had not emphasised the *dynamic* perspective of education nor had they paid sufficient attention to the differential character of some factors (Creemers & Kyriakides, 2006). Moreover, this graduate

movement also saw an interest develop in investigating the question of changes in the effectiveness of schools, rather than exploring the extent of stability in effectiveness (Kyriakides & Creemers, 2009). The move away from seeing effectiveness as an essentially stable characteristic of different schools or teachers to one that varies across years and may differ for different student outcomes or in relation to different groups of students places change at the heart of EER. As such, the field became increasingly linked with the growth of larger scale, systematic evaluations of the long term effect of teachers and schools and of local and national reform policies (Kyriakides, Antoniou, & Maltezos, 2009; Pustjens, Van de Gaer, Van Damme, & Onghena, 2004).

AIMS OF THE PAPER

In this paper, we discuss the impact of EER on the design and the evaluation of reform policies at local and national level and the establishment of strategies to improve practice. It is argued that EER can contribute in the development of theory-driven evaluation studies which will serve both policy-makers and practice as well as promoting further theoretical development of the field. Thus, we provide guidelines on how to design theory-driven evaluation studies by taking into account the knowledge-base of EER. It is also claimed that these studies will contribute to the establishment of an evidence-based approach in policy making and a theory-driven approach towards improving education.

THEORY-DRIVEN EVALUATION STUDIES

Theory driven evaluation is a collection of different methodological approaches that can be used by evaluators in trying to understand the impact of a reform policy evaluation such as those of

programme theory, theories-of-change, and realism (Bledsoe & Graham, 2005; Rosas, 2005). In all of these perspectives, social programmes are regarded as products of the human imagination; they are hypotheses about social betterment (Bickman, 1985). Programmes chart out a perceived course where wrongs might be put right, deficiencies of behaviour corrected, and inequalities of condition alleviated. Programmes are thus shaped by a vision of change and social justice and they succeed or fail according to the veracity of that vision. In respect to these, evaluation has the task of testing out the underlying programme theories (Chen & Rossi, 1987) but also identifying unintended consequences, which may or may not be beneficial. When one evaluates, he/she always returns to the core theories about how a programme is supposed to work and then interrogates it by asking whether the basic plan is sound, plausible, durable, practical, and above all, valid.

Evaluation projects that are theory driven take into account the needs and issues raised by the various stakeholders associated with an innovation, such as the practitioners and the policy-makers. However, the evaluation agenda behind these projects are also not entirely defined by the stakeholders. The overall agenda is expanded in such a way as to allow evaluators to not only provide answers to the questions raised by stakeholders but also help them understand the reasons why a reform is more or less effective (Weiss, 1997). In this paper, it is argued that in order to provide such answers, evaluators in education should make use of the growing knowledge base of EER as it is concerned with the correlates of effective practice and provides theories about their relationships with each other and with student outcomes. Educational effectiveness can be seen as a theoretical foundation upon which can be built better evaluation studies in education. Further, programmes are embedded in social systems as they are delivered (Shaw & Replogle, 1996). As a result, it is through the workings of entire systems of social

relationships in and outside the classroom and/or the school that any changes in behaviours, events, and social conditions in education are put into effect. Serving to aide an understanding of variation within an effective implementation of a reform, theories of educational effectiveness can help evaluators identify factors most closely associated with the effective implementation. Moreover, in making use of these theories evaluators may also contribute to the development of the knowledge base of EER itself.

A typical example of a theory driven evaluation is the evaluation of a 1998 Cypriot reform that concerned the use of schema theory in teaching mathematics (Kyriakides, Charalambous, Philippou, & Campbell, 2006). Five years after the introduction of the reform, an evaluation study was conducted in order to determine its current implementation. The study aimed to examine the main stakeholders' (i.e. teachers' and students') reaction to the reform and the factors influencing its effectiveness. The study not only provided answers to policy-makers but also revealed that student achievement was determined by a number of factors related to teachers' and students' personal characteristics and teachers' reaction towards the reform itself. The research verified the decisive role of teachers in implementing any reform. Based on the findings of this study and drawing on the theoretical assumptions of the "emergent design" research model, a conceptual framework for conducting program evaluations was proposed which attributes a central role to teachers' classroom behaviour. It was claimed that teacher effectiveness research could be a foundation upon which to design studies regarding the evaluation of reforms. In turn, this study revealed that EER can be seen as a foundation upon which a theory driven evaluation project can be designed.

This study reveals that it is possible to combine theoretical models of EER with evaluation projects that have their agendas defined by different stakeholders for political and

practical reasons. Such projects contribute to the development of the knowledge base of EER and provide elaborate and better answers to the questions posed by the various stakeholders of education. In this context, this paper provides an overview of the recent theoretical model of EER that takes into account the dynamic nature of effectiveness and was developed in order to establish stronger links between EER and improvement of practice. This paper also draws implications of using the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008) for designing theory driven evaluation studies

THE DYNAMIC MODEL OF EDUCATIONAL EFFECTIVENESS: AN OVERVIEW

The development of the dynamic model is based on the results of a critical review of the main findings of EER and of the theoretical models of effectiveness which were developed in 1990s (Creemers & Kyriakides, 2006). This section refers to the main assumptions and elements of the dynamic model and to the main factors included in the model.

A) The Rationale of the Model

The dynamic model is based on the following three main assumptions. First, the fact that most of the effectiveness studies are exclusively focused on language or mathematics rather than on the whole school curriculum aims (cognitive, metacognitive and affective) reveals that the models of EER 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 are used in order to specify variables associated with the quality of teaching. Second, an important constraint of the existing approaches of modelling school

effectiveness is the fact that the whole process does not contribute significantly to the improvement of school effectiveness. Thus, the dynamic model is 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 (Creemers & Kyriakides, 2010a). Finally, the dynamic 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 expected to be interrelated within and/or between levels.

B) The Essential Characteristics of the Dynamic Model

The main characteristics of the dynamic model are as follows. First, the dynamic model takes into account the fact that effectiveness studies conducted in several countries reveal that the influences on student achievement are multilevel (Teddlie & Reynolds, 2000). Therefore, the model is multilevel in nature and refers to factors operating at the four levels shown in Figure 1. Figure 1 reveals the main structure of the dynamic model. Teaching and learning situation is emphasised and the roles of the two main actors (i.e., teacher and student) are analysed. Above these two levels, the dynamic model also refers to school-level factors. It is expected that school-level factors influence the teaching-learning situation by developing and evaluating the school policy on teaching and the policy on creating a learning environment at the school. The system level refers to the influence of the educational system through a more formal way, especially through developing and evaluating the educational policy at the national/regional level. It also is taken into account that the teaching and learning situation is influenced by the wider educational

context in which students, teachers, and schools are expected to operate. Factors such as the values of the society for learning and the importance attached to education play an important role both in shaping teacher and student expectations as well as in the development of the perceptions of various stakeholders about effective teaching practice.

Insert Figure 1 about here

Second, Figure 1 does not only refer to the four levels of the dynamic model and each level's association with student outcomes. The interrelations between the components of the model are also illustrated. In this way, the model supports that factors at the school and system level have both direct and indirect effects on student achievement since they are able to influence not only student achievement but also the teaching and learning situations.

Third, the dynamic model assumes that the impact of the school- and system- level factors has to be defined and measured in a different way than the impact of classroom-level factors. Policy on teaching and actions taken to improve teaching practice must be measured over time and in relation to the weaknesses that occur in a school. The assumption is that schools and educational systems which are able to identify their weaknesses and develop a policy on aspects associated with teaching and the school learning environment are also able to improve the functioning of classroom-level factors and their effectiveness status. Only changes in those factors for which schools face significant problems are expected to be associated with the improvement of school effectiveness. This implies that the impact of school and system level factors depends on the current situation of the objects under investigation (Creemers & Kyriakides, 2009). This characteristic of the proposed dynamic model does not only reveal an

essential difference in the nature of this model with all the integrated models of EER but, as it is explained in the next section, it has also some significant implications for using the dynamic model for improvement purposes.

Fourth, the dynamic model is based on the assumption that the relation of some effectiveness factors with achievement may not be linear. This assumption is supported by results of quantitative syntheses investigating the effect of some effectiveness factors upon student achievement. These studies revealed that although these variables 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. 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 of teacher direction, drill or recitation becomes dysfunctional (Soar & Soar, 1979). This implies that optimal points for the functioning of factors in relation to student outcomes have to be identified. By doing so, different strategies focusing on the improvement of specific factors for each teacher/school could emerge (Creemers & Kyriakides, 2006).

Fifth, the model assumes that there is a need to carefully examine the relationships between the various effectiveness factors which operate at the same level. Walberg's (1984) model, which

is one of the most significant educational productivity models, attempts to illustrate such relationships. Aptitude, instruction and the psychological environment are seen as major direct causes of learning. They also influence one another and are in turn influenced by feedback on the amount of learning that takes place. The Walberg's model was tested as a structural equation model on science achievement, indicating more complex, indirect relationships (Reynolds & Walberg, 1990). This implies that there is a need to refer to the relationships between the effectiveness factors which operate at the same level. Such approach to modelling school effectiveness reveals grouping of factors that make teachers and schools effective (see Kyriakides, Creemers, & Antoniou, 2009). Therefore, strategies for improving effectiveness which are comprehensive in nature may emerge.

Finally, the dynamic model is based on the assumption that different dimensions for measuring the functioning of effectiveness factors are used. The integrated models do not explicitly refer to the measurement of each effectiveness factor. On the contrary, it is often assumed that these factors represent unidimensional constructs. For example, the comprehensive model of educational effectiveness states that there should be control at school level, meaning that goal attainment and the school climate should be evaluated (Creemers, 1994). In line with this assumption, studies investigating the validity of the model revealed that schools with an evaluation policy focused on the formative purposes of evaluation are more effective (e.g., Kyriakides, Campbell, & Gagatsis, 2000; Kyriakides, 2005) However, the examination of evaluation 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 aspect of the functioning of evaluation such as the procedures used to design evaluation instruments, the forms of record keeping, and the policy on reporting results to parents and pupils.

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. More specifically, each factor is defined and measured using five dimensions: *frequency*, *focus*, *stage*, *quality*, and *differentiation*. *Frequency* is a quantitative way to measure the functioning of each effectiveness factor. The other four dimensions examine qualitative characteristics of the functioning of the factors and help us describe the complex nature of effective teaching. A brief description of these four dimensions is given below. Specifically, two aspects of the *focus* dimension are taken into account. The first one refers to the specificity of the activities associated with the functioning of the factor whereas the second one to the number of purposes for which an activity takes place. The *stage* at which tasks associated with a factor take place is also examined. 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. The *quality* refers to the properties of the specific factor itself, as these are discussed in the literature. Finally, *differentiation* refers to the extent to which activities associated with a factor are implemented in the same way for all the subjects involved with it (e.g., all the students, teachers, schools). It is expected that adaptation to specific needs of each subject or group of subjects will increase the successful implementation of a factor and will ultimately maximize its effect on student learning outcomes. The use of different measurement dimensions reveals that looking at just the frequency of an effectiveness factor (e.g., the quantity that an activity associated with an effectiveness factor is present in a system/school/classroom) does not help us identify those aspects of the functioning of a factor which are associated to student achievement. Considering effectiveness factors as multidimensional constructs not only provides a better

picture of what makes teachers and schools effective but may also help us develop specific strategies for improving educational practice (Kyriakides & Creemers, 2008a).

C) Classroom Factors of the Dynamic Model

Based on the main findings of teacher effectiveness research (e.g., Brophy & Good, 1986; Muijs & Reynolds, 2001; Rosenshine & Stevens, 1986), the dynamic model refers to factors which describe teachers' instructional role and are associated with student outcomes. These factors refer to observable instructional behaviour of teachers in the classroom rather than on factors that may explain such behaviour (e.g., teacher beliefs and knowledge and interpersonal competences). The eight factors included in the model are as follows: *orientation, structuring, questioning, teaching-modelling, applications, management of time, teacher role in making classroom a learning environment, and classroom assessment*. These eight factors do not refer only to one approach of teaching such as structured or direct teaching (Joyce, Weil, & Calhoun, 2000) or to approaches associated with constructivism (Schoenfeld, 1998). An integrated approach in defining quality of teaching is adopted.

D) School Factors of the Dynamic Model

School factors are expected to influence classroom-level factors, especially the teaching practice. Therefore, the dynamic model gives emphasis to the following two main aspects of the school policy which affect learning at both the level of students and teachers: a) school policy for teaching and b) school policy for creating a learning environment at school. Guidelines are seen as one of the main indications of school policy and this is reflected in the way each school level factor is defined. However, in using the term guidelines we refer to a range of documents, such

as staff meeting minutes, announcements, and action plans, which make the policy of the school more concrete to the teachers and other stakeholders. These two factors do not imply that each school should simply develop formal documents to install its policy. The factors concerned with the school policy mainly refer to the actions taken by the school to help teachers and other stakeholders have a clear understanding of what is expected from them to do. Support offered to teachers and other stakeholders to implement the school policy is also an aspect of these two school factors (Creemers & Kyriakides, 2010b).

Based on the assumption that the essence of a successful organization in the modern world is the search for improvement, the dynamic model is also concerned with the processes and the activities which take place in the school in order to improve the teaching practice and its learning environment. For this reason, the processes which are used to evaluate the school policy for teaching and the SLE are investigated. It is expected that evaluation mechanisms will generate data that will help schools to take decisions on how to improve the functioning of school factors. Thus, the following four overarching factors at the school level are included in the model:

- a. school policy for teaching and actions taken for improving teaching practice,
- b. evaluation of school policy for teaching and of actions taken to improve teaching,
- c. policy for creating a SLE and actions taken for improving the SLE, and
- d. evaluation of the SLE

EVALUATION STUDIES: LESSONS DRAWN FROM THE DYNAMIC MODEL

The fourth part of this paper is an attempt to identify implications of the dynamic model for the design of theory-driven evaluation studies. Four major implications of the dynamic model for the development of evaluation studies are drawn. These refer not only to the criteria that can be used

to evaluate a reform but also to the content of the evaluation (i.e., the aspects that can be covered). Moreover, we raise methodological issues associated with the measurement of factors that define the effective implementation of the reform and the quantitative analysis of data. Finally, a framework that can be used by evaluators to build a theory driven evaluation in line with the dynamic model is presented.

A) Criteria of Evaluation Studies Based on the Assumptions of the Dynamic Model:

Measuring the Impact of the Reform on Student Achievement

Theory-driven evaluation studies that take into account the main principles of the dynamic model are expected to recognise the importance of measuring the impact of any reform on student achievement. Irrespective of the nature of the reform, it is considered essential that evaluators adopting the dynamic model should search for the impact of the reform on student achievement. Given that the use of value-added assessment for measuring effectiveness is recommended, it is expected that evaluation studies should identify the extent to which the organisation units (e.g., schools or educational systems) which implement a reform improve their effectiveness status by looking at the progress that their students made during the implementation of the reform.

To achieve this purpose, both direct and indirect effects of the reform on student achievement can be measured. In the case of indirect effects, those evaluators who make use of the knowledge base of the dynamic model should try to find out whether the reform has any positive impact on the functioning of any effectiveness factor(s) of the dynamic model. For example, evaluators searching for the impact of a national reform in using IT in teaching mathematics should not only search for a direct impact of the reform on student achievement in mathematics. They should also find out whether the reform has any effect on teacher behaviour

in the classroom in respect to any of the eight classroom-level factors included in the dynamic model. In case that the reform helps teachers to improve their generic teaching skills, we could consider the reform effective since such a positive impact on teaching practice is expected to improve student learning.

B) Building an Evaluation Study that takes into account the Multilevel Nature of Education

Evaluators, in their attempt to identify the impact of reform on student achievement, should take into account the multilevel structure of education. This does not only mean that multilevel statistical modelling approaches should be used to measure the impact of a reform on student achievement. It also implies that evaluators should search for factors operating at different levels, specifically at the school, teacher, and student levels, which are likely to influence the effective implementation of a reform. It is also acknowledged that an effective reform policy may not necessarily have direct effects on student achievement, but it is more likely to have indirect effects. Therefore, in our attempt not only to design reforms but also to evaluate them for formative reasons, we should examine the extent to which the reform takes into account the skills of those associated with the reform to implement it. For example, a curriculum reform is expected to take into account not only the skills of teachers to teach the new teaching content or to use the proposed teaching approaches but also their generic teaching skills as these are defined by the dynamic model. The extent to which the reform helps teachers to improve these skills is seen as crucial for the effectiveness of the reform. Therefore, evaluators who conduct formative evaluation should search for ways that help policy-makers improve the effective implementation of the reform associated with improving its impact on quality of teaching. In this way the formative evaluation could be directed in the implementation of the reform in a rather specific

way since the extent to which the implementation of the reform contributes to the improvement of factors operating at the different levels of education can be examined.

If we turn back to the example given above, a significant aspect of a formative evaluation of a reform policy on using IT to teach mathematics is concerned with the impact of the reform on improving the learning environment of the school. Improvement of this school-level factor may allow teachers to cooperate and solve together the difficulties that they may have to face in introducing IT in their teaching practice. This example reveals that data of formative evaluation can be used by those who are responsible for designing the reform in order to make decisions on how to improve the impact of the reform on effectiveness factors associated with the nature of the reform. This assumption is one of the major lessons drawn from the evaluation of the curriculum reform in mathematics concerned with the schema theory mentioned above.

Specifically, the need to incorporate teacher effectiveness research in reform evaluation studies has been stressed (Kyriakides et al., 2006). Till recently, much emphasis was placed on the role of effective schools, in an attempt to specify the criteria that made a school able to introduce a reform and improve its effectiveness. However, during the last few years, there was a remarkable change: Attempts at policy and practice level have been made to focus upon teacher effects and generally on issues related to the effectiveness of teachers' work. This implies that those responsible for designing reform should bear in mind how the reform can help teachers improve their behaviour in the classroom.

C) Implications of the Dynamic Model for the Aspects of the Reform Covered by the Evaluation: Going Beyond Stakeholders' Reactions towards the Reform

Evaluators should take for granted that EER reveals that irrespective of the nature of the reform, there will be variation in the ability of teachers and schools to effectively implement the reform. This is attributed to the fact that teachers and schools have significant effects on student achievement. This implies that evaluators should search for characteristics of teachers and schools that make them more or less effective in implementing the reform. In this context, evaluation studies which are able to explain most of the variance in implementing the reform are, undoubtedly, essential in any reform movement since they provide a wealth of information that is helpful for any stage of reform implementation (Worthen et al., 1997) and thereby achieve the formative purpose of evaluation (Kyriakides, 2004).

Yet, when evaluating an educational reform, researchers often face difficulties in reducing their scope of examination since reforms constitute complex phenomena, and thus, numerous aspects of them are worth examining. Elaborating on the impact that different education actors (i.e., district administrators, inspectors, consultants, principals, parents, teachers, and students) have on the implementation of reform has been considered a sine qua non element of any reform evaluation (Amit & Fried, 2002; Atkin, 1998; Fullan, 1991; Kelly, 1989). Acknowledging the difficulties in studying the whole spectrum of factors operating in the school environment during the introduction and subsequent implementation of the reform, we argue that evaluators should at least focus their attention on how teachers' and students' reactions toward the reform affected the effective implementation of the reform.

Teachers' and students' reactions towards the reform

Having told teachers how to teach, what texts to use, and what theory of learning to believe in and follow, reform designers have usually taken the success of a reform for granted (Campbell,

1985). Repeated unsuccessful attempts to implement reforms in such a way suggested that this approach was nothing else but a prescription of failure. Indeed, research findings during the last two decades underlined that teachers can play a decisive role in the implementation and future success of an innovation (e.g., Kyriakides, 1997; Poletti, 2000; Ponte, Matos, Guimaraes, Leal, & Canavarro, 1994; Sztajn, 2003; van den Berg, Slegers, Geijsels, & Vandenberghe, 2000). As the picture constantly changes, teachers are increasingly considered by most policy-makers and school change experts to be the centrepiece of educational change (Datnow, Hubbard, & Mehan, 2002). Therefore, examining teachers' attitudes, thoughts, and criticism regarding a reform is judged imperative. Yet, the same emphasis does not seem to be placed on the students' role since students' impact on the implementation of the reform continues to be considered marginal, as evident by the scarcity of research in this domain. Acknowledging the significant role of both constituents of the teaching and learning process in affecting the implementation and success of a reform, the importance of investigating how teachers' and students' reactions may have influenced the effectiveness of a reform is pointed out.

To date, it is accepted that teachers do not passively respond to the directives mandated from higher levels of organisations; rather, they respond in a variety of ways to such directives--through advancing reform efforts, symbolically displaying reforms, or resisting them overtly or covertly (Datnow et al., 2002; Tyack & Cuban, 1995). This diversity of reactions can be attributed to the impact of characteristics related to the reform per se (or the way teachers conceive the reform) and/or to teachers' individual characteristics. Namely, research has so far illustrated that the magnitude of changes that teachers need to introduce into their traditional way of working, the estimation of the extra effort needed to address new requirements, the extent to which the reform addresses important needs, the way the reform is diffused, teachers'

involvement in the initiation of the innovation, and their content and pedagogical content knowledge regarding the reform are among the factors that exert great influence on teachers' reactions towards the reform (Fullan, 1991; Ghaith & Yaghi, 1997; Kelly, 1989; Levenberg & Sfar, 1996; Lloyd & Wilson, 1998; Polettini, 2000; Ponte et al., 1994). It has also been demonstrated that when teachers are not committed to the reform, they see the additional demands associated with them as increasing the stress and pressure of their already difficult jobs (Datnow et al., 2002; Duke, 2004). Though we do not underestimate the effect of these factors, we believe that emphasis should also be given to teachers' beliefs with respect to the reform since there is evidence suggesting that these beliefs serve as inhibitors or promoters of reform efforts (Datnow et al., 2003; Fullan, 1991).

Recently, research has suggested that teachers' perspectives and responses to a reform should not be considered independently of the context in which they operate. Therefore, it was stressed that rather than examining teachers' reactions uniformly, emphasis should also be given to the culture of schools since some schools seem to implement reforms in supportive ways while others seem to resist more to the introduction of the reform (Datnow et al., 2003). For example, knowing that colleagues in a school are implementing the reform successfully creates a productive atmosphere for teachers to experiment with the reform (Datnow, Borman, & Stringfield, 2000). This suggestion seems to be in line with the way the school level has been described in the dynamic model, especially since the model refers to the learning environment of the school. It is important to note, though, that educators' responses to a reform are never homogeneous, even within a school (Datnow et al., 2002). For instance, a recent survey research study by Beerens (2000) suggests that the variation that exists within schools can be greater than the variation across schools.

As far as the importance of investigating students' reactions towards the reform is concerned, we would like to point out that attempts to implement an educational reform are often linked to the following paradox: Even though students are regarded as potential beneficiaries of change, rarely is their attitude toward the reform taken into account (Fullan, 1991). However, during the last decade, the importance of investigating students' beliefs regarding the reform was underlined (Ponte et al., 1994); emphasis was also placed on students' self-efficacy beliefs (Pajares, 1999). Research is replete with evidence that students' efficacy beliefs are related to the goals they set; the activities they choose to engage in; their effort, energy expenditure, and persistence when pursuing certain outcomes; their use of cognitive strategies and self-regulated learning approaches; and their motivation and interest in certain domains. Moreover, there is evidence that high efficacious students have fewer adverse emotional reactions when they encounter difficulties; display lower levels of anxiety, stress and depression than students who doubt their capabilities; and possess intrinsic rather than extrinsic motives (Pajares, 1999; Pajares & Miller, 1994; Pintrich, 1999). It is pointed out that some of these factors affecting students' beliefs are also included in the dynamic model, such as their motivation and expectations.

Looking at the impact of the reform on the behaviour of stakeholders: The importance of looking at changes in teacher behaviour in the classroom

Yet, we acknowledge a number of limitations related to the idea that formative evaluation should be based on teachers' and students' reactions towards the reform. The findings of evaluation studies looking at the perceptions of stakeholders usually cannot help us explain the differences in the effective implementation of the reform from classroom to classroom and school to school (Kyriakides et al., 2006). This implies that instead of putting the blame for the ineffectiveness of

a reform on the process followed for its design and diffusion, evaluators should try to explain the differences in the way teachers react to a reform and to the different impact that the reform exerts on their existing teaching practices.

Looking at the way the factors included in the dynamic model have been defined, one could argue that the overemphasis witnessed in previous years on teachers' attitudes towards the reform should be replaced by a balanced emphasis on attitudes and behaviour in classroom. Evaluation studies reveal that a reform is seldom implemented as planned; teachers often make adaptations to the proposed reforms either to fit with their professional judgment and ideologies or to match the realities of their experiences and meet their students' needs (e.g., Campbell, 1985; Datnow et al., 2002, 2003; Kyriakides, 1997; Pollard, Broadfoot, Croll, Osborn, & Abbott, 1994; Tyack & Cuban, 1995; Woods, 1994). For instance, spending more time than that prescribed in the reform manuals was the main adaptation witnessed in a series of research conducted by Datnow and her colleagues (Datnow et al., 2002). It may, therefore, be concluded that teachers have a predominant contribution to the effectiveness of a reform. What matters in an educational reform is not the availability of supporting resources but the *quality* of teachers themselves and their generative role in the curriculum change, which determines the *quality of teaching* and, consequently, the effectiveness of any curriculum change. This argument is reflected in the fact that the dynamic model is based on the assumption that the classroom level is more significant than either the school or the context level. In addition, the classroom level is entirely defined in relation to the behaviour of the teacher in the classroom (Kyriakides & Creemers, 2008a).

With all this said, it seems pertinent to claim that EER should comprise another theoretical strand upon which reform evaluation studies could be based. The question is,

though, “In what ways could the description of teacher effectiveness in the dynamic model be helpful in evaluating a reform?” We assume that the dynamic model could be informative in at least two ways. First, it could provide a list of criteria that can help researchers elaborate on and study teachers’ practices during the reform implementation. Second, it could facilitate the investigation of teachers’ professional development, if there is any, during the implementation of the reform. We discuss each theme in turn.

Educational effectiveness research could aid in focusing on *teachers’ practice* and on the *quality of teaching* rather than solely elaborating on teachers’ beliefs (as a lot of evaluation studies did in the past) and their knowledge (as some recent studies have done). This suggestion is in accord with current research findings showing that teacher practices exert a stronger effect on students’ outcomes than their beliefs and knowledge (Seidel & Shavelson, 2007). Therefore, rather than perceiving teachers as a unified body of people which responds similarly when receiving the same stimulus (i.e., a reform), the dynamic model could provide criteria upon which teacher practices and effectiveness could be studied. For instance, we could focus on the *structure* of their lessons (e.g., Do teachers call attention to the main ideas underlying the theoretical background of the reform? Do they outline the content to be covered and signal transitions between lesson parts?) or on their *questioning techniques* (e.g., If the implementation of the reform requires the use of specific types of questions, do teachers use them successfully? What type of feedback do they provide to students’ answers?).

Were the dynamic model suggesting only a number of criteria for studying teachers’ practice in implementing a reform, it would offer nothing else but a blueprint for conducting observations. Fortunately, the contribution of the dynamic model is not so limited since it has the potential to help with studying teachers’ development as regards the reform, especially

since at both the school and the context level factors associated with the learning environment of the schools are included. This implies that empirical evidence is also needed to examine whether the introduction of a reform encourages teachers to modify their existing practice. Beyond portraying teachers' practices at a specific time, the dynamic model implies that we should also collect longitudinal data of teachers' classroom behaviour. Hence, we argue that the dynamic model may also help in providing a historical perspective of teachers' practice as regards the reform rather than a motionless picture of teachers' implementation of the reform at a specific time. This argument is partly based on the fact that a measurement framework is proposed by the dynamic model which, among others, refers to the stage of the functioning of a factor. Some further implications of the measurement framework of the dynamic model for the development of evaluation studies are drawn below.

D) Using the Measurement Framework of the Dynamic Model to Design Evaluation Studies and Build a Meta-evaluation System

The dynamic model is based on the assumption that each effectiveness factor operating at the classroom, school, and context level can be measured in relation to five dimensions. Therefore, evaluators could use this measurement framework in their attempt to measure each factor which may be associated with the effective implementation of a reform. It is also important to note that this framework can be used in analysing the characteristics of the reform. For this reason, we illustrate how we could measure the characteristics of the reform focused on the use of IT in teaching mathematics, mentioned above. As far as the frequency dimension is concerned, we may raise questions such as how many types of software are recommended to be used or how many lesson plans are offered to teachers in order to support them in using IT in teaching

mathematics. Second, the focus dimension is measured by investigating whether the guidelines given and/or the actions taken for improving teaching practice are too specific (or too general) in terms of what each teacher is expected to do. We also examine the purpose(s) that are expected to be achieved by the reform policy, and especially whether the reform aims to achieve a single-purpose (e.g., improving teaching practice) or more purposes. The reform may not only aim to solve a problem but to improve effectiveness of education in a more general way (e.g., the reform may be related not only to quality of teaching but also to the establishment of a learning environment in each region through the establishment of networks). Third, the stage dimension is seen as a very important measurement dimension since according to the dynamic model, the reform policy should be flexible, as this is reflected in the fact that, from time to time, changes in the reform policy take place. Continuity of the reform policy is also examined. Therefore, we measure the extent to which changes in the reform policy emerge from the results of a systematic evaluation of the reform.

Fourth, quality is measured by investigating the properties of the guidelines on using IT in teaching mathematics, especially whether these are clear, concrete, in line with the literature, and provide support to teachers, students, and administrators to implement the reform policy. Concrete guidelines include the kind of measures that should be taken in instances where teachers find that a problem with the implementation of the reform is about to be created. Finally, differentiation is measured by investigating the extent to which school policy is designed in such a way that more support is given to teachers and students who have difficulties in implementing the policy on provision of learning opportunities. For example, novice teachers could be asked to produce different type(s) of short term planning or to produce it more often

than other teachers. It may also be pointed out that students who need more learning opportunities should receive them.

The proposed measurement framework can also be used in investigating the implementation of a reform in different organisations (e.g., school units or classrooms). Specifically, frequency can be measured by looking at the extent to which the reform policy is implemented in a classroom or a school. As far as the focus dimension is concerned, we can investigate the extent to which teachers or schools implement a policy by following exactly what they have been asked to do or whether they are more flexible in implementing the policy in different organisation units (e.g., classrooms or schools). The second aspect of the focus dimension concerned with the number of purposes that are expected to be achieved can be examined by looking at the extent to which teachers/schools implement a reform to help them achieve the purposes mentioned by the policy-makers or their expectations for a reform are expanded to cover other aspects/problems of the functioning of their units. The stage dimension is measured by looking at the period in which the reform policy is implemented. It is expected that teachers and schools will try to implement the reform policy during the whole school year and not only at a certain period. Quality is measured by investigating the properties of the behaviour of teachers and schools in implementing the policy guidelines. Finally, differentiation is measured by investigating the extent to which the reform policy is implemented in such a way that more emphasis on the implementation of the reform is given to teachers and students who need the reform more. For example, a reform focused on realistic teaching in mathematics is expected to be implemented more frequently in those schools and classrooms that have high percentages of students with learning difficulties. It may also be pointed out that the reform is

implemented in such a way that students/teachers who need the reform more should receive it more than others.

We would finally like to mention that although an integrated aspect of the development of an evaluation system is the development of its meta-evaluation mechanisms, only rarely are such meta-evaluation mechanisms built. However, those evaluators who may like to develop a meta-evaluation mechanism to help them improve the quality of their evaluation studies could consider the possibility to build this mechanism around the five measurement dimensions of the dynamic model. This is also reflected in the fact that the factors concerned the evaluation of either school or national policy have been described in relation to these five dimensions (see Creemers & Kyriakides, 2008). The development of such mechanism can help the evaluators generate data which will help the various stakeholders to increase the impact of the reform on student achievement. For example, looking at the quality dimension of the evaluation of the reform will help policy-makers identify not only the psychometric properties of the evaluation and improve them but also the extent to which data gathered from evaluation help them make decisions on how to improve the reform and its impact on effectiveness.

A FRAMEWORK FOR CONDUCTING THEORY-DRIVEN EVALUATIONS: THE CONTRIBUTION OF THE DYNAMIC MODEL

The last section of this paper is an attempt to propose a possible theoretical framework for conducting theory-driven evaluations (see Figure 2), incorporating the elements of the dynamic model discussed above. Acknowledging the contribution of different factors identified in previous studies as predictors of the effective implementation of a reform, we considered it imperative to include those factors in the proposed model.

Insert Figure 2 about here

The following observations arise from Figure 2. First, we recommend that evaluators need to reformulate the research questions that policy makers may have in relation to a reform process. In doing so, the theory upon which the reform is based and the main characteristics of the dynamic model are taken into account. This implies that the multilevel structure of education and the factors operating at different level should at least be considered. For example, a reform programme looking at the reduction of the class size could look at its impact on quality of teaching and especially on the eight classroom-level factors before investigating the impact of the reform on student outcomes. The reformulation of the evaluation questions of the stakeholders can be the starting point of designing the evaluation plan. This plan is expected not only to look at the summative aspect of evaluation but also to the formative one. The later one is closely related to the implementation of the reform. Given that we expect that variation on the implementation of the reform will exist, we propose that evaluators need to focus their attention on the behaviour of those expected to make use of the reform. Data concerning the impact of the reform on teachers and students behaviour as well as on the behaviour of other stakeholders may help us identify factors associated with the effective implementation of the reform. The dynamic model may be of use to search for the impact of such factors and may also provide suggestions on how the reform can be redesigned and provide further support to those who need it. Rather than discussing issues related to the existence of prescribed plans for implementing the reform, we propose that we need to examine how teachers use and modify these plans to meet student needs and promote learning. Instead of placing so much emphasis

on students' reactions towards a reform, we consider it important to examine what learning opportunities students are provided by participating in the reform.

Furthermore, the model proposed here suggests that beyond examining students' progress in terms of learning outcomes, we need to collect longitudinal data for both teachers and students. Namely, we suggest that it is worth examining both the short term as well as the long-term effects of the reforms on students since there is evidence that reforms and intervention programs may not have enduring effects on student learning (Kyriakides & Creemers, 2008b; Plewis, 2000). The model also suggests that evaluators could examine whether teachers improve their practices throughout the years as a consequence of implementing the reform (i.e., the reform itself could be considered a force able to bring change in teachers' practices).

The proposed framework does not aim to provide a comprehensive model for evaluating educational reforms. Rather, it aims to incorporate different theoretical frameworks into a single model, acknowledging the fact that each theoretical framework could illuminate different aspects of the reform. It is also argued that the dynamic model could have an important role in this process which ultimately is expected to contribute to the improvement of quality in education by using an evidence-based and theory-driven approach (Creemers & Kyriakides, 2012).

References

Amit, M., & Fried, M.N. (2002). Research, Reform, and Times of Change. In L. D. English (Ed.), *Handbook of International Research in Mathematics Education* (pp. 355-381). New Jersey: Lawrence Erlbaum Associates.

- Atkin, J.M. (1998). The OECD study of innovations in science, mathematics and technology education. *Journal of Curriculum Studies*, 30(6), 647-660.
- Beerens, D.R. (2000). *Evaluating teaching for professional growth*. Thousands Oaks, California: Corwin Press, Inc. Sage.
- Bickman, L. (1985). Improving established statewide programs – a component theory of evaluation. *Evaluation Review*, 9(2), 189–208.
- Bledsoe, K.L., & Graham, J.A. (2005). The use of multiple evaluation approaches in program evaluation. *American Journal of Evaluation*, 26(3), 302-319.
- Borich, G.D. (1992) (2nd Ed). *Effective teaching methods*. New York: Macmillan Publishing Company.
- Brookover, W.B., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School systems and student achievement: schools make a difference*. New York: Praeger.
- Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed, pp. 328–375). New York: MacMillan.
- Campbell, R.J. (1985). *Developing the Primary School Curriculum*. London: Cassell.
- Chen, H.T., & Rossi, P.H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning*, 10(1), 95-103.
- Coleman, J.S., Campbell, E.Q., Hobson, C.F., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.
- Creemers, B. (1994). *The Effective Classroom*. London: Cassell.

- Creemers, B.P.M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366.
- Creemers, B.P.M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Creemers, B.P.M., & Kyriakides, L. (2009). Situational effects of the school factors included in the dynamic model of educational effectiveness. *South African Journal of Education, 29*(3), 293–315.
- Creemers, B.P.M., & Kyriakides, L. (2010a). Using the dynamic model to develop an evidence-based and theory-driven approach to school improvement. *Irish Educational Studies, 29*(1), 5–23.
- Creemers, B.P.M., & Kyriakides, L. (2010b). School factors explaining achievement on cognitive and affective outcomes: establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research, 54*(3), 263-294.
- Creemers, B.P.M., & Kyriakides, L. (2012). *Improving Quality in Education: Dynamic Approaches to School Improvement*. London: Routledge.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives, 8*(1), <http://epaa.asu.edu/epaa/v8n1/>.
- Datnow, A., Borman, G., & Stringfield, S. (2000). School Reform through a Highly Specified Curriculum: Implementation and Effects of the Core Knowledge Sequence. *The Elementary School Journal, 101*(2), 167–191.
- Datnow, A., Borman, G., Stringfield, S., Overman, L.T., & Castellano, M. (2003). Comprehensive School Reform in Culturally and Linguistically Diverse Contexts:

- Implementation and Outcomes form a Four-Year Study. *Educational Evaluation and Policy Analysis*, 25(2), 143–170.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending Educational Reform: From one school to many*. New York: RoutledgeFalmer Press.
- De Jong, R., Westerhof, K.J., & Kruiter, J.H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement*, 15(1), 3–31.
- Driessen, G.W.J.M., & Mulder, L.W.J. (1999). The enhancement of educational opportunities of disadvantaged children. In R.J. Bosker, B.P.M. Creemers, & S. Stringfield (Eds.), *Enhancing educational excellence, equity and efficiency: evidence from evaluations of systems and schools in change*, (pp 37-64). Dordrecht: Kluwer Academic Publishers.
- Duke, D. (2004). *The Challenges of Educational Change*. Boston: Allyn and Bacon.
- Edmonds, R.R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15–27.
- Fullan, M. (1991). *The New Meaning of Educational Change*. New York: Cassell.
- Ghaith, G., & Yaghi, H. (1997). Relationships among Experience, Teacher Efficacy, and Attitudes towards the Implementation of Instructional Innovation. *Teaching and Teacher Education*, 13(4), 451-458.
- Jencks, C., Smith, M., Acland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: a Reassessment of the Effects of Family and Schooling in America*. New York: Basic Books.
- Joyce, B., Weil, M., & Calhoun, E. (2000). *Models of teaching*. Boston: Allyn & Bacon.

- Kelly, A.V. (1989). *The Curriculum: Theory and Practice*. London: Paul Chapman Publishing.
- Kyriakides, L. (1997). Primary teachers' perceptions of policy for curriculum reform in Mathematics. *Educational Research and Evaluation*, 3(3), 214–242.
- Kyriakides, L. (2004). Investigating Validity from Teachers' Perspective through their engagement in Large-Scale Assessment: the Emergent Literacy Baseline Assessment Project. *Assessment in Education: Principles, Policy and Practice*, 11(2), 143-165.
- Kyriakides, L. (2005). Extending the Comprehensive Model of Educational Effectiveness by an Empirical Investigation. *School Effectiveness and School Improvement*, 16(2), 103-152.
- Kyriakides, L. (2006). Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis of TIMSS 1999 data, *Educational Research and Evaluation*, 12(6), 513–534.
- Kyriakides, L., Antoniou, P., & Maltezou E. (2009, April). *Investigating the short- and long-term effects of secondary schools upon academic success and development*. Paper presented at the 90th Annual Meeting of the American Educational Research Association. San Diego, USA.
- Kyriakides, L., Campbell, R.J., & Gagatsis, A. (2000). The Significance of the Classroom Effect in Primary Schools: An Application of Creemers' Comprehensive Model of Educational Effectiveness. *School Effectiveness and School Improvement*, 11(4), 501–529.
- Kyriakides, L., Charalambous, C., Philippou, G., & Campbell, R.J. (2006). Illuminating reform evaluation studies through incorporating teacher effectiveness research: a case study in Mathematics. *School Effectiveness and School Improvement*, 17(1), 3–32.
- Kyriakides, L., & Creemers, B.P.M. (2008a). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the

- validity of the dynamic model. *School Effectiveness and School Improvement*, 19(2), 183–205.
- Kyriakides, L. & Creemers, B.P.M. (2008b). A longitudinal study on the stability over time of school and teacher effects on student learning outcomes. *Oxford Review of Education*, 34(5), 521-545.
- Kyriakides, L., & Creemers, B.P.M. (2009, August). *Explaining Stability and changes in schools: A follow-up study testing the validity of the dynamic model*. Paper presented at the EARLI conference. Amsterdam, the Netherlands.
- Kyriakides, L., Creemers, B.P.M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education*, 25(1), 12–23.
- Levenberg, I., & Sfard, A. (1996). When change becomes the name of the game: Mathematics teachers in transition to a new learning environment. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education*, vol. 3, (pp. 249-256). Valencia: University of Valencia.
- Levine, D.U., & Lezotte, L.W. (1990). *Unusually effective schools: a review and analysis of research and practice*. Madison (USA): National Center for Effective Schools Research and Development.
- Lloyd, G. M., & Wilson, M. (1998). Supporting Innovation: The Impact of a Teacher's Conceptions of Functions on His Implementation of a Reform Curriculum. *Journal for Research in Mathematics Education*, 29(3), 248-274.

- MacDonald, B. (1991). Critical introduction from innovation to reform – a framework for analysing change. In Rudduck, J. (Ed.), *Innovation and Change: Developing Involvement and Understanding*, Milton Keynes: Open University Press.
- Madaus, G.G., Kellagham, T., Rakow, E.A., & King, D. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review*, 4, 207-230.
- Monk, D.H. (1994). Subject matter preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Somerset, UK: Open Books.
- Muijs, D., & Reynolds, D. (2001). *Effective Teaching: evidence and practice*. London: Sage.
- Pajares, F. (1999). Current Directions in Self-Efficacy Research. In M. Maehr, & P.R. Pintrich (Eds.), *Advances in motivation and achievement*. (pp. 1–49). Greenwich, CT: JAI Press.
- Pajares, F., & Miller, M. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86(2), 193-203.
- Pintrich, P.R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31, 459-470.
- Plewis, I. (2000). Evaluating Educational Interventions Using Multilevel Growth Curves: The Case of Reading Recovery. *Educational Research and Evaluation*, 6(1), 83-101.
- Polettini, A.F.F. (2000). Mathematics teaching life histories in the study of teachers' perceptions of change. *Teaching and Teacher Education*, 16, 765–783.
- Pollard, A., Broadfoot, P., Croll, P., Osborn, M., & Abbott, D. (1994). *Changing English Primary Schools? The Impact of the Education Reform Act at Key Stage 1*. London: Cassell.

- Ponte, J.P., Matos J.F., Guimaraes, H.M., Leal, L.C., & Canavarro, A.P. (1994). Teachers' and Students' Views and Attitudes towards a New Mathematics Curriculum: A Case Study. *Educational Studies in Mathematics*, 26, 347–365.
- Pustjens, H., Van de Gaer, E., Van Damme, J., & Onghena, P. (2004). Effect of Secondary Schools on Academic Choices and on Success in Higher Education. *School Effectiveness and School Improvement*, 15(3–4), 281–311.
- Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (Eds.) (2002). *World Class Schools: International Perspectives on School Effectiveness*. London: RoutledgeFalmer.
- Reynolds, A.J., & Walberg, H. J. (1990). *A Structural Model of Educational Productivity*, Illinois: Northern Illinois University.
- Rosas, S.R. (2005). Concept mapping as a technique for program theory development – an illustration using family support programs. *American Journal of Evaluation* 26(3), 389-401.
- Rosenshine, B., & Stevens, R. (1986). Teaching Functions. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 376–391). New York: Macmillan.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. London, Office for Standards in Education and Institute of Education.
- Sammons, P., Power, S., Elliot, K., Campbell, C., Robertson, P., & Whitty, G. (2003). *New Community Schools in Scotland: Final Report – national evaluation of the pilot phase*. Edinburgh: Scottish Executive Education Department.

- Scheerens, J. (1992). *Effective Schooling: Research, Theory and Practice*. London: Cassell.
- Scheerens, J., & Bosker, R.J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Schoenfeld, A.H. (1998). Toward a theory of teaching in context. *Issues in Education*, 4(1), 1–94.
- Schon, D.A. (1971). *Beyond the Stable State*. Harmondsworth: Penguin.
- Schunk, D.H. (1991). Self-Efficacy and academic motivation. *Educational Psychologist*, 26(3), 207–231.
- Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education*, 8, 151–184.
- Seidel, T., & Shavelson, R.J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454-499.
- Shaw, K.M., & Replogle, E. (1996). Challenges in evaluating school-linked services – toward a more comprehensive evaluation framework. *Evaluation Review*, 20(4), 424–469.
- Soar, R.S., & Soar, R.M. (1979). Emotional climate and management. In P. Peterson & H. Walberg (Eds.), *Research on Teaching Concepts: Findings and Implications*. Berkeley, CA: McCutchan.
- Stevenson, H.W., Chen, C., & Lee, S.Y. (1993). Mathematics Achievement of Chinese, Japanese and American Children: Ten Years Later. *Science*, 259, 53–58.
- Stringfield, S.C., & Slavin, R.E. (1992). A hierarchical longitudinal model for elementary school effects. In B.P.M. Creemers & G.J. Reezigt (Eds.), *Evaluation of Educational Effectiveness*, pp. 35–69. Groningen: ICO.

- Sztajn, P. (2003). Adapting Reform Ideas in Different Mathematics Classrooms: Beliefs Beyond Mathematics. *Journal of Mathematics Teacher Education*, 6, 53–75.
- Taggart, B., & Sammons, P. (1999). Evaluating the impact of raising school standards initiative. In R.J. Bosker, B.P.M. Creemers, & S. Stringfield (Eds.), *Enhancing educational excellence, equity and efficiency: evidence from evaluations of systems and schools in change*, (pp. 137-166). Dordrecht: Kluwer Academic Publishers.
- Teddlie, C. (1994). The integration of classroom and school process data in school effectiveness research. In D. Reynolds et al. (Eds). *Advances in school effectiveness research and practice* (pp. 111-133). Oxford: Pergamon.
- Teddlie, C., & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer Press.
- Teddlie, C., & Stringfield, S. (1993). *Schools make a difference. Lessons learned from a 10-year study of school effects*. New York: Teachers College Press.
- Tyack, D. & Cuban, L. (1995). *Tinkering Toward Utopia: A Century of Public School Reform*. USA: Harvard University Press.
- van den Berg, R., Slegers, P., Geijsel, F., & Vandenberghe, R. (2000). Implementation of an Innovation: Meeting the Concerns of Teachers. *Studies in Educational Evaluation*, 26, 331–350.
- Walberg, H.J. (1984). Improving the productivity of America's schools. *Educational Leadership*, 41(8), 19–27.
- Weiss, C.H. (1997). How can theory-based evaluation make greater headway. *Evaluation Review*, 21(4), 501-524.

Woods, P. (1994). Adaptation and self-determination in English Primary Schools. *Oxford Review of Education*, 20(4), 387-410.

Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L. (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines* (2nd ed.). USA: Longman Publishers.

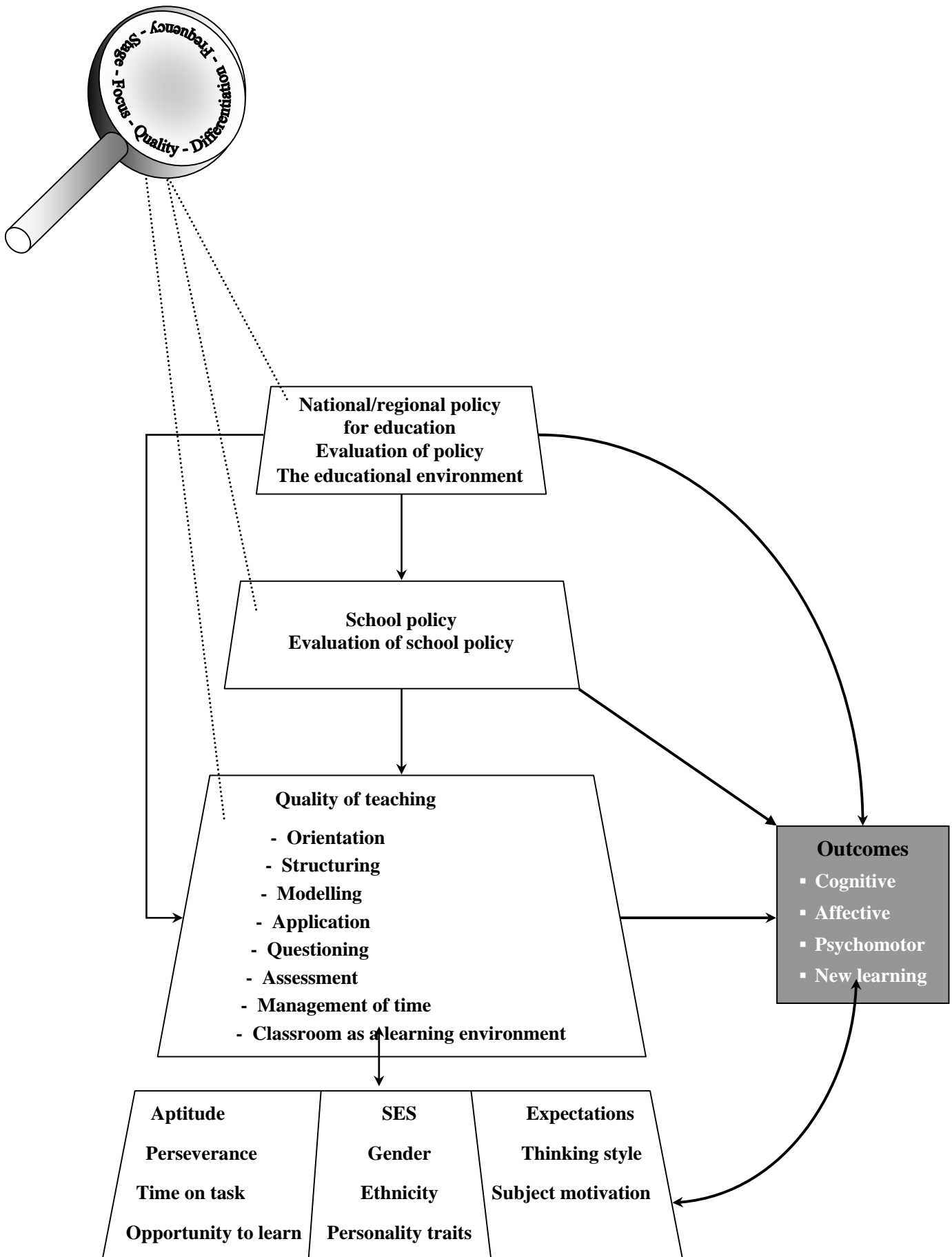


Figure 1: The dynamic model of educational effectiveness

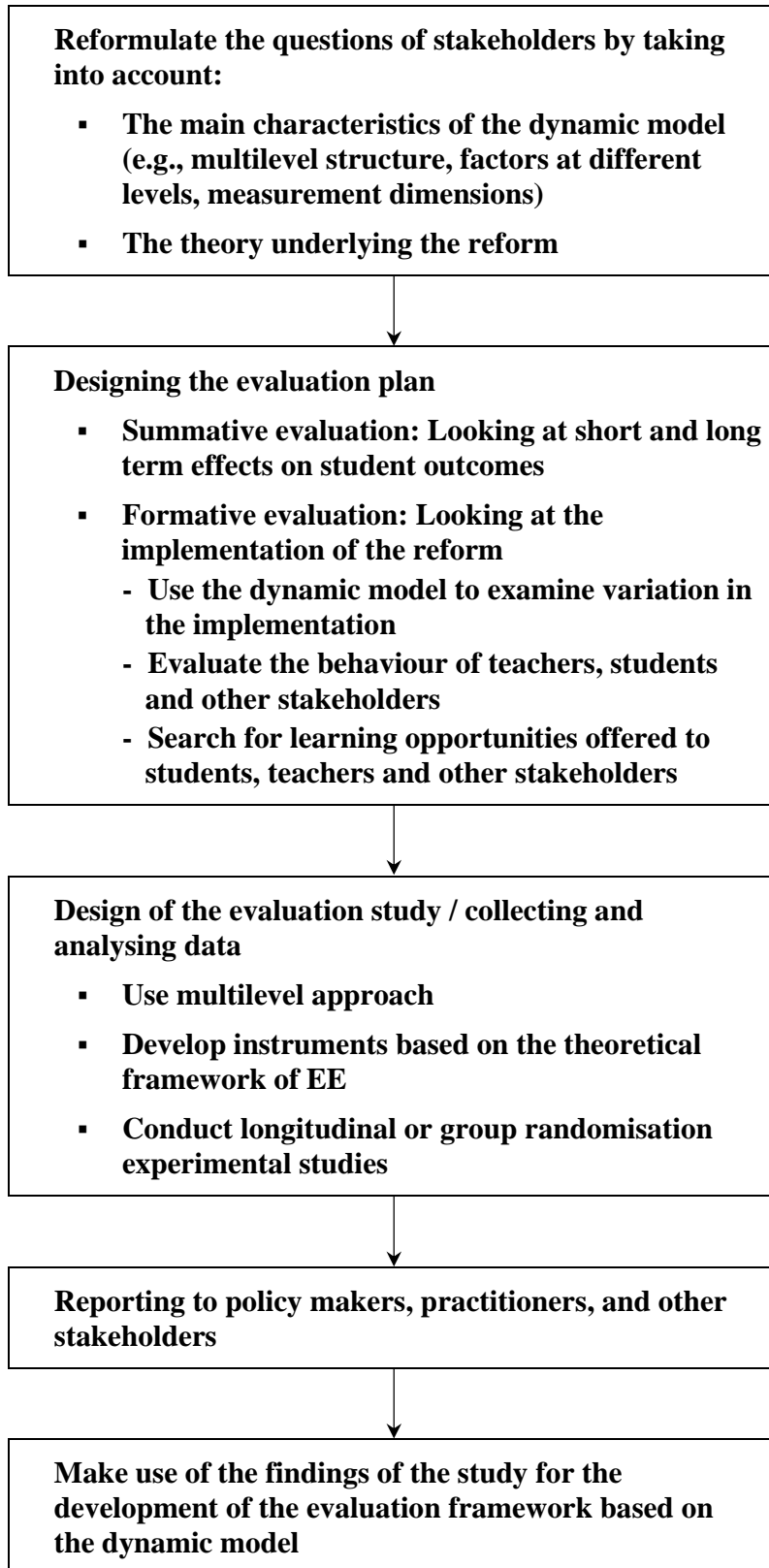


Figure 2: An evaluation framework inspired by the dynamic model