Educational Effectiveness Research (EER):

A State of the Art Review

David Reynolds, University of Southampton, UK
Pam Sammons, University of Oxford, UK
Bieke De Fraine, Catholic University of Leuven, Belgium
Tony Townsend, University of Glasgow, UK
Jan Van Damme, Catholic University of Leuven, Belgium

Paper presented to the annual meeting of the International Congress for School Effectiveness and Improvement, Cyprus, 2011

Any comments, suggestions about material that may have been omitted and views about the conclusions of this review are very welcome, and should be sent to D.Reynolds@soton.ac.uk or pamela.sammons@education.ox.ac.uk
**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 summarised as:

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

It looks at all the factors within schools in particular, and the educational system in general, that might affect the learning outcomes of students in 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 and the effects of educational ‘learning environments’ in general.

To undertake this complex task, EER involves sorting out the effects of schools from other effects such as those of prior attainment and educational background, which therefore requires modelling tools that involve comparable complexity (Goldstein, 1995). 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;
- 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 of the breadth which 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 (Gray et al., 1996; Mortimore, 1991; Sammons, 1999; Townsend, 2007; Reynolds, 2010b; Teddlie, 2010; Creemers, Kyriakides & Sammons, 2010; Reynolds et al., 1994; Teddlie & Reynolds, 2000), so we will only briefly summarise the phases that EER has gone 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), who concluded that schools had little effect upon the outcomes of their students by comparison to the effects of their 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 & 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, 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, the consistency of school effects upon different outcome domains, the differential effects of school upon students of different background characteristics, the size of school effects and the long term effects of schools (see reviews in the Special Country Reports Issue of School Effectiveness and School Improvement, Vol. 7 (2), 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 United States, and work in the United Kingdom into subject Department effects upon performance and also upon school effects (Sammons et al, 1997). These years also saw a number of influential reviews of the field such as by Scheerens & Bosker, (1997) and Reynolds et al, (1996).

The fourth phase, which began in the mid to late 1990s, ran for a decade and is still in evidence today, 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. 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 SE and 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 organisational 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.

The fifth phase has begun only recently in the late 2000s and is still developing rapidly, focussing 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.

However, 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 (ICSEI), a journal located in the ISI *Current Contents* databases within six years of its launch and a volume of research books and papers that justified over 1,500 references in the Teddlie & Reynolds (2000) *International Handbook of School Effectiveness Research*, and probably double that in the replacement volume currently being written and planned for 2012 publication. All these factors suggest considerable scientific success.

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 of course easy to understand, given the need to demonstrate beyond doubt that schools varied in their effects, 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 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 recognised need for 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, for policymakers and practitioners particularly. 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 standardised systems of measurement, very few of which even had the qualitative data as examples of what the quantitative instruments were measuring, although there were some early mixed methods approaches (e.g. Teddlie & Stringfield, 1993; Sammons, Thomas & Mortimore, 1997).

Secondly, 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 a separate tradition of teacher effectiveness that has developed in parallel. Creemers, (1994) was one of the first to point this out, but it has taken another decade for ‘teaching’ to receive anything like the attention given to ‘schooling’. Given the clear evidence that teacher effects greatly exceed school effects when progress over time is studied (Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000; Muijs & Reynolds, 2010), 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.

Of course 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. 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.

Thirdly, the rise of multilevel 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 utilised – it made possible the handling of multiple variations, differential effects and cross level interactions.

However, MLM only made it possible to study direct educational effects – of teachers on students for example – rather than indirect or reciprocal relationships, of students on teachers in their behaviours for example. Additionally, in its early development it 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 was not a focus of early 'MLM. 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, though 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.

Fourthly, 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 without itself, given that there was little need, given the intellectual and geographical variance that was already on offer, to learn from outside. The rise of a number of critical perspectives on the discipline (see the papers from the AERA debate such as Thrupp, 2001, 2002; Slee & Weiner, 2001; Teddlie & Reynolds, 2001; and Reynolds & Teddlie, 2001; together with the overview of Townsend, 2001) 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 and cliques);
- Cognitive neuroscience (on 'brain based' interventions);
- Psychology of education (on multiple outcomes);
- Humanistic psychology (on self conception).

Fifthly, 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 conceptualised 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. Additionally, 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.

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

- 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 helping to show 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 which could avoid the need for the reinvention of the wheel by the teaching professions of different countries.

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 ‘organise’ 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 which was used to discuss contextual variation, or
coalition building from political science which was 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 & 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 & Creemers, 2008; Kyriakides, 2008; Kyriakides & Creemers, 2009), with promising results. Other interesting theoretical speculations of a different kind are in Luyten et al, (2005) and Van De Grift, (2006).

Secondly, 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 & Sammons (2010), have also discussed methodological issues.

- Multi-Level Modelling. 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 recognises 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, 1996, 1998; that demonstrated not only that teacher effects tend to be larger than school effects, but also that in combination they could account for a substantial proportion of the variance in student outcomes).

Improvements in the modelling of measurement error and interaction effects, cross-classified models that examine multiple institutional membership and regression discontinuity to study the size of the school 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 Leeuv, 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 & 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 & 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.

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


EER investigates the effects of schools on student outcomes. 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 broad 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 (Teddle & Reynolds, 2000). It was found that the primary school can have a long-lasting, but small, effect on student achievement in secondary education (Sammons, Nuttall, Cuttance, & Thomas, 1995; Thomas, 2001) and that 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 (like PISA, TIMSS, PIRLS, etc.) 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).

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. 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).

There is also a trend to investigate absolute school effects instead of only relative school effects. Educational effectiveness research 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.

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 intraclass correlation). The size of the school effect is largely affected by the outcome under study. Schools have a larger effect on student achievement than on non-cognitive outcomes (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 are given less emphasis in the curriculum. Second, the
measurement of these non-cognitive outcomes is less precise as compared to the measurement of achievement. Also, school effects tend to be larger for subjects such as mathematics and science, as compared to school effects for language (Thomas, Sammons, Mortimore & Smees, 1997a). Also, longitudinal studies examining student growth are more likely to demonstrate school effects of larger magnitude (Teddlie & Reynolds, 2000; Van de gaer et al., 2009).

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 were found. However, some studies (especially in secondary education) found that some schools can be effective for one subject and ineffective for another (Thomas, 2001; Thomas, Sammons, Mortimore & Smees, 1997b). Inconsistency in primary school could indicate that the teacher is better in one subject than the other. 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 (van der Wal, & Waslander, 2007).

To summarise, 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).
Moving on to look at stability over time, this 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 is some stability for general school results, but much more fluctuations for specific subjects that are interpreted as departmental effects (Thomas et al., 1997b).

Absolute stability over several years is impossible because schools are changing organisations. There can be changes in schools' policy, 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.

We conclude that school effects are relatively stable over time. The correlations between effects in 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).

Moving on to look at 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 socio-economic status. School effects can be a function of student ability level or prior achievement (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 (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 some differential effects with regard to student socio-economic status (Strand, 2010; Thomas, 2001; Thomas et al., 1997).

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.

We now move on to consider 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 have been the initial wave of small scale case studies from the USA (e.g. Weber, 1971), the mixed method longitudinal studies from London (Rutter et al, 1979; Mortimore et al, 1988) into both primary and secondary schools and the Louisiana School Effectiveness Studies (Teddlie & Stringfield, 1993), and a wide range of other studies that 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 & 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 reviews of even more contemporary literature by Marzano (2003, 2007), undertaken from the different paradigm of educational administration, show 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 United States (Hallinger & Murphy, 1986; Teddlie & Stringfield, 1993) focused upon the differences in the processes of effective schools that occurred in different socio economic 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 (Reynolds et
al, 2001; Muijs et al, 2004; Harris et al, 2006), 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 socio economic 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. religiosity 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 leadership (directive in Oriental cultures, more 'lateral/vertical' in Anglo Saxon ones) is context dependent.

Whilst another 'State of the Art' review will be 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.

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 60 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 one 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.

Further recent research (Day, Sammons & Kington, 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 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 EPPE (Effective Pre-School and Primary Education 3-11) Project shows that the influence of overall Teaching Quality upon Maths 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 VITAE (Variation in Teachers Work, Lives and Effectiveness) Project 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).

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'.

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 & Franeschini (2007) generated a study and review of research on the American school superintendency. Marzano & 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 since Tymms et al, (2008) show minimal District effects the usefulness of this new emphasis - intellectually and in policy terms - may need to be established.

Having now reviewed the scientific properties of EER, and the processes associated with effectiveness, we now move on 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 which 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 United States 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 (see Reynolds, 2010a; Sammons, 2008), but the association with ‘prescription’ meant that the influence was short lived. However, the English inspection agency OFSTED utilised EER in its Inspection Framework (see Sammons, Hillman & Mortimore, 1995), and the documentation upon school improvement that each English school fills in 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) shows interest - but not mainstreaming - in many societies. EER is bolt-on, not bloodstream, for the policy making communities.

Reasons for this may be as follows:

- The quantitative statistical knowledge required to fully access some of the knowledge base;
- The considerable volume of criticisms that have 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 policymakers seem happier operating at school rather than classroom level;
- The reluctance to embrace a discipline which increasingly argues for 'contextually specific' policies, given historic policymaker commitment to 'steam press', universal or 'one size fits all' ones;
- The continued tendency for EER to generate findings that are in the United Kingdom 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 importance of the early years rather than later years of education.

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, 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 the 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 of 'teaching';
• 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.

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 foci 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 possible change fully into the field, since it has been seen as a stand alone issue, and there need to be 'more studies where leadership is integrated within a model of school effectiveness which is theorised and takes into account the ways in which leadership interacts with other key school factors.' (Muijs, forthcoming);
• 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 which can detail the processes creating stability and change in schools.' (Stringfield & Teddlie, forthcoming);
• Making sampling across socio economic contexts, kinds of school governance, school types and school Districts axiomatic, rather than attempting to control out such variation in the interests of easier 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 organised 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 United Kingdom 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 utilised, but not the school level factors that are associated with its scope. The theoretical modelling of Creemers & 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 costs 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, limited the extent to which EER is relevant for practitioners who work in the niches in 'whole schools' and has 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, 2010). 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 which 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/her social background.

- 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, behaviour etc
  - 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 outcome - 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 etc - 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, which if it 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, might paint an interestingly different picture of the reasons behind supposedly high achieving school systems, or highly successful individual schools.

**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 judgments 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 imposition of national targets 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 upskill their professionals, then there needs to be attention to ensuring that firstly ‘the knowledge base’ of EER is well communicated, but secondly with equivalent attention being 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, the real impact of EER started to change how school systems operated in the 1990s. Townsend (2009), documents patterns of change in education over the course of history 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, first locally for more than a hundred years, than more recently nationally, and then 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 thirty years (from 1943 until 1973) and then consider its progress over the last thirty years (from the 1970s on), we might start to ponder how the next thirty years of school effectiveness research might further alter the landscape of education. Townsend (2009), 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 and individual systems and the country as a whole.

An international perspective is of vital importance, since EER may not mean the same thing in different parts of the world. Bisschoff & Rhodes (forthcoming) provide ample evidence that the same rules cannot apply equally in western countries, that had a hundred years of developing a universal education system before the tsunami of recent change, and 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 (mostly sub-Saharan Africa) remain pretty much where they were 50 years ago. The same can be said for schooling. There are still many millions of young people who do not go to school, and school effectiveness 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 which indicates 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, 2010).

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 mechanism whereby 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).

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 school which serve disadvantaged groups are likely to increase. 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


Townsend, T. (Ed) (2001). 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*, Vol. 12, No. 1, pp. 3-5.


Van der Werf, G., Opdenakker, M.-C., & Kuyper, H. (2008). Testing a dynamic model of student and school effectiveness with a multivariate multilevel...


