Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. & Stringfield, S. (2014) Educational effectiveness research (EER): a state-of-the-art review, *School Effectiveness and School Improvement*, 25:2, 197-230.

Introduction

Educational effectiveness research (EER) has shown rapid growth in the quantity and quality of the research answers it can produce to its core, foundational questions, which can be simply summarized as:

- What makes a "good" school?
- How do we make more schools "good"?

It seeks to investigate all the factors within schools in particular, and the educational system in general, that might affect the learning outcomes of students in both their academic and social development, which means it encompasses a wide range of factors such as teaching methods, the organisation – formally and informally – of schools, the curriculum, the role of leadership and the effects of educational "learning environments" in general, whether schools, Districts or Nations.

To undertake this complex task, EER involves distinguishing the effects of schools from other effects such as those of student intake and educational background, which therefore requires modelling tools that involve comparable complexity (Creemers, Kyriakides & Sammons, 2010); (Goldstein, 2003). Thus, the methodological issues of how we analyse complex data from multiple levels of the educational system have had, and continue to have, a salience in EER more than in many other educational specialities.

In this review we consider briefly after this introduction:

- The history or "phases" of EER;
- Theoretical and methodological advances in EER;
- The scientific properties of EER on the size of school effects, differential effects, their consistency, their stability over time, and the different outcomes of education;
- The processes of educational effectiveness, and ineffectiveness;
- The historical links between EER, and policy and practice;
- The future research needs and possible policy and practice links of EER;
- The future directions for EER, in a changing international society.

We should note that we are using the term "Educational Effectiveness Research" rather than "School Effectiveness Research" here as a reflection of the wide range of topics that the field has now encompassed in its forty year existence, and the breadth it now covers (Muijs, 2006; Creemers, Kyriakides & Sammons, 2010).

The history of EER

There are a large number of publications which review the field's history (Creemers, Kyriakides & Sammons, 2010; Gray et al., 1996; Mortimore, 1991; Reynolds, 2010b; Reynolds et al., 1994; Sammons, 1999; Scheerens & Bosker, 1997; Teddlie, 2010; Townsend, 2007; Teddlie & Reynolds, 2000), so we will only briefly summarize the main phases that EER has evolved through here. There are probably five.

The first phase emerged as a reaction to the seminal studies of Coleman et al. (1966) and Jencks et al. (1972) that concluded that schools had little effect upon the outcomes of their students in comparison with the effects of students' own ability and social backgrounds. The belief was commonplace that "schools make no difference" and that "education cannot compensate for society" (Bernstein, 1968). Into this climate came the empirical studies of Edmonds (1979), Rutter et al. (1979), Smith and Tomlinson (1989), Mortimore et al. (1988) and the smaller, "one off" case studies of researchers such as Weber (1971) and Reynolds (1976).

The second phase ran from the mid-1980s, in which the use of multilevel methodologies (Goldstein, 2003, 1995) and methodologically sophisticated studies began to show the scientific properties of school effects in such areas as the stability of school effects over time. These areas include their consistency upon different outcome domains, their differential effects upon students of different background characteristics, their size and their impact in the long-term (see reviews in Reynolds, 1996).

Beginning probably in the early to mid-1990s was EER's third phase, in which there were numerous attempts to explore the reasons *why* schools had their different effects. Influential work here was the Louisiana School Effectiveness Studies of Teddlie and Stringfield (1993) in the US, and work in the UK into subject department effects upon performance and also upon school effects (Sammons, Thomas & Mortimore, 1997). These years also saw a number of

influential reviews of the field such as by Scheerens and Bosker (1997), Reynolds et al. (1996) and Teddlie and Reynolds (2000). The importance of the classroom level was also noted by Creemers (1994) and Teddlie and Stringfield (1993).

The fourth phase, which began in the mid- to late-1990s, ran for a decade and is still in evidence today. This is the marked internationalisation of the field, together with the merger or synergy of approaches generated by having, for example, school effectiveness researchers in close intellectual proximity to school improvement researchers and practitioners. Examples of some large-scale studies are the studies of Brandsma and Knuver (1988) on primary schools and those of Bosker and van der Velden (1989) on secondary schools in the Netherlands, the study of Grisay (1996) on secondary schools in France, of Hill and Rowe (1996) on primary and secondary schools in Australia and those of Van Damme et al. (2006), De Fraine et al. (2007) and Verachtert et al. (2009) in Flanders. The international opportunities for networking, for joint research in multiple countries and the powerful effects of different research and country traditions in EER being opened up for learning from, for view, and for choice, meant that the field developed rapidly. Also, after calls for a merger of school effectiveness (SE) and school improvement (SI) concerns (e.g. Reynolds, Hopkins, & Stoll, 1993), many SE researchers became more comfortable with SI's qualitative methodology, its commitment to more cultural views of school processes instead of to the formal organizational factors that had been the SE commitment, and its commitment to the importance of seeing teachers as something other than mere "empirical/rational" educational actors. In this phase, there was growing concern to develop the theoretical underpinnings of the field and to develop theoretical models. Methodologically, there was pioneering mixed method work involving large-scale quantitative analysis combined with in-depth case study work into particular schools and departments.

The fifth phase has begun in the late 2000s and is still developing rapidly, focusing on EER as a dynamic, not static, set of relationships and moving away from seeing education in particular as an inherently stable set of arrangements towards one that sees the various "levels" of the educational system interacting and achieving variable outcomes (Creemers & Kyriakides, 2008). Additionally, and linked with this more dynamic perspective, is a commitment to newer forms of statistical analyses that can permit the establishment of indirect as well as direct relationships between educational factors and student outcomes, and which can permit reciprocal

relationships between educational factors to be established, both permitted by the increased popularity of Structural Equation Modelling (SEM). We consider these issues in more detail below.

In parallel with the phases described, there was an international trend towards more evidence-based education. From the early "best-evidence syntheses" of Slavin and others in the eighties, the development of "Success for All" of the same group and the STAR-experiment on the effect of class size in Tennessee, the movement led to many intervention studies (real experiments or cluster-randomized controlled trials, where possible) to "prove" that some practices were effective, and to reviews to inform practitioners and policy makers about what is working in education (Borman et al., 2003). The *"What works" Clearinghouse* and the *Best evidence encyclopaedia* in the USA, and the Interactive Best-evidence synthesis sponsored by the Ministry of Education in New Zealand, together with the book *Visible learning* by Hattie (2009), are well-known examples of these.

In cases where experiments were not possible, new techniques (mostly taken from econometrics) could help to allow causal inferences: the instrumental variables approach, regression discontinuity, difference-in-difference and propensity scores and other matching techniques.

Another way to inform the policy makers has been the international studies of educational achievement, from IEA and the OECD. These studies are mostly cross-sectional, but the analyses can also be longitudinal at the country level (Gustafsson, 2007). In many countries where the results were unexpectedly low, these studies were a major stimulus to do more research on the effects of the educational system at all levels. This was especially the case in Germany, which started with extensions of PISA (PISA – Konsortium Deutschland, 2006), and followed this by a series of longitudinal studies.

We should note that the early phases of EER may have laid down certain foundations that may have critically affected the intellectual and practical growth of EER. Of course, EER generated an international organisation – the International Congress for School Effectiveness and Improvement (ICSEI) – a journal (*School Effectiveness and School Improvement*) located in the ISI *Current Contents* databases within six years of its launch, and a volume of research books. These and papers justified over 1,500 references in the Teddlie and Reynolds (2000)

International handbook of school effectiveness research, and probably double that in the replacement volume currently being written and planned for 2012 publication. In more recent years, new organisational structures were created, especially the Society for Research on Educational Effectiveness (SREE) in the USA, with first one and now two annual conferences in Washington DC, the SIG on Educational Effectiveness in the European Association for Learning and Instruction with bi-annual conferences, and the network on Educational Effectiveness and Quality Assurance of ECER, the other European conference. ICSEI has also developed a strand of researchers involved in the Methodology of Research in Educational Effectiveness (MORE). All these factors suggest considerable scientific success, which is also indicated by new journals such as the *Journal of Educational Effectiveness, Effective Education* and, in Germany, the *Journal for Educational Research/Journal für Bildungsforshung* online.

However, much of the content of the "paradigm" of the discipline was in retrospect reactive, not purposive, and may have had negative consequences. Firstly, the commitment to quantitative methods within EER was easy to understand, given the need to demonstrate clearly that schools varied in their effects upon student outcomes, and given that few education researchers anywhere believed this in the 1980s. The methodological apparatus that was necessary to sort out the varying influences of school intakes of students upon outcomes, the need for large sample sizes, the need for individual data on each student at intake and outcome and the increasingly recognized value of multiple measures of educational outcomes that reflected the multiple nature and goals of schooling, all necessitated sophisticated quantitative methods and approaches.

However, the absence of much qualitative data that would have exemplified what the quantitative data were showing, and the absence of many "mixed methods" studies where quantitative data could "*demonstrate*" relationships between educational factors and qualitative data could "*explain*" the relationships, made the field and its findings difficult to access, particularly for policy makers and practitioners. Put simply, EER lacked rich, thick descriptions. Even where factors were used which had often been the concern of qualitative educational researchers historically – teacher attitudes, school ethos, learning environment, school culture and students' views, for example – and which had been the subject of their "rich, thick descriptions", many EER studies substituted questionnaires, formal interviews and standardized

systems of measurement, very few of which even had qualitative data as examples of what the quantitative instruments were measuring, although there were some innovative mixed methods approaches (e.g. MacBeath & Mortimore, 2001; Sammons, Thomas & Mortimore, 1997; Teddlie & Stringfield, 1993; Townsend 1994; Stringfield & Teddlie, 1988). Of course, over time, the intention of a growing number of quantitative studies was to "explain" why certain educational processes may be working by analysing data from intervention studies and including "mediators" to explain how the observed effects can be explained (e.g. Raudenbush, 2011).

Second, because EER grew up in a climate where it was believed "schools make no difference", our assertion that schools *did* make a difference meant that EER locked itself into an almost exclusive concern with the school, rather than with the District/Local Authority and particularly rather than with the classroom and with the teacher. This is despite the separate tradition of teacher effectiveness that has developed in parallel. Teddlie and Stringfield (1993) and Creemers (1994) were among the first to point this out, but it has taken another decade for "teaching" to receive anything like the attention given to "schooling" in EER. Given the clear evidence that teacher effects greatly exceed school effects when progress over time is studied (Muijs & Reynolds, 2010; Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000), this school-based concern may well have hindered the explanatory power of EER and possibly its practitioner take-up too, given practitioners' liking for discussions about pedagogy more than school organisation.

In Europe, the study of teacher effects has been hampered by the lack of suitable data, given the need to collect data at the start and end of the school year to assess student progress. The clearest contrasting case is the US, where many Districts (and, under No Child Left Behind legislation, all States) test all students in several consecutive grades every year, thus explaining why the US moved ahead more rapidly in the area of teacher effects (Brophy & Good, 1986; Rosenshine & Stevens, 1986). Many assessment systems only collect data at particular ages and, of course, the concerns that value-added studies on teacher effects may feed into judgements of teacher performance make this whole area a difficult one. It is increasingly recognized that, due to the numbers of students in classes and issues of statistical uncertainty, value-added estimates of teacher effects are much better used for research and improvement purposes to establish the features of effective practice than for decisions about teacher competence, pay or progression

(McCaffrey et al., 2004). For some recent examples of attention given to teacher effects and their implications for improving practice, see Day et al. (2006); Muijs and Reynolds (2010); Opdenakke and Van Damme (2006); Teddlie et al. (2006); and Van der Grift et al. (2007).

Third, the rise of Multi-level Modelling (MLM) was clearly essential to faithfully represent the reality of educational systems in which students learned or not in accordance with variation in their background characteristics, and in which they attended classes nested in schools, which in turn were nested in Districts/Local Authorities, and which in turn were nested in regions and nations. MLM made possible much more than the simple historical analysis of a "means on means" variety that the multiple regression methodologies of the early studies had utilized – it made possible the handling of multiple variations, differential effects and cross-level interactions.

However, the early MLM approaches tended to study *direct* educational effects – of teachers on students, for example – rather than indirect or reciprocal relationships, of students on teachers through their behaviours, for example. Additionally, in its early development MLM did not itself permit study of the interaction *between* levels that could be seen as the true "black box" of education – the decisions made at school level, for example, that have major effects upon the classroom level through the allocation of teachers, their professional development or their "empowerment". Put simply, the transactions between levels were not a focus of early MLM, since MLM was "sold" to EER quite consciously as a solution to the "means on means" problem.

Nonetheless, in more recent years as MLM has developed in sophistication there has been increasing interest in the study of cross-level interactions and reciprocal relationships, although the demands in terms of sample size are limiting factors. In addition, the use of cross-classified models has enabled the simultaneous study of primary and secondary school effects, and of neighbourhood influences also.

Fourth, the rapid internationalisation and growth of the field potentiated its chances of learning *from* itself. It is also likely that it reduced its chance of learning even more from other fields, given that there was little apparent need, given the intellectual and geographical variance that was already on offer, to learn from fields outside. The rise of a number of critical perspectives on the discipline (see the papers from the AERA debate such as Reynolds &

Teddlie, 2001; Slee & Weiner, 2001; Teddlie & Reynolds, 2001; Thrupp, 2001, 2002, together with the overview by Townsend, 2001a, 2001b) probably encouraged the discipline to focus inwards to address these issues through international collaboration, even though there were useful perspectives to be had from disciplines such as:

- Sociology of education (on school culture)
- Educational administration (on leadership)
- Human relations (on within-school relationships)
- Cognitive neuroscience (on "brain-based" interventions)
- Psychology of education (on multiple outcomes)
- Humanistic psychology (on self-conception).

Fifth, the very rapid growth in the number of researchers and studies over a short thirty-year time period probably made it difficult for EER to be in its knowledge base cumulative. Studies in EER often conceptualized the factors involved in EER differently – each study often then measured the factors differently, and then often employed different analytic methods. Whilst none of this is surprising in the development of an infant discipline, the sheer variability and its lack of cumulative work made the "body of knowledge" in EER a difficult one to assess and weigh. The *International handbook* (Teddlie & Reynolds, 2000) provided a good foundation and more recent work has further strengthened this through a focus on international research (Townsend, 2007), theory (Creemers & Kyriakides, 2008) and methodological advances (Creemers, Kyriakides, & Sammons, 2010), but these reviews all came relatively late in the development of the field.

Of course, there have been numerous criticisms of EER throughout its history, going back to the Rutter et al. (1979) research, concerning small sample size and lack of theoretical orientation to contemporary concerns that EER is largely post-positivistic in nature and ignores critical theory approaches. Teddlie (2010) identified 15 specific criticisms of EER and divided them into three major areas: theoretical, methodological and ideological/political.

The long-term criticism of EER as "atheoretical" has been blunted in the last few years as investigators have consciously employed theories in empirical research or reviews of the literature. One example is the Muijs et al. (2004) employment of three theoretical frameworks (contingency theory, compensation hypothesis and additivity hypothesis) in organizing their

review of the literature on "improving schools in socioeconomically disadvantaged areas". Similarly, Van de Grift and Houtveen (2006) used the same three theoretical frameworks in interpreting the data they gathered in a study of underperformance in Dutch elementary schools.

Thus, it appears that researchers in the areas of EER have taken criticism of their areas being "atheoretical" seriously and have started to address the issue, using existing and new theoretical frameworks. More details about this development are found in the next section of this paper.

There have been a number of methodological advances in EER over the past forty years beginning with criticisms of the methods employed in the Coleman Report and other studies from that era. Some of the earlier advances included the development of social psychological scales to measure school processes, inclusion of more sensitive measures of classroom input, the trend toward more longitudinal studies of school processes, and so forth. Details about more recent state-of-the-art advances in EER methodology (e.g. growth curve modelling) are found in the next section of this paper. Methodological advances in EER have been a continued hallmark of its progress.

Ideological/political criticisms of EER have been more difficult to address. As noted earlier in this section, the internationalisation of the field and the influence of perspectives from other disciplines have led researchers in EER to reconsider some of the implicit assumptions of their philosophical or conceptual orientations. One criticism, that the EER field is philosophically homogeneous (i.e. dominated by post-positivists), has been blunted by continued emphases on qualitatively orientated research-based school improvement studies and by recent EER involving mixed methods. Other criticisms, such as EER having a pervasive and biased effect on educational policy (either liberal or conservative), have tended to fade, as they have been inconsistent across time and within different country contexts.

It would be wrong to be too critical of EER, however. Overall it has achieved much, notably:

- In encouraging a focus on a range of important outcomes and on the potential role of education in improving outcomes for disadvantaged groups of students
- In helping to counter the mistaken belief that schools could do nothing to change the society around them

- In helping to rigorously study "what worked", rather than be prone to follow fads and fashions about this
- In demonstrating to practitioners they had power that could be used for the good over young people
- In creating a valid, although as we noted above, somewhat limited knowledge base which could act as a foundation for training and professional development and which could avoid the need for the reinvention of the wheel by the teaching professions of different countries
- In stimulating evidence-based approaches to school improvement.

For the remainder of this paper we will concentrate – positively – on the future.

Theoretical and methodological advances in EER

Recent years have seen considerable advances in both theoretical formulation and in methodology. Firstly, possessing theories that explain the relationships between variables – like those between schools and students – is essential for any successful field. Theories "organize" findings in ways that help new entrants, they provide clear explanations for people inside and outside a field and provide rationales for any practitioner or policy take-up of findings. But having "theory" that moves beyond associations to "causality" is a problem throughout the social and behavioural sciences, of course, as well as in EER.

However, EER in its early stages only had theoretical perspectives that were the results of the borrowing of theories from other disciplines (like contingency theory, used to discuss contextual variation, or coalition building from political science, used to discuss successful leadership in effective schools), together with some preliminary attempts to "causally order" educational effectiveness factors (Scheerens & Bosker, 1997; Bosker & Scheerens, 1994).

However, explaining in an integrated theoretical fashion the associations between variables has recently been the focus of the dynamic theory of educational effectiveness of Creemers and Kyriakides (2008). Their theory is comprehensive in nature and looks simultaneously at all the different levels of the educational system – the student, the classroom, the school and the context. Crucially, it is dynamic and seeks to place the study of change at its

heart, since its proponents rightly believe that the lack of appropriate models of change has hindered the uptake of EER by practitioners in schools. Longitudinal research is favoured, obviously, as this makes the study of change easier, and each educational factor is argued to possess five dimensions – frequency, focus, stage, quality and differentiation. Factors at different levels are seen as having both direct and indirect effects upon student outcomes. There is also a particular focus upon the classroom, and upon teachers' behaviours there. The theory is being tested in multiple studies (e.g. Kyriakides, 2008; Kyriakides & Creemers, 2008, 2009), with promising results. Other interesting theoretical speculations of a different kind are in Luyten et al. (2005) and Van De Grift (2009).

Second, at the same time as the move towards increasing theoretical sophistication, methodological advances have been taking place over time, of which the main ones have been multilevel modelling, meta-analysis, structural equation modelling, growth curve modelling and mixed method research. A Special Issue of the *SESI* journal has been devoted to this topic (Sammons & Luyten, 2009). Creemers, Kyriakides and Sammons (2010), have also discussed methodological issues and new advances in depth.

- Multi-level Modelling (MLM). As we noted above, there have been a number of important methodological achievements in EER, particularly related to the use of multilevel models and large-scale longitudinal research that recognizes the complexity and hierarchical structure of most educational systems. The advent of accessible software packages such as HLM and MLWin encouraged improvements in the size, scale and statistical approaches used in EER during the late 1980s and 1990s (for example, see work examining both teacher and school effects by Hill and Rowe in 1996 and 1998 that demonstrated not only that teacher effects tend to be larger than school effects but also that, in combination, teacher and school effects could account for a substantial proportion of the variance in student outcomes).

  Improvements in the modelling of measurement error and interaction effects, and cross-classified models that examine multiple institutional membership and regression discontinuity to study the size of the schooling effect and its variation in the absence of longitudinal data, provide examples of recent developments in EER that are proving fruitful.

- Meta-analysis. Further refinements in multilevel approaches include multilevel meta-analysis that has the potential to provide better estimates of the size and variation in educational effectiveness for a range of outcomes, phases of education and contexts (Hox & De Leeuw, 2003). Meta-analysis uses statistical results from a range of studies that address a similar research question and often seeks to establish an average effect size and estimate of the statistical significance of a relationship. In EER this might be the effects attributable to a particular approach to teaching or of a school reform programme. This can be seen as a major advancement in the field. It has promoted the refining of theory and enables researchers to identify generic and more specific factors, the impact of which is dependent on the educational setting in which they are operating (Scheerens & Bosker, 1997).
- Structural Equation Modelling (SEM). We noted earlier that SEM permits the study of indirect and reciprocal effects, and is increasingly popular in social and behavioural science. SEM has also become central to many EER studies in seeking to measure the construct validity of the instruments that have been employed. SEM is often used as an umbrella concept to denote a wide range of models.
- Growth Curve Modelling. Growth curve modelling is becoming more widely applied and represents a further refinement on more traditional multilevel analysis by modelling student growth in academic or in attitudinal or social and behavioural outcomes across more than two time points. Guldemold and Bosker (2009) illustrate the curvilinear nature of growth curves for children's academic outcomes from kindergarten through to Grade 6 and show how growth rates differ for low SES students compared with others. Van de gaer et al. (2009), by contrast, examine non-cognitive outcomes and secondary age students using multivariate latent growth curve approaches to the study of developments in student motivation and academic self-concept. This has been particularly important in identifying non-linear processes over time.
- Mixed Methods Research. Teddlie and Sammons (2010) argue that the flexibility of mixed methods research in simultaneously addressing multiple and diverse questions through integrated qualitative (QUAL) and quantitative (QUAN) techniques is one of its attractions. Mixed method research data adds "extra value" to EER research that seeks to

better describe, predict and understand the variation in, and contributors to, differences in educational effectiveness. The integration and synthesis of QUAL and QUAN evidence can foster mutual illumination and so has the potential to enable the development of EER theories and is also necessary to inform and support closer links with applied research and evaluations that can promote effective school improvement initiatives and teacher development programmes. For examples of such mixed methods research, see Sammons (2010b).

We now move on to consider EER's findings in terms of what can be called the scientific properties of schools, and, after this, its descriptions of effective school and classroom processes.

Scientific properties of school effects (including consistency, size, stability, differential effects, and the different outcomes of education)

School effects have been generally measured through the application of cross-sectional multilevel models, correcting for student background characteristics and usually also prior achievement. These classic value-added models yield an estimate of the percentage of the total variation in students' scores that is situated at the school level, which gives an idea of the relative importance of schools to the outcomes of individual students.

Student achievement is still the predominant effectiveness criterion in EER. However, in recent years, researchers have been investigating a broader range of outcomes of education. These include non-cognitive outcomes such as student well-being (Konu, Lintonen & Autio, 2002; Van Landeghem, Van Damme, Opdenakker, De Fraine, & Onghena, 2002) and achievement motivation (Van de gaer, De Fraine, Van Damme, De Munter, & Onghena, 2009; Van der Werf, Opdenakker & Kuyper, 2008).

Long-term effects across phases of schooling are increasingly studied in EER. Here, the effect of the school is studied on a student outcome that is measured after the students have left the school. This is related to the issues of predictive validity and the generalizability of school effects across time (Teddlie & Reynolds, 2000). It was found that the primary school can have a long-lasting, but small, effect on student achievement in secondary education (Goldstein & Thomas, 1996; Sammons, Nuttall, Cuttance, & Thomas, 1995; Thomas, 2001), while Leickie

(2009) has used cross-classified models to study primary, secondary and neighbourhood effects simultaneously. Secondary schools differ in the extent to which their students obtain a degree in higher education (Pustjens, Van de gaer, Van Damme, & Onghena, 2004).

The two general dimensions of school effectiveness, quality and equity, are still at the heart of the research domain. The international comparative studies tend to report the effectiveness of educational systems regarding these two dimensions. School quality is seen as the degree to which a school scores better than other schools, corrected for student intake characteristics. The equity dimension refers to the compensatory power of schools, indicating that some schools are better at compensating for input characteristics (such as SES, gender and ethnicity) than others. Thus, the quality dimension refers to between-school differences, while the equity dimension refers to within-school gaps (Strand, 2010), although of course gaps between schools which can attract different kinds of students have equity connotations, too.

Recent studies in educational effectiveness stress the importance of (1) studying growth in student achievement and (2) studying absolute school effects. Growth in student achievement has been gaining acceptance as the essential criterion for assessing school effectiveness because learning involves changing (Teddlie & Reynolds, 2000). In this way, the terms "progress", "growth" and "learning gains" are often regarded as synonyms. Longitudinal data (at least two measurement occasions) are a necessary condition for measuring the learning gains of students and schools (Singer & Willett, 2003). Many educational effectiveness researchers are convinced that learning gains should be measured using growth curve models over several time points. These models are becoming increasingly popular (e.g. De Fraine, Van Damme & Onghena, 2007; Palardy, 2008; Van de gaer et al., 2009; Van der Werf, Opdenakker & Kuyper, 2008) and show that change in outcomes is often non-linear, with the possibility for social and affective outcomes of relative decline at certain time points.

There is also a trend to investigate absolute school effects instead of only relative school effects. EER has traditionally focused on variation in learning outcomes between schools (relative effect or comparative effect). The value-added measures from these studies express to what extent learning outcomes deviate from the performance level expected (Van de Grift, 2009). However, there is a growing interest in assessing the impact of education on the development of children in its own right. The regression-discontinuity approach allows for the

assessment of the absolute effect of schooling (Luyten, 2006; Luyten, Peschar & Coe, 2008; Luyten, Tymms, & Jones, 2009). In addition, it can study relative variation between schools in this absolute effect (in other words, what is the absolute effect of an extra year of schooling and how does this vary between schools).

In terms of the size of school effects, in the field of EER nowadays there is no doubt that schools can make a difference. Even rather small school effects are considered important, because they might be cumulative and they refer to a large number of students. Moreover, when measurement error is controlled estimates of school effects tend to be larger, suggesting that historic EER may have underestimated the size of the overall school effects.

Recent studies differ in the size of the school effect they report. Generally, this size is estimated as the percentage of the differences in the student outcome that can be attributed to the school (the intra-class correlation). The size of the school effect is affected by the outcome under study. Schools tend to have a larger effect on student achievement than on non-cognitive outcomes (Martin et al., 2011; Opdenakker & Van Damme, 2000; Thomas, 2001). Two main hypotheses have been generated to explain the relative small effects on non-cognitive outcomes. First, these non-traditional outcomes may be given less emphasis in the curriculum. Second, the measurement of these non-cognitive outcomes is less precise than the measurement of achievement. Martin et al. (2011) suggest that, to improve our understanding, research needs to focus more on teaching and schools as experienced by individual students to improve our understanding. Also, there is variation in the size of school effects for different academic outcomes. School effects tend to be larger for subjects such as mathematics and science, which are typically learned largely at school, as compared with school effects for language (Thomas, Sammons, Mortimore & Smees, 1997a). Also, longitudinal studies examining student growth are more likely to demonstrate school effects of greater magnitude (Teddlie & Reynolds, 2000; Van de gaer et al., 2009). Additionally, it seems school effects are larger in subject areas where the school forms the main avenue of exposure – as with mathematics, for example, where exposure is limited in the family and the community (see Mortimore et al., 1988; Teddlie & Reynolds, 2000).

Moving on to look at consistency across outcomes, several criteria can be used to investigate differences between schools. Consistency refers to the correlation between school

effects on these various outcomes. Consistency helps to answer whether school effects are an overall or a specific phenomenon.

In general, school effects seem to show some degree of coherence. Small to moderate correlations are found, so we can conclude that effective schools in one area tend to be more effective in other areas. However, some studies (especially in secondary education) found that in spite of this general trend some schools can be effective for one subject but ineffective for another (Thomas, 2001; Thomas, Sammons, Mortimore & Smees, 1997b). Inconsistency in primary school effects could indicate that the teacher is better at teaching in one subject than another (Mortimore et al., 1988). Inconsistency in secondary schools could indicate differences in teacher effectiveness or departmental effectiveness. Van de gaer et al. (2009) investigated consistency between two non-cognitive outcomes and found that the consistency between school effects for motivation and academic self-concept resulted largely from intake differences between schools.

Researchers in EER are generally interested in the consistency between cognitive and non-cognitive outcomes. A negative correlation might indicate a trade-off (competition between cognitive and non-cognitive goals), whereas a positive correlation would support the occurrence of complementarity (Van der Wal & Waslander, 2007). However, the results remain inconclusive (*ibid*), suggesting that a school's effectiveness in one domain is largely dependent on effectiveness in the other.

To summarize, since there is not perfect consistency, the evaluation of a school's effectiveness should be based on more than one effectiveness criterion, because a single criterion can only highlight one particular aspect of schooling (Sammons, 1996; Teddlie & Reynolds, 2000).

Stability over time is displayed by the correlation between school effects at different moments in time (Doolaard, 2002; Thomas et al., 1997b). Most studies indicate rather stable school effects. Thomas et al. (1997b) found correlations between school effects in three successive years between 0.82 and 0.85. There was high stability over three years for general secondary school academic results, but fluctuations for specific subjects that are interpreted as departmental effects (Sammons, Thomas & Mortimore, 1997; Thomas et al., 1997b).

Absolute stability over several years is impossible, because schools are changing organisations. There can be changes in schools' policies, a new school leader, changes in staff and student body, etc. Instability can thus indicate changes or improvements in schools. Many education systems have strong pressures to improve school results, especially for those seen as weak or poorly performing, and this is likely to reduce stability in schools' results over time, especially at the lower achievement range.

We conclude that school effects are typically relatively stable over short periods of time. The correlations between overall effects across several years are high but far from perfect, which makes it difficult to predict results (Mangan, Pugh & Gray, 2005). Thus, judging a school's effectiveness should be based on data from several years (Sammons, Thomas & Mortimore, 1997; Thomas, 2001).

When considering differential effects, a schools' overall effect refers to the impact of this school for an 'average' student. But the schools' effect can vary across students. Schools may be more effective for one group of students than for another group. Several studies have investigated differential school effects, addressing effects for different student groups in terms of prior achievement, gender, ethnicity and socioeconomic status.

School effects can be a function of student ability level or prior achievement (Sammons et al., 1993; Strand, 2010; Thomas, 2001; Thomas et al., 1997a), which means that the relationship between prior achievement and later achievement is smaller in some schools than in others.

With regard to gender, some studies found that schools are equally effective for boys and girls (Sammons et al., 1993; Thomas, 2001), whereas others found that the gender gap differs from school to school (Strand, 2010). Also, for ethnicity some studies found that a school can be more effective for one ethnic group than for another group, while other studies found no evidence of differential effectiveness regarding ethnic background of students. For example, Strand (2010) found that schools that were strong in facilitating the progress of White British pupils were equally strong in facilitating the progress of Black Anglo-Caribbean pupils. And several studies found only modest differential effects with regard to student socioeconomic status (Strand, 2010; Thomas, 2001; Thomas et al., 1997a).

To conclude, there is some evidence for differential effectiveness, but there is still much research needed on these within-school differences, together with the other scientific properties of school effectiveness that we have reviewed here.

The next section considers the processes of educational effectiveness.

The processes of educational effectiveness

Given the historic origins of the field in improving the life chances of children, it is not surprising that a considerable volume of effort has gone into describing the characteristics of the schools and classrooms that "add value" to the outcomes of students. There has been the initial wave of small-scale case studies from the US (e.g. Weber, 1971), the mixed method longitudinal studies from London (Mortimore et al., 1988; Rutter et al., 1979; Sammons, Thomas & Mortimore, 1997) into both primary and secondary schools, the Louisiana School Effectiveness Studies (Teddlie & Stringfield, 1993), and research in Scotland (MacBeath & Mortimore, 2001). A wide range of other studies are mentioned in the reviews at the beginning of the History section of this paper.

What is interesting is the extent to which the original five "correlates" of effectiveness at school level in the foundational Edmonds (1979) study appear to have survived over time, in multiple countries and multiple settings within countries, as valid explanations, although research now is much more complex and multi-faceted in all these areas.

Edmonds (1979) noted five "correlates":

- Strong Principal leadership
- An emphasis upon basic skill acquisition
- An orderly climate that facilitated learning
- High expectations of what students would achieve
- Frequent monitoring of the progress of students.

The quite comprehensive Teddlie and Reynolds (2000) review, based upon analysis of literally hundreds of "process-based" studies, identified nine similar global factors:

(1) Effective leadership that was:

Firm

Involving

Instrumentally orientated

Involving monitoring

Involved staff replacement

(2) A focus upon learning that involved:

Focusing on academic outcomes

Maximized learning time

(3) A positive school culture that involved:

Shared vision

An orderly climate

Positive reinforcement

(4) High expectations of students and staff

(5) Monitoring progress at school, classroom and student level

(6) Involving parents through:

Buffering negative influences

Promoting positive interactions

(7) Generating effective teaching through:

Maximizing learning time

Grouping strategies

Benchmarking against best practice

Adapting practice to student needs

(8) Professional development of staff that was:

Site located

Integrated with school initiatives

9. Involving students in the educational process through:

Responsibilities

Rights.

Interestingly, the recent review of even more contemporary literature by Marzano (2003, 2007), undertaken from the different paradigm of educational administration, shows remarkably similar findings. His school level factors were:

- Professional behaviours involving leadership and cooperation
- Guaranteed curriculum offerings that involved high time available for learning and opportunity to learn within that time
- A safe and orderly classroom climate
- Challenge, involving pressure to achieve and frequent monitoring
- Parental and community involvement
- An effective "classroom" or "instructional" level.

Over time, however, there has been increasing interest in more complex formulations of the "correlates" that reflect the possible effects of variation in the contexts in which schools are situated – the so-called "context specific" models of effectiveness. Early work in this area tended to look at the school composition effect in terms of how the composition of the entire body of students in a school had effects upon outcomes *in addition to* the effects of the students as individuals (Murnane, 1981; Willms, 1986).

Later work, particularly in the US (Hallinger & Murphy, 1986; Teddlie & Stringfield, 1993) focused upon the differences in the processes of effective schools that occurred in different socioeconomic status areas, with the particularly interesting finding that the schools in low SES areas actively pursued policies to dis-involve their parents from their children's education!

More recently, the distinct characteristics of what is needed to improve in very socially challenged communities has been a focus in the UK (Harris et al., 2006; Muijs et al., 2004; Reynolds et al., 2001) with hints that, whilst many of the effective practices needed are in line with the "global" correlates outlined earlier, three specific additional areas seem particularly important:

- Making the school a learning community that can in a lateral fashion identify and transmit "good practice"
- Support from outside the school in key areas
- Additional resources to potentiate innovation and change.

Contextual effects of course need not be restricted to those of socioeconomic backgrounds only. They could be associated with:

- Urban/rural differences
- Differences in school improvement trajectories
- Differences in school initial effectiveness level
- Differences in school "types" (e.g. religiousity factors).

In recent years, the study of effective processes has been given an international "dimension" by the increased focus upon country differences emanating from the PISA studies, particularly. There have been productive reviews of the literature from multiple countries that show interesting similarities – and differences – in "what works" (Townsend, 2007). There have been ambitious attempts to look at the student experience in selected countries to see whether the same factors explain variance as in our above reviews (e.g. Reynolds et al., 2002). Interestingly, in this latter study the usual teacher or instructional level factors did "travel" internationally, both conceptually and operationally, but the school level factors only "travelled" conceptually, meaning for example that whilst the leadership of the Principal "mattered" in different contexts, the precise characteristics of that effective leadership (directive in Oriental cultures, more "lateral/vertical" in Anglo Saxon ones) is context-dependent.

Whilst another "state-of-the-art" review in this issue is concentrating upon teacher effectiveness and development, it is important to note that the study of effective classroom practices has also been central to the search for "what works" in EER, with many studies simultaneously studying the school and the classroom "levels".

One of the most important historical studies in this area was that of Mortimore et al. (1988), that collected an immensely rich database of information on children, their classrooms, their primary schools and their individual background characteristics, using a cohort of children followed through the four years of British junior school education. Generally, Mortimore and colleagues found that teachers were in those days spending much more time communicating with individual children than they were doing whole-class teaching or facilitating collaborative group work. At classroom level, the effective teacher characteristics were:

- Teachers having responsibility for ordering activities during the day for students (i.e. structured teaching)

- Students having some responsibility for their work and independence within working sessions
- Teachers covering only one curriculum area at a time
- High levels of interaction with the whole class
- Teachers providing ample, challenging work
- High levels of student involvement in tasks
- A positive atmosphere in the classroom
- Teachers showing high levels of praise and encouragement.

Mortimore and his colleagues also showed that teachers who spent a lot of time with individual students were using most of the time in routine (i.e. non-work) matters and there was less use of higher-order questioning, while teachers who used class discussions as a teaching strategy tended to make rather more use of higher-order communication.

Mortimore concluded that the classroom factors contributing to effective student outcomes were structured sessions, intellectually challenging teaching, a work-orientated environment, communication between teachers and students, and a limited focus within the sessions. It should also be acknowledged that a larger field of TE was evolving in the US where, in the 1970s and 1980s, there was a great deal of research. Three useful reviews are in Brophy and Good (1986), Rosenshine and Stevens (1985) and Brophy (1988). Brophy observed that a great deal of the "teacher effectiveness" data across many large-scale studies should more accurately be described as "teacher ineffectiveness research", because there were clearly many ways to be "effective," but a specific series of correlates of "teacher ineffectiveness.".

In the 2000s, major teacher effectiveness research studies were built on the evaluation of a specialist mathematics intervention, the Mathematics Enhancement Programme (see Muijs & Reynolds, 2000, 2003; Reynolds & Muijs, 1999). This work was based upon testing the entire student population of 35 British primary schools on mathematics and using a standardized observation instrument that measured teachers' behaviours, students' behaviours and lesson structure.

Nearly sixty different behaviours by teachers in classrooms, concerning their classroom management, management of behaviour, the quality of their direct instruction, the interactivity of their teaching, the attention given to individual review and practice, the variation in the teaching

methods, the use of "connectionist" teaching methods and the classroom climate created in lessons, were related to improvement in performance over the year. Rather than any single teacher behaviour being strongly related to achievement, lots of small correlations were found, indicating that effective teaching is not being able to do a small number of "big" things right but is rather doing a large number of "little" things well. In our more advanced analysis, the factor of "effective teaching" was the most important determinant of how children did, after the influence of their own achievement level, reinforcing what we noted earlier about how important teaching is.

In these studies, though, teaching behaviours were not the only factor of importance to student achievement. Teachers' beliefs about teaching, their subject knowledge and their self-efficacy (or their views about their own power as teachers) all also mattered, in the way that they encouraged teachers to adopt the more effective teaching methods that have powerful effects in improving students' achievements.

The development of observation instruments based on the TER knowledge bases has facilitated the study of teaching in different contexts and how far the characteristics of effective practice travel. For example, Teddlie et al. (2006) developed the International Schedule for Teacher Observation and Feedback (ISTOF).

Further recent research (Day et al., 2008) involved mixed methods work in English primary and secondary schools that described, analysed and sought to explain the variation in primary and secondary teachers' classroom practices using two different observational instruments (the ISTOF and the Quality of Teaching schedule developed by Van de Grift et al., 2007) and pupil and teacher perceptions, focusing on English and mathematics teaching. It also explored typical and more effective classroom practice of teachers across different school contexts, career phases and ages.

Results reveal that the sample of effective teachers scored highly in terms of the following factors, based on observation of the quality of their teaching:

- Supportive lesson climate
- Proactive classroom management
- Clarity of objectives and well-organized lesson structure
- Environmental and teacher support

- Engaging students with assignments and activities
- Positive behaviour management
- Purposive learning
- High-quality questioning and feedback for students.

These features can be seen as necessary characteristics of effective teaching across different sectors, subjects and contexts.

In the last few years there has been considerable growth in research, in reviews of research and in models about effective practices in this area (see, for example, Campbell et al., 2004; Muijs et al., 2004). As an example, the Teddlie and Reynolds (2000) review of school effectiveness research had no chapter on teacher effectiveness, but the upcoming replacement volume will have two.

There have been also important recent studies in the field. The Effective Pre-School and Primary Education 3-11 Project (EPPE) shows that the influence of overall teaching quality upon mathematics and reading outcomes is stronger than the net influence of some background factors such as gender and family disadvantage. In detail, this covers and relates to the richness of instructional methods, a positive classroom climate, productive use of instructional time, the use of feedback to students, teacher sensitivity to their students and a lack of teacher detachment. An organized classroom where there is a calm, orderly climate is also important (see Sammons et al., 2008, for full information).

The Variation in Teachers Work, Lives and Effectiveness Project (VITAE) has also found an association between teachers' commitment to their jobs, their resilience in resisting stressors and the improvement of their students on the English national tests at ages 7, 11 and 14 (Day et al., 2006, 2007).

One of the key research needs – as with the school level – is to explore the extent to which effective teaching is a set of "generic" behaviours and attributes that "work" across all kinds of educational contexts, and the extent to which a more differential model, in which effective teachers have to do different things in different contexts, may be necessary (Campbell et al., 2004). In this latter perspective, teaching effectiveness is seen as a multi-dimensional construct and a variable factor rather than a universal "given". Interesting attempts have been

made to study the extent that different aspects of effective teaching apply in different contexts using the ISTOF instrument (Teddlie et al., 2006).

Generally, more and more evidence is accumulating of the need for differentiated explanations of good teaching, as shown in the following areas:

- Differences between subjects – the major studies on teacher effectiveness commissioned by the British Teacher Training Agency (Askew et al., 1997; Wray and Medwell, 2001) showed that subject knowledge mattered less in teaching numeracy than literacy. Classroom grouping of tasks by ability was more prevalent in literacy teaching that was effective.
- Differences between students of varying SES – low SES students generally need teacher behaviours that generate a warm and supportive climate by letting children know help is available, elicit a response (any response) before moving on to the next bit of material, show how bits fit together before moving on, emphasize knowledge and applications before abstraction (putting the concrete first), give immediate help (through use of peers perhaps) and which generate strong structure, ground-flow and well-planned transitions (from Borich, 1996).
- The effectiveness level of the school – with more effective institutions needing a more "collegial" approach to performance enhancement by comparison with the teachers in less effective schools, who require more "assertive" kinds of leadership (Hopkins & Reynolds, 2001).
- The trajectory that a school is on – with schools already on a steep curve of improvement needing less provision of basic foundations than those yet to start that journey (Hopkins & Reynolds, 2001).

New foci of interest have been appearing in recent years. At the level of the District/Local Authority there has been increased attention to school boards and their effects (Alsbury, 2008; Stringfield & Yakimowski, 2005; Stringfield, 2008; Land, 2002), and similarly Glass and Franeschini (2007) generated a study and review of research on the American school superintendency. Marzano and Waters (2009) recently reviewed all the District/Local Authority effects studies in order to make recommendations integrating schools and district local authorities to generate High Reliability Organisations (HROs), although Tymms et al. (2008)

show minimal District effects. Shelton (2010), by contrast, shows substantial District differences in the shape of student achievement over a 10 year period, related to school board/superintendant relationships. This is clearly a field ripe for further research.

Educational ineffectiveness

As long as the fundamental antithesis forwarded by Coleman et al. (1966) that schools had little to no differential effects on students' levels of achievement held sway, there was little reason for an effectiveness field to evolve. From its roots, the various "effectiveness" fields have focused the great majority of their scientific endeavour on identifying characteristics of relatively more effective teachers, schools, districts and countries. Being a young and perhaps somewhat insecure science, the effectiveness field has focused most of its energies – and writings – on the positive side of its discoveries.

Two important concepts have tended to be overlooked in this desire to please. First, the identification of relatively positive characteristics implies the presence of negative ones, but the negatives are not necessarily the polar opposites of positives. Second, in planning to respond to any problem it is as important to understand the specifics of the problem as the range of solutions. For example, in the evolution of the medical sciences, the field had to develop a deep understanding of the differentiating nature and causes of diseases before it could develop equally differential cures (Thomas, 1979).

In a groundbreaking but under-discussed paper entitled "Research on teacher effects: Uses and abuses", Brophy (1988) observed that most of what was known about the "teacher effectiveness" field was drawn from process–product studies that more clearly described (but did not adequately discuss) what was known about teacher *in*effectiveness. Similarly, Edmonds' famous "five factors" are more accurately understood and sound less like truisms if they are used to implicitly describe characteristics of *in*effective schools. Ineffective schools have weak Principal leadership, a lack of emphasis on the acquisition of basic skills, a disorderly climate, low or uneven expectations, and inconsistent or no monitoring of student progress.

In any science, it is important – if not always popular – to make the implicit explicit. A few studies have been conducted that have formally addressed the topics of school

ineffectiveness, school decline and processes that kill school improvement efforts. Each of these areas is briefly noted below, and each area merits additional research.

Reviewing research on the topic of school ineffectiveness, Stringfield (1998) described ineffectiveness in schools as being observable at school, teacher and student levels. Students in ineffective schools were characterized as spending considerably less time per hour and day engaged in academic learning. Further, the time they did spend was more likely to be characterized as "intellectual anarchy". Tasks were put in front of students with little explanation of why the students were being asked to complete the tasks or how the tasks related to the larger processes of understanding coherent fields of knowledge.

At the classroom level, ineffective schools were characterized by a leisurely pace, minimal moderate-to-long term planning, low or uneven rates of interactive teaching, a preponderance of "ditto sheets" and other relatively unengaging tasks, a failure to cover all of the year's assigned content, and teachers teaching in isolation from one another.

Finally, at the school level, ineffective schools held most or all of seven characteristics: lack of academic focus, regular disruptions to and wasting of academic time, resources working at crossed purposes, Principals who were not conversant with the specifics of their schools' curricula and were relatively passive in the key processes of recruiting new teachers and providing accurate feedback to current teachers, the inefficient use of school libraries/media centres, and a lack of public celebration of student successes.

Additionally, Hochbein (2011) noted that Brookover and Lezotte (1979) conducted the first rigorous study that included schools in decline. While their sample included six improving and only two declining schools, Brookover and Lezotte noted that in declining schools:

> The most pervasive finding was the one concerning teachers' and Principals' attitudes toward student achievement. The staff in the declining schools had low opinions of their students' abilities, while staff in the improving schools had high opinions of student abilities (Brookover & Lezotte, 1979, abstract).

In an article largely focused on school improvement, Stringfield and Teddlie (1988) provided a somewhat detailed roadmap for the creation of ineffective schools. They described a process beginning with the introduction of a new Principal lacking academic focus, declining attention to student learning and coherence among school processes, the choices made by

competent, experienced teachers to leave the increasingly dysfunctional environment, and a haphazard attitude toward hiring new professional staff.

Both Grant (1988) and Duke (1995) provided detailed case studies of American high schools that had fallen from excellence to sub-mediocrity. Both stories featured substantial declines in Principals' leadership, declining academic standards among faculty, declining school climates, and increases in the percentages of students who were more at risk when entering the schools. In England the Forging Links research on academic effectiveness of secondary schools studied ineffective and more effective schools and subject departments (Sammons, Thomas & Mortimore, 2007), plus schools that might appear average but in which both effective and ineffective subject departments co-existed. The results drew attention to the importance of Principal and head of department leadership, academic emphasis, quality of teaching and behavioural climate.

In terms of the processes which can kill school improvement efforts, Charles M. Payne's (2008) *So much reform, so little change* is a sobering reminder that most efforts to improve schools do not produce desired results in the first place, and/or are not sustained. In the educational effectiveness field, the need to study and understand why literally hundreds of well-intended and often seemingly well-designed school reform efforts have fallen into the dustbin of history is long overdue. From the "Eight year study" of the 1930s (Aikin, 1942) through to Tyack and Cuban's (1995) analysis of a century of reform efforts, to Supovitz and Weinbaum's (2008) *The Implementation gap,* a discouraging summary could be stated as, "There is a lot more said than done, and a lot more begun than sustained". A prudent would-be reformer would ask why this is so before initializing another reform effort.

After participating in a three-year study of diverse externally developed efforts to improve 25 high poverty schools, Nesselrodt, Schaffer and Stringfield (1997) concluded that the 25 participating schools had experienced 10 potentially overlapping and over-determining potential causes of reform failure. Listed in order of frequency, the potential change killers were:

- inability to sustain funding (in 8 out of 25 schools)
- inability to sustain teacher commitment (8)
- unresolved issues with curriculum alignment (6)
- challenges in recruiting and developing teachers and other key staff (5)

- racial disharmony on the staff (3)
- parent or community perceptions that the school faced too many deep problems (3)
- management, communication and scheduling problems (3)
- the schools' physical facilities presented challenges to offering the reform that the school and/or district did not address (2)
- and other contextual or political problems such as the arrival of a new district superintendant who saw no value in continuing the particular reform (2).

(summarized in Stringfield, 1998).

After five years of leading a team studying six different, externally designed reforms in one large system, Datnow (2005) concluded that reforms that last actively assist school leaders in adapting to ever-changing district and state/national policy demands and make few long-term financial demands on the school and system. She observed that policy people need to be more aware of their impacts, intended and unintended, on schools and their various reform efforts. Datnow concluded with several pertinent observations. First, schools not firmly committed to seeing specific reforms through for the long haul probably should not begin the reforms at all. If a school's leaders believe a reform to be producing desired results, they would be well advised to keep their central administration appraised of the reform and what it takes to sustain it. "Finally, it would be wise to choose a reform that can help the school improve on state and district measures of accountability" (Datnow, 2005, p. 148).

Having reviewed the scientific properties of EER, and the processes associated with effectiveness and ineffectiveness, we move to look at the extent to which EER has been successful in communicating these insights to broader constituencies.

The historical links between EER and policy/practice

As a discipline that has generated a valid body of knowledge about "what works" at school, classroom and increasingly at country and educational system level, one might have expected a considerable take-up of EER insights internationally. That has not happened, except in a few individual countries. SER "boomed and busted" in the US in the 1980s, largely due to its adherence to a very simple model of effective school practice independent of context (Edmonds,

1979). It had influence through the national school effectiveness programme of the Australian Federal Government (in which the knowledge base was taken into all State systems), but was replaced later by emphases from the systemic change literature and the sociology of education.

In the UK, the "New Labour" government in the late 1990s used school effectiveness and teacher effectiveness research as the foundations of its National Strategies and some of its policies to improve weaker schools (see Reynolds, 2010a; Sammons, 2008), but the association with "prescription" meant that the influence was relatively short-lived. However, the English inspection agency OFSTED utilized EER in its Inspection Framework (see Sammons, Hillman & Mortimore, 1995), and the documentation upon school improvement that each English school fills in for school self-evaluation drew on the evidence upon within school variation (Reynolds, 2007).

There have been links between EER and the Dutch National Inspection Framework, and some evidence of governmental interest in Germany, Finland and some Latin American societies, but it is only in Wales (with its educational outcomes apparently falling rapidly down the PISA "league table") that there appears to be systematic use of EER findings currently (Reynolds, 2008). Townsend (2007) documents interest – but not mainstreaming – in many societies. EER is bolt-on, not bloodstream, for the policy making communities.

In addition, value-added measures of school effectiveness based upon multilevel analysis using EER approaches were introduced in 2002 and contextual value-added measures after 2005 to supplement raw league tables in England. However, after a change of government they were abolished in 2010, because they recognized the link between school results and student intake characteristics such as ethnicity and socioeconomic status, a topic regarded as politically unacceptable.

Reasons for this lack of policy/practice reach may be as follows:

- The quantitative statistical knowledge required to fully access some of the knowledge base
- The considerable volume of criticisms of EER which has emerged, given that politicians may tend to gravitate to the popular

- The reluctance to embrace a discipline that now repeatedly argues for the primacy of teacher effects rather than school effects, given that policy makers seem happier operating at school rather than classroom level
- The reluctance to embrace a discipline which increasingly argues for "contextually specific" policies, given historic policy maker commitment to "steam press", universal or "one size fits all" ones
- The continued tendency for EER to generate findings that are, in the UK at least, "inconvenient", from the Gray et al. (1999) finding of a negative association between school development planning and improvement in student achievement over time, to the relatively small effects associated with factors such as school type, governance and practices such as setting and streaming.

At the level of practice, it would also be difficult to find evidence of substantial take-up of the insights of EER at practitioner level in many countries although, of course, any practitioners doing award-bearing courses in areas such as educational administration, educational leadership and educational improvement would have been exposed to it. Training teachers may gain some familiarity with material on effective educational practices, but generally the volume of practitioner take-up does not match the volume of useful knowledge available. Again, EER is bolt-on, not bloodstream, for many in practitioner communities.

The reasons for this are likely to reside in:

- The historic, mostly quantitative, orientation of EER which makes access difficult
- The historic neglect of practitioners' core concerns which are likely to be their teaching methods
- The limited attention given to theory that would help to explain patterns of results of individual studies
- The static nature of the historic EER analyses, in a world that is dynamic and fast moving
- The need to prepare practice constituencies for complex educational changes inherent in the EER knowledge bases
- The slowly melting divide between educational effectiveness and educational improvement research fields (see Hopkins & Reynolds, 2001).

Greater understanding of how to improve policy and practice "reach" in the fields of EER, using insights from the fields of professional development, educational change, school improvement and related fields, may help us change this situation.

EER: future research needs and possible policy/practice links

Future Research Needs

Future directions for research probably grow naturally from the findings and emphases of EER over time. They involve:

- Further concentration upon teaching and teachers, moving beyond the historic focus upon only their behaviours to focii such as their attitudes and values, in which may lie some of the "levers" for changing their practices and behaviours.
- The integration of leadership, its characteristics and its future possible changes fully into the field, since it has been seen as a stand-alone issue, and there needs to be "more studies where leadership is integrated within a model of school effectiveness which is theorized and takes into account the ways in which leadership interacts with other key school factors." (Day et al., 2011; Sammons et al., 2011).
- More longitudinal studies that study the same students and teachers over time, that permit the study of the "naturally occurring experiments" that comprise the "day to day" and "year to year" life in educational systems, and that can detail the processes creating stability and change in schools.
- Making sampling *across* socioeconomic contexts, kinds of school governance, school types and school Districts axiomatic, rather than attempting to control-out such variation in statistical analyses. This would facilitate the much needed production of "contextually specific" accounts of schools.
- More international comparative work, which will benefit the field since the range of "school factors" and teacher behaviours is likely to be much greater in such work than within-the-one-country work that is still the foundation of the field. This expansion of variance in processes at the school level may indeed expand the variance explained at

school level above the present rather low level. It may also show interesting educational factors for experimentation within different societies.

- More work into the links between the school and the classroom level, where much ignorance still reigns. It is clear that there is variation within schools according to the background of the students and indeed for all students attending different subject departments/faculties within schools (at secondary level, obviously, where teaching is organized by different groups of teachers). This variation is itself variable in different schools, but the research enterprise has continued to adopt a "whole school" perspective, which fails to look at the variable processes actually experienced by different pupils of different background characteristics and in different subjects. Students do not experience a "whole school" – they experience different niches within it, yet in virtually all existing EER their schools are seen as a common factor.

  There have been some limited attempts to handle these issues, but none have been entirely satisfactory. In some of the American school effectiveness studies (e.g. Teddlie & Stringfield, 1993) there have been attempts to study the "range" or "dispersal" of teachers, and in the UK some acknowledgement of differences between departments (e.g. Sammons et al., 1997), yet this work focuses on the effective practices of the departments more than the school level factors associated with them. Range is utilized, but not the school-level factors that are associated with its scope. The theoretical modelling of Creemers and Kyriakides (2008) also studies within-school variation in terms of consistency, constancy and cohesion, but is less useful in terms of the factors at school level that may produce the range itself.

  The cost of the absence of fine-grained understanding of the experience of different subjects, different student groups and different student groups within different subjects has limited our understanding of schools. It has also limited the extent to which EER is relevant for practitioners who work in the niches in "whole schools", and impoverished school improvement even more, leading to a fondness for whole school solutions and/or school to school transfer in which policies are thrown at whole institutions in the hope that they reach all the (highly variable) internal school niches. Further speculations on these themes are in Reynolds (2007, 2010a, 2010b). The use of

more specific measures of the educational environments inhabited by students could be a further step along the road towards "student specific" school factors, in which students *as individuals* accrue educational experiences that are measured and tagged to them individually, permitting a much fairer test of the power of the "educational" factors against those connected with the student and his or her social background. This would fit with the increasing tendency to listen to student "voice".

- More work into IT, a major component of school/classroom instructional methods but one that has not generated comparable research effort. The "bolt-on" nature of IT and its lack of close relationship to pedagogy in many societies may explain this, as may the disappointment at its impact, leading to researcher's unwillingness to study the educational equivalent of a "train wreck" because of likely negative (and therefore difficult to publish) findings. The difficulty of measuring IT use in any other than the basic "quantity/quality of kit" utilized is probably responsible for this.
- The adoption of further outcome measures in addition to those upon academic achievement, a cry that has been routinely made now for over a decade (Teddlie & Reynolds, 2000). These could cover:
  - educational "academic" outcomes such as attendance and behaviour
  - more social and affective outcomes.
- The utilization of an "efficiency" as well as an "effectiveness" orientation, which will have the effects of multiplying considerably the range of possible relationships that can be found in EER studies. If, for example, financial "cost" were factored in as an input – and it could be done easily at a whole-school level in terms of educational resources consumed in terms of staff salaries, books and equipment, and so on – then all the multiple relationships that can be seen in studies between educational processes and value-added achievement outcomes would be added to immediately, with the addition of the "cost" factor. If the "cost" factor could be calculated at class level, and particularly if the calculation could be made at individual student level (difficult, but not impossible) then even more interesting analyses could be done.

  Efficiency measures do not, of course, have to be solely concerned with "cost" – they could involve "time", say – another scarce resource. If time were measured and used in

the customary value-added fashion to generate time-adjusted effectiveness scores that reflected the volume of time it took to generate different levels of educational achievement, it might paint a different picture of the reasons behind supposedly high-achieving school systems, or highly successful individual schools.

- More studies of the long-term ineffectiveness of schooling are required to understand their continued dysfunctionality (Teddlie & Reynolds, 2000) and how that cycle might be broken. Intensive longitudinal case studies of samples of these low performing schools might help us to better understand the complex relationship between ineffectiveness and effectiveness. One research question might concern the relationship patterns among teachers at less effective as opposed to more effective schools, which could be examined using sociograms and other measurement techniques. While Luyten, Visscher and Witziers (2005) initially called the suggestion to focus on dysfunctional schools a radical recommendation, they later supported the call "to pay more attention to clearly ineffective schools as a starting point for expanding the school improvement knowledge base" (p. 267).
- Spanning studies of the effects of socioeconomic status, school effectiveness and school improvement, there is a great need for case studies and pro-active change studies of efforts to improve chronically low-performing schools. Herman et al. (2008) conducted a wide-ranging search for evidence of successful efforts at "turning around" schools in which the great majority of students had been performing at very low levels for several years. Tellingly, they could only identify a scattering of case studies and no focused, pro-active, multi-year research. Edmonds (1979) famously observed that we could turn around such schools "whenever and wherever we choose." Four decades later, systemic data to substantiate that claim is in strikingly short supply, and is needed for both scientific and ethical reasons.

Possible policy/practice links

It may be that a number of factors make the situation for EER very promising in the 2010s at the level of links with the policy making communities:

- The problematic economic situation internationally means that public expenditure is under considerable pressure, resulting in an increasing emphasis upon educational *quality* rather than the *quantity* of resources;
- The arrival of societies such as China, India, Brazil and Russia at precisely the time to embrace EER, since there is much dissatisfaction within these societies with the results of their educational systems, and the clear recognition that their industrialization generated by rapid urbanization and the application of limited capital needs to be replaced by more technologically advanced, educationally more intensive production, necessitating improved outcomes from their school systems;
- The recognition in a number of societies that the use of unusually effective or "star" Principals and teachers may not bring reliable knowledge because of the unusual nature of the people concerned, necessitating larger sample sizes, as it were, upon which to base judgements and policies;
- The need for "thought leadership", given the exhaustion of many of the educational paradigms that have been utilized in the last decade, seen in the lack of the "demand side" levers of parental choice, publication of assessment data, creation of multiple kinds of schools and creation of educational markets having much traction in the societies where they have been employed (see the UK and USA results in PISA for example, or the analyses internationally). It may be that governments will be sensibly thinking again about getting their "supply side" right.

At the level of better links with practice and practitioners, it is difficult to see substantial changes in the present limited take up. If societies under the kind of economic pressures outlined above turn to programmes/policies to up-skill their professionals, then there needs to be attention given to ensuring that firstly, "the knowledge base" of EER is well communicated, but second, that equivalent attention is given to ensuring practitioners can "create" better and more valid knowledge in their particular educational contexts, through high-quality data systems, for example. These two sets of policies have been rarely combined.

EER in a changing international society

Although EER can be traced back in origins to the middle to late 1970s, we noted earlier that the real impact of EER started to change how school systems operated in the 1990s. Townsend (2010) documents patterns of change in education over the course of time where education has moved from thousands of years of being an individual activity for the rich and privileged (Beare, 1997), to where communities have taken responsibility for education of their populations. This was first locally, for more than a hundred years, than more recently nationally, and now to the present where international comparative testing has created new imperatives for schools and school systems worldwide. The EER research has been used as justification for many changes in education, in governance, in the way schools are managed, the way in which school leaders and teachers go about their work and how that work is judged.

However, if we think back and look at the progress of the computer over its first 30 years (from 1943 until 1973) and then consider its progress over the last 30 years (from the 1970s on), we might start to ponder how the next 30 years of school effectiveness research might further alter the landscape of education. Townsend (2010) argues that we are approaching a time where we need to think and act both locally and globally, so it might be instructive to think of how that may play out when considering school effectiveness. There are different ways in which local and global can be interpreted and it is possible to do this at different levels of operation. In a classroom, local is the individual student and global is the class as a whole. So to think and act both locally and globally at this level is to consider how to cater to the needs of the individual while ensuring that the whole class moves forward. Similar perspectives might be given for individual classrooms (local) and the school as a whole (global), individual schools and the system (e.g. Local Authority, District) as a whole, individual systems and the country as a whole, and individual countries and groups of countries.

An international perspective is of vital importance, since EER may not mean the same thing in different parts of the world. Bisschoff and Rhodes (2011) provide ample evidence that the same rules cannot apply equally to western countries, which have had a hundred years of developing a universal education system before the whirlwind of recent change, and to other countries that are still dealing with the problem of access for every child. As Hans Rosling, a Swedish health professor, has shown for health improvements (TED, 2007), some countries have made more progress than others, with Asia and India particularly coming from way behind the

OECD countries to having comparable levels of health data, but some countries (mostly sub-Saharan Africa) remain pretty much where they were fifty years ago. The same can be said for schooling. There are still many millions of young people who do not go to school, and EER as the West knows it has no impact on these young people. To think and act both locally and globally suggests that the developed world may have some responsibility to support poorer countries to enable their young people to attend school in the first place, and then to maximize their school quality. There is interesting EER evidence that disadvantaged children are more susceptible to educational effects. Similarly, for such children, high-quality pre-school can act as an effective intervention with lasting effects. The EER field's historic interest in promoting equity and effectiveness needs to receive more attention (Sammons, 2010a).

The concern for education in all countries may need to be paralleled by a concern for addressing schooling for sustainability. The mechanism of "school choice" is the current way that many Western governments are responding to the individualization of responsibility for economic prosperity and well-being, and it is interesting that parents and those choosing schools may have views about the importance of "schooling for sustainability" well ahead of those of governments and, possibly, educational professionals (Kelly, 2009). However, the school practices necessary to develop "eco-literacy" have not been developed, a consequence of little being done to connect students to ideas that see them as humans on an interconnected planet (Clarke, 2009).

However, sustainability, once thought of as being only about the environment, has now developed into a broader concept. Townsend (2010, pp. 23-24) argued:

> Elkington's (1994) definition of the triple bottom-line argues we must consider the economic and the social environment as well as the natural environment for true sustainability to emerge. If we consider only the natural environment and the economic environment we will have viability but not sustainability, if we consider only the natural environment and the social environment we will have bearability but not sustainability and if we consider only the social environment and the economic environment we will have equity but not sustainability. Only when all three are in balance can we have sustainability.

If we suggest that the natural environment of education is schools, we can ask ourselves these questions:

- Are schools equitable for all students?
- Are schools bearable for all students?
- Are all schools viable in the current climate?

We might suggest that schools, as they are currently constructed, managed and operated, are not sustainable. Something needs to change and perhaps EER can contribute to this, by adding cost-effectiveness to its focus on equity and learning.

At a time of recession and public spending cuts in many societies, additionally, the importance of education for disadvantaged groups may be neglected and the challenges facing schools which serve disadvantaged groups are likely to increase. The bottom line for EER and the challenge for the future, locally and globally, is to deliver a quality outcome in ways that are cost effective and supportable by the community as a whole and by politicians. EER has the potential to study such changes and to be an advocate to promote equity and opportunities for the disadvantaged. In the 1970s, EER stood against the currents of the time, intellectually and politically. Maybe it needs to rediscover that radical spirit again, in this and in other areas related to the future of international society.

References

Aikin, W.M. (1942). *The story of the eight year study.* New York: Harper.

Aitkin, M. & Longord, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society.* Series A, 149(1), 1-43.

Alsbury, T.L. (2008). *The future of school board governance.* Lanham, MD: Rowhan & Littlefield.

Askew, M., Rhodes, V., Brown, M., William, D., & Johnson, D. (1997). *Effective teachers of numeracy. Report of a study carried out for the Teacher Training Agency.* London: King's College London, School of Education.

Beare, H. (1997). Enterprise: The new metaphor for schooling in a post-industrial society, in T. Townsend (Ed.), *The primary school in changing times: the Australian experience* (London and New York: Routledge).

Bernstein, B. (1968). Education cannot compensate for society. *New Society, 387,* 344-347.

Bisschoff, T., & Rhodes, C. (2011). Good schools for some but why not better schools for all? Sub-Saharan Africa in transition. In C. Day (Ed.), *International handbook on teacher and school development*, (pp. 901–924). London: Routledge.

Borman, G., Hewes, G., Overman, L., & Brown, S (2003) .Comprehensive school reform and achievement: a meta-analysis. *Review of Educational Research, 73*(2), 125–230.

Borich, G. (1996). *Effective teaching methods* (3rd ed.). New York: Macmillan.

Bosker, R.J., & van der Velden, R.K.W. (1989). The effects of schools on the educational career of disadvantaged pupils. In B.P.M. Creemers & D. Reynolds (Eds.), *Proceedings of the 2nd International Congress for School Effectiveness.* Rotterdam.

Bosker, R., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. In R.J. Bosker, B. P. M. Creemers & J. Scheerens (Eds.), Conceptual and methodological advances in educational effectiveness research (Special Issue). *International Journal of Educational Research, 21,* 159-180.

Brandsma, H.P., & Knuver, J.W.M. (1988). Organisatorische verschillen tussen basisscholen en hun effect op leerlingprestaties. (Organisational differences between primary schools and their effects on pupil achievement). *Tijdschrift voor Onderwijsresearch, 13,* 201-212.

Brookover, W., & Lezotte, C. (1979). *Changes in school characteristics coincident with changes in student achievement.* East Lansing: Michigan State University.

Brophy, J. (1988). Research on teacher effects: uses and abuses. *Elementary School Journal, 89*(1), 3-22.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching, third edition* (pp. 328-376). New York: Macmillan.

Campbell, J., Kyriakides, L., Muijs, D. ,& Robinson, W. (2004). *assessing teacher effectiveness: developing a differential model.* London: Routledge Falmer.

Clarke, P. (2009). Sustainability and improvement: a problem 'of' education and 'for' education. *Improving Schools, 12*(1), 11-17.

Coleman, J.S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Washington: DC: US Government Printing Office.

Cooper, L., Hedges, L.V., & Valentine, J.C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage.

Creemers, B.P.M. (1994). *The effective classroom.* London: Cassell.

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., & Sammons, P. (2010). *Methodological advances in school effectiveness research.* London: Routledge.

Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Education Administration Quarterly, 44*(1), 121-153.

Day, C., Stobart, G., Kington, A., Gu, Q., Smees, R., & Mutjuba, T. (2006). *Variation in teachers' work, lives and effectiveness (VITAE).* Nottingham: DCFS Publications.

Day, C., Sammons, P., Stobart, G., Kington, A., & Gu, Q. (2007). *Teachers matter: connecting work, lives and effectiveness.* Maidenhead*:* McGraw-Hill International.

Day C., Sammons, P., Kington, A., Regan, E., Ko J., Brown E., Gunraj J., & Robertson D. (2008). *Effective classroom practice (ECP): a mixed-method study of influences and outcomes*. Report submitted to the Economic and Social Research Council (RES-000-23-1564). School of Education, University of Nottingham.

Day, C., Sammons, P., Leithwood, K., Hopkins, D., Gu, Q., Brown, E., & Ahtaridou, E. (2011). *Successful school leadership: linking with learning.* Maidenhead*:* McGraw-Hill International/Open University Press.

De Fraine, B., Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: a multivariate multilevel latent growth approach. *Contemporary Educational Psychology, 32*, 132-150.

Doolaard, S. (2002). Stability and change in results of schooling. *British Educational Research Journal, 28,* 773- 787.

Duke, D.L. (1995). *The school that refused to die: continuity and change at Thomas Jefferson High School.* Albany, NY: SUNY Press.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership,* 37, 15-27.

Elkington, J. (1994). Towards the sustainable corporation: win-win-win business strategies for sustainable development. *California Management Review, 36*(2), 90-100.

Glass, T.E. & Franceschini, L.A. (2007). *The state of the American school superintendency.* Lanham, MD: Rowhan & Littlefield.

Goldstein, H. (2003). *Multilevel models in educational and social research: 3rd ed.* London: Edward Arnold.

Goldstein, H., & Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society,* (Series A), *159*(1), 149-163.

Gray, J., Reynolds, D., Fitz-Gibbon, C., & Jesson, D. (1996). *Merging traditions: the future of research on school effectiveness and school improvement.* London: Cassell.

Gray, J., Hopkins, D., Reynolds, D., Wilcox, B., Farrell, S., & Jesson, D. (1999). *Improving schools: performance and potential.* Buckingham: Open University Press.

Grisay, A. (1996). *Evolution des acquis cognitifs et socio-affectifs des eleves au cours des annees de college.* (Cognitive and socio-affective development during the years in the middle school). Liege: Universite de Liege.

Guildemold, H., & Bosker, R. (2009). School effects on students' progress – a dynamic perspective. *School Effectiveness and School Improvement, 20*(2), 255-268.

Gustafsson, J.E. (2007). Understanding causal influences on educational achievement through analysis of differences over time within countries. In T. Loveless (Ed.), *Lessons learned: what international assessments tell us about math achievement* (pp. 37-63). Washington, D.C.: The Brookings Institution.

Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education, 94,* 328-355

Harris, A., Chapman, C., Muijs, D., Russ, J., & Stoll, L. (2006). Improving schools in challenging contexts: exploring the possible. *School Effectiveness and School Improvement, 17*(4), 409-424.

Hattie, J.A.C. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., & Redding, S. (2008). *Turning around chronically low performing schools: a practice guide* (NCEE # 2008-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from: http://ies.ed.gov/ncee/wwc/practiceguides.

Hill, P., & Rowe, K. (1996). Multilevel modeling in school effectiveness research. *School Effectiveness & School Improvement, 7,* 1, 1-34.

Hill, P., & Rowe, K. (1998). Modelling student progress in studies of educational effectiveness. *School Effectiveness & School Improvement, 9,* 3, 310-333.

Hopkins, D., & Reynolds, D. (2001). The past, present and future of school improvement: towards the third age, *British Educational Research Journal, 27*(4), 459-475.

Hox, J., & De Leeuw, E.D. (2003). Multilevel models for meta-analysis. In Reise & N. Duan (Eds.), *Multilevel modeling: methodological advances, issues and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jencks, C.S., Smith, M., Ackland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: a reassessment of the effect of the family and schooling in America.* New York: Basic Books.

Kelly, A. (2009). *Education choice and schooling theory: adapting Sen's early work on 'capability' to choice and sustainability*. Unpublished paper.

Konu, A.I., Lintonen, T.P., & Autio, V.J. (2002). Evaluation of well-being in schools: a multilevel analysis of general subjective well-being. *School Effectiveness and School Improvement, 13*, 187-200.

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: a step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement, 19*(4), 501-529.

Kyriakides, L., & Creemers, B.P.M. (2008). 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), 183-206.

Kyriakides, L., & Creemers, B.P.M. (2009). The effects of teacher factors on different outcomes: two studies testing the validity of the dynamic model. *Effective Education, 1*(1), 61-85.

Land, D. (2002). Local school boards under review: their role and effectiveness in relation to students' academic achievement. *Review of Educational Research, 72*(2), 229-278.

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement, *Journal of the Royal Statistical Society, 172*(3), 537-554.

Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. *Oxford Review of Education, 32*, 397-429.

Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: from a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement, 16*(3), 249-279.

Luyten, H., Peschar, J., & Coe, R. (2008). Effects of schooling on reading performance, reading engagement and reading activities of 15-year-olds in England. *American Educational Research Journal, 45*, 319-342.

Luyten, H., Tymms, P., & Jones, P. (2009). Assessing school effects without controlling for prior achievement? *School Effectiveness and School Improvement, 20*, 145-166.

MacBeath, J., & Mortimore, P. (2001). *Improving school effectiveness.* Buckingham: Open University Press.

Mangan, J., Pugh, G., & Gray, J. (2005). Changes in examination performance in English secondary schools over the course of a decade: searching for patterns and trends over

time. *School Effectiveness and School Improvement, 16*, 19-50.

Martin, A., Bobis, J., Anderson, J., Way, J., & Vellar, R. (2011). Patterns of multilevel variance in psycho-educational phenomena: comparing motivation engagement, climate and achievement factors. *Zeistschrift fur Padagogische Psychologie, 25*(1), 1-13.

Marzano, R.J. (2003). *What works in schools: Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R.J. (2007). Leadership and school reform factors. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 597- 614). New York: Springer.

Marzano, R.J. & Waters, T. (2009). *District leadership that works.* Bloomington, IN: Solution Tree.

McCaffrey, D.F., Lockwood, J.R., Koretz, D.M., Louis, T.A. & Hamilton, L.S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*(1), 67-101.

Mortimore, P. (1991). School effectiveness research: which way at the crossroads? *School Effectiveness and School Improvement, 2*(3), 213-229.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Somerset: Open Books (reprinted in 1995 by Paul Chapman: London).

Muijs, D. (2006). New directions for school effectiveness research: towards school effectiveness without schools. *Journal of Educational Change, 7*. 141-160.

Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness. some preliminary findings from the evaluation of the mathematics enhancement programme. *School Effectiveness and School Improvement, 11*(2), 323-327.

Muijs, D., & Reynolds, D. (2003). Student background and teacher effects on achievement and attainment in mathematics. *Educational Research and Evaluation. 9* 1), 289-313.

Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas: a review of research evidence. *School Effectiveness and School Improvement, 15*(2), 149-175.

Muijs, D., & and Reynolds, D. (2010). *Effective teaching: research and practice.* London: Paul Chapman.

Murnane, R.J. (1981). Interpreting the evidence on school effectiveness. *Teachers College*

Record, 83, 19-35.

Opdenakker, M.-C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement, 11,* 165-196.

Opdenakker, M.-C., & Van Damme, J. (2006). Teacher characteristics and learning styles as effectiveness enhancing factors of classroom practice. *Teaching and Teacher Education, 22*(1), 1-21.

Palardy, G. (2008). Differential school effects among low, middle, and high social class composition schools: a multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement, 19,* 21-49.

Payne, C.M. (2008). *So much reform, so little change.* Cambridge MA: Harvard Education Press.

PISA-Konsortium Deutschland. (2006). *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres.* (Research on the development of competencies during one school year). Munster: Waxmann.

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*, 281-311.

Raudenbush, S. (2011). Modelling mediation – causes, markers and mechanisms. Opening address of the Conference of the Society for Research on Educational Effectiveness (SREE). Washington DC, March.

Reynolds, D. (1976). The delinquent school. In M. Hammersley & P. Woods (Eds.), *The process of schooling*. London: Routledge and Kegan Paul.

Reynolds, D. (1996). Special Issue: Country reports from Australia, the Netherlands, the UK and the United States of America. *School Effectiveness and Improvement, 7* (2), 111-195.

Reynolds, D. (2007). *Schools learning from their best – the within school variation project*. Nottingham: NCSL.

Reynolds, D. (2008). New Labour, education and Wales: the devolution decade. *Oxford Review of Education, 34*(6), 753-765.

Reynolds, D. (2010a). Smart school improvement: towards schools learning from their best. In M. Fullan, A. Hargreaves & D. Hopkins (Eds.), *the international handbook of*

educational change, 2nd ed. Dordrecht: Kluver.

Reynolds, D. (2010b). *Failure free education? The past, present and future of school effectiveness and school improvement.* London: Routledge.

Reynolds, D., Hopkins, D., & Stoll, L. (1993). Linking school effectiveness knowledge & school improvement practice: towards synergy. *School Effectiveness & School Improvement, 4*(1), 37-58.

Reynolds, D., Creemers, B.P.M., Stringfield, S., Teddlie, C., Schaffer, E., & Nesselrodt, P.S. (1994). *Advances in school effectiveness research and practice.* Oxford: Pergamon Press.

Reynolds, D., Bollen, R., Creemers, B.P.M., Hopkins, D., Stoll, L., & Lagerweij, N. (1996). *Making good schools: linking school effectiveness and school improvement*. London: Routledge.

Reynolds, D., & Muijs, D. (1999). *Numeracy: an annotated bibliography for schools and colleges*. London: DfEE.

Reynolds, D., Hopkins, D., Potter, D., & Chapman, C. (2001). *School improvement for schools facing challenging circumstances: a review of research and practice.* London: Department for Education and Skills.

Reynolds, D. & Teddlie, C. (2001). Reflections on the critics, and beyond them. *School Effectiveness & School Improvement.* 12, 99-114.

Reynolds, D., Creemers, B.P.M., Stringfield, S., Teddlie, C., & Schaffer, G. (2002). *World class schools: international perspectives in school effectiveness*. London: Routledge Falmer.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. Wittrock (Ed.) *Handbook of research on teaching,* 3rd ed., pp. 376-391. New York: Macmillan.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. with Smith, A. (1979). *Fifteen thousand hours: secondary schools and their effects on children*. London: Open Books and Boston, MA: Harvard University Press.

Sammons, P. (1996). Complexities in the judgment of school effectiveness, *Educational Research and Evaluation, 2*(2), 113-149.

Sammons, P. (1999). *School effectiveness: coming of age in the 21st century*. Lisse: Swets & Zeitlinger.

Sammons, P. (2008). Zero tolerance of failure and New Labour approaches to school

improvement in England. *Oxford Review of Education, 34*(6), 651-664.

Sammons, P. (2010a). Equity and educational effectiveness. In P. Peterson, E. Baker & B. McGaw (Eds.) *International encyclopedia of education, Vol. 5 , Leadership and Management* – Politics and Governance (pp. 51-57). Oxford: Elsevier.

Sammons, P. (2010b). The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods Research* 2nd ed., pp. 697-723. Thousand Oaks: Sage.

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: results from a reanalysis of the Inner London Education Authority's Junior School project data, *British Educational Research Journal*, 19 (4) 381-405.

Sammons, P., Hillman, J. & Mortimore, P. (1995). *Key characteristics of effective schools: a review of school effectiveness research.* London: OFSTED.

Sammons, P., Nuttall, D., Cuttance, P., & Thomas, S. (1995). Continuity of school effects: a longitudinal analysis of primary and secondary school effects on GCSE performance. *School Effectiveness and School Improvement, 6,* 285-307.

Sammons, P., Thomas, S., & Mortimore, P. (1997). *forging links: effective departments and effective schools.* London: Paul Chapman.

Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., Barreau, S., & Grobbe, S. (2008). *Effective Pre-School and Primary Education 3-11 Project (EPPE): The Influence of School and Teacher Quality on Children's Progress in Primary School.* Nottingham: DCSF.

Sammons, P., & Luyten, H. (2009) Editorial article for special issue on alternative methods for assessing school effects and schooling effects. *School Effectiveness and School Improvement, 20*(2), 133-143.

Sammons, P., Gu, Q., Day, C., & Ko, J. (2011). Exploring the impact of school leadership on pupil outcomes: results from a study of academically improved and effective schools in England. *International Journal of Educational Management, 25*(1), 83-101.

Schaffer, E., Nesselrodt, P., & Stringfield, S. (1997). *Impediments to reform: an analysis of destabilizing issues in ten promising programs.* Arlington, VA: Educational Research Science.

Scheerens, J., & Bosker, R. (1997). *The foundations of school effectiveness*. Oxford: Pergamon Press.

Shelton, T. (2010). *The effects of school system superintendents, school boards, and their interactions on longitudinal measures of districts' students' mathematics achievement.* Unpublished doctoral dissertation. Louisville, KY, USA: University of Louisville.

Singer, J., & Villett, J. (2003). *Doing data analysis in the multilevel model for change.* Cambridge, MA: Havard University Graduate School of Education.

Slee, R., & Weiner, G. (2001). Education reform and reconstruction as a challenge to research genres: Reconsidering school effectiveness research and inclusive schooling. *School Effectiveness and School Improvement, 12*, 83-98.

Smith, D.J., & Tomlinson, S. (1989). *The school effect. a study of multi-racial comprehensives.* London: Policy Studies Institute.

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement, 21,* 289-314.

Stringfield, S. (1998). An anatomy of ineffectiveness. In L. Stoll & K. Myers (Eds.), *No quick fixes: perspectives on schools in difficulties.* (pp. 209-221). London: Falmer.

Stringfield, S. (2002). Science making a difference: let's be realistic! *School Effectiveness and Improvement, 13*(1), 15-28.

Stringfield, S. (2008). School Boards and raising student outcomes: reflections (confessions?) of a former urban School Board member. In T. Alsbury (Ed.), *The future of School Board governance: relevancy and revelation.* Lanham, MD: R & L Education.

Stringfield, S., & Teddlie, C. (1988). A time to summarize: six years and three phases of the Louisiana School Effectiveness Study. *Educational Leadership, 46*(2), 43-49.

Stringfield, S., & Yakimowski, M. (2005). The promise, progress, problems and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore City Public Schools. *American Educational Research Journal,* 42 (1) 43-76.

Supovitz, J.A., & Weinbaum, E.H. (Eds.). (2008) *The implementation gap.* New York, NY: Teachers' College Press.

TED (2007) Hans Rosling's new insights on poverty (videotape), available at

http://www.ted.com/talks/hans_rosling_reveals_new_insights_on_poverty.html. Retrieved 12 August 2011.

Teddlie, C. (2010). The legacy of the school effectiveness research tradition. In A. Hargreaves, A. Lieberman, M. Fullan & D. Hopkins (Eds.), *The second international handbook of educational change.* Dordrecht: Springer.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: lessons learned from a ten year study of school effects*. New York: Teachers College Press.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer Press.

Teddlie, C., & Reynolds, D. (2001). Countering the critics: responses to recent critics of school effectiveness research. *School Effectiveness & School Improvement, 12*, 41-82.

Teddlie, C., Creemers, B.P.M., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for teacher observation and feedback: evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation, 12*(6), 561-582.

Teddlie, C., & Sammons, P. (2010). Applications of mixed methods to the field of educational effectiveness research, In B.P.M. Creemers, L. Kyriakides & P. Sammons (Eds.), *Methodological advances in educational effectiveness Research.* London: Routledge Falmer.

Thomas, L. (1979). Medical lessons from history. In L. Thomas (Ed.). *The medusa and the snail.* New York: Viking.

Thomas, S. (2001). Dimensions of secondary school effectiveness: comparative analyses across regions. *School Effectiveness and School Improvement, 12*, 285-322.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997a). Differential secondary school effectiveness: comparing the performance of different pupil groups. *British Educational Research Journal, 23*, 451-469.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997b). Stability and consistency in secondary schools' effects on students' GCSE outcomes over three years. *School Effectiveness and School Improvement, 8*, 169-197.

Thrupp, M. (2001). Sociological and political concerns about school effectiveness research: time for a new research agenda. *School Effectiveness and School Improvement,* 12, 7-40.

Thrupp, M. (2002). Why "meddling" is necessary: a response to Teddlie, Reynolds, Townsend , Scheerens, Bosker & Creemers. *School Effectiveness and School Improvement,* 13(1), 1-14.

Townsend, T. (1994). *Effective Schooling for the Community*, London and New York: Routledge, 246 p.

Townsend, T. (Ed.) (2001a). The background to this set of papers on two decades of school effectiveness research, Special Issue critique and response to twenty years of SER. *School Effectiveness and School Improvement, 12*(1), 3-5.

Townsend, T. (2001b). Satan or saviour? An analysis of two decades of school effectiveness research. *School Effectiveness and School Improvement, 12*(1), 115-129.

Townsend, T. (Ed.) (2007). *The international handbook of school effectiveness and school improvement.* New York: Springer.

Townsend, T. (2010). Thinking and acting both locally and globally: new issues for school development planning. a keynote paper presented to the International Professional Development Association (IPDA) annual conference. November, Birmingham, England.

Tyack, D., & Cuban, L. (1995). *Tinkering towards Utopia: a century of public school reform.* Cambridge, MA: Harvard University Press.

Tymms, P., Merrell, C., Heron, T., Jones, P., Albone, S., & Henderson, B. (2008). The importance of districts. *School Effectiveness and School Improvement, 19*, 261-274.

Van Damme, J., Opdenakker, M., Van Landeghem, G., De Fraine, B., Pustjens, H., & Van de gaer, E. (2006). *Educational effectiveness: an introduction to international and Flemish research on school, teachers and classes.* Leuven: Acco.

Van de gaer, E., De Fraine, B., Van Damme, J., De Munter, A., & Onghena, P. (2009). School effects on the development of motivation toward learning tasks and the development of academic self-concept in secondary school. *School Effectiveness and School Improvement, 20*, 235-253.

Van de Grift, W., & Houtveen, A.A.M. (2006). Underperformance in primary schools. *School Effectiveness and School Improvement, 17*(3), 255-273.

Van de Grift, W. (2007) Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. *Educational Research, 49*(2), 127–152.

Van de Grift, W. (2009). Reliability and validity in measuring the value added of schools. *School Effectiveness and School Improvement, 20,* 269-286.

Van der Wal, M., & Waslander, M. (2007). Traditional and non-traditional educational outcomes: trade-off or complementarity? *School Effectiveness and School Improvement, 18,* 409-428.

Van der Werf, G., Opdenakker, M.-C., & Kuyper, H. (2008). Testing a dynamic model of student and school effectiveness with a multivariate multilevel latent growth curve approach. *School Effectiveness and School Improvement, 19,* 447-462.

Van Landeghem, G., Van Damme, J., Opdenakker, M.-C., De Fraine, B., & Onghena, P. (2002). The effect of schools and classes on noncognitive outcomes. *School Effectiveness and School Improvement, 13,* 429-451.

Verachtert, P., Van Damme, J., Onghena, P. & Ghesquiere, P. (2009). A seasonal perspective on school effectiveness: evidence from a Flemish longitudinal study in kindergarten and first grade. *School Effectiveness and School Improvement, 20*(2), 215-233.

Weber, G. (1971). *Inner city children can be taught to read: four successful schools.* Washington DC: Council for Basic Education.

Willms, J.D. (1986). Social class segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review, 51* (2), 224-241.

Wray, D. & Medwell, J. (2001). *What can teachers of literacy learn from a study of effective teachers?* Paper presented at the 12th European Reading conference, July, Dublin, Ireland.