THE ROUTLEDGE INTERNATIONAL HANDBOOK OF EDUCATIONAL EFFECTIVENESS AND IMPROVEMENT. RESEARCH, POLICY AND PRACTICE
By: Christopher Chapman, Daniel Muijs, David Reynolds, Pam Sammons, and Charles Teddlie

Review by: Jaap Scheerens

Introduction

The purpose of the Routledge International Handbook of Educational Effectiveness and Improvement, to be further indicated as "the Handbook", is not explicitly stated in the introduction, but it is obvious that the aim has been to update the state of the art of the field of educational effectiveness and improvement research (EEIR). A similar Handbook was published in 2000 (Teddlie and Reynolds, 2000). The outlook of the Handbook is International, an ambition that is reasonably met in the content covered, despite of the fact that in 15 of the 17 Chapters first authors are based in the UK.

Educational effectiveness research is about goal attainment in education and finding out which conditions, amenable by policy, school management and teaching, work best. This means that "independent" variables like the degree of autonomy of schools in a country, school leadership style and specific teaching strategies are correlated with outcome indicators, like graduation rates and student achievement. This leads to a type of research that is defined "on the skin" of educational policy and practice. The means to ends relationships of policy and practice are interpreted as cause and effect relationships, making the challenge of educational effectiveness research the establishment of causal relationships. Educational effectiveness can therefore be pursued as a scientific research discipline aimed at generalizable knowledge about "what works" and concerned with theory embedded explanatory mechanisms. The combination of a scientific research paradigm and closeness to applied contexts is a central issue in understanding possibilities and limitations of "the field" (as the Handbook calls it). The notion to combine educational effectiveness research with school improvement is a manifestation of this dual interest. The most straightforward way to see this combination is that school improvement applies the results of effectiveness research and school improvement research is involved with the design and monitoring of implementation of change processes. But alternative perspectives on this sometimes uneasy marriage exist, and will be touched upon later on as they show up in the course of this review.

The Handbook consists of 17 chapters (439 pages of text and 78 pages of references), has a clear index and a helpful list of abbreviations.

With an eye to enhancing readability and specific interconnectivity among the chapters a categorization has been followed in this review that differs from the order in which the chapters appear in the book. The sequence in this review is as follows:
I Introduction (Chapters 1 and 2)
II Research methodology (Chapters 7, 8 and 16)
III Theory development (Chapter 6)

IV Knowledge base (Chapters 3 and 4)
V School improvement (Chapters 5 (first part), 13 and 14)
VI System level effectiveness and reform (Chapters 5, (second part), 9, 10, 11 and 12)
VII Conclusion (Chapter 17)

Occasionally in this review Educational Effectiveness Research will be abbreviated as EER, and Educational Effectiveness and Improvement Research as EEIR.

Review per chapter

Part I INTRODUCTION

Chapter 1: Educational effectiveness and improvement research and practice
Christopher Chapman, David Reynolds, Daniel Muijs, Pam Sammons, Sam Stringfield and Charles Teddlie

The introductory chapter speaks of an emergent paradigm, of which five "commitments" have been realized to a larger or smaller degree: 1) A commitment to theory 2) A commitment to methodological sophistication 3) A commitment to the learning level and multiple levels 4) A commitment to measuring the multiple outcomes of education, and finally 5) An outward facing discipline, responding to criticism. Although theory development has been slow in coming the tone of the narrative on these point is quite positive on what has been achieved. In the second part of the chapter a number of challenges for future development are mentioned: 1) To have greater policy and practice impact 2) To understand the application of technology in educational settings 3) To develop a more forward thinking perspective , and 4) Re-conceptualizing the failing school. All of these commitments and challenges are addressed in subsequent chapters and comments will be given when these are reviewed.

Chapter 2: Methodological and scientific properties of school effectiveness research
Pam Sammons, Susila Davis, and John Gray

This chapter offers a comprehensive overview of conceptual and research methodological issues in educational effectiveness research over a period of almost five decades. This means that attention is given to fundamental issues, like the definition of educational effectiveness and the mission of effectiveness research, the existence and magnitude of school effects, the stability of school effects over time, the consistency of effectiveness across sub-systems of the school (classrooms, departments, subject matter domains), differential effectiveness (for different sub-groups of students), and the association between outcomes on cognitive and non-cognitive measures. As far as methodology is concerned the chapter rightly addresses the rather slow and partial progress made in the development of more standardized measures of process variables, mentions progress in the application of statistical analyses techniques (multi-level modelling, multi-level structural equation modeling, and longitudinal studies). The authors have high expectations of "mixed" methods research, in which quantitative and qualitative data are somehow combined and they touch upon the pros and cons of experimental

research design versus the more common correlational designs in educational effectiveness research. On these issues the chapter provides an overview, while the next chapter (chapter 7) is more comprehensive and detailed. The structure of the chapter, based on a development in phases (including a "midlife crisis"), appears a bit of an overdone metaphor. Practically all core-issues for advancement of the field were formulated much earlier. Rather than "a journey across phases" all core issues have more or less been and remained on the agenda across this whole period (including the agreement on a core set of factors at school and teaching level that work), and some areas have shown more progress than others. The way the role of theory in educational effectiveness research is addressed in this chapter is debatable (see the discussion section)

Part II: RESEARCH METHODOLOGY

Chapter 7: Quantitative methods
Daniel Muijs and Annie Brookman

This chapter provides a comprehensive overview of quantitative methods used in educational effectiveness research. Questions of statistical analysis, instrument development and design are addressed systematically. The chapter describes the important progress that was made by the introduction of multi-level modelling, followed by applications of structural equation modelling, and the combination of both in multi-level SEM. Apart from describing important technical developments in educational measurement, generalizability theory and item response theory, the chapter rightly underlines the importance of a long term challenge for the field, namely the relative lack of generally accepted constructs, and standardized instruments to measure them. Next, on research design, the possibilities and limitations of "true" experiments in the field of EER are discussed, as compared to the more common non-experimental, "correlational" designs. The chapter pays considerable attention to mixed methods designs, in which quantitative and qualitative approaches are combined, and, ideally, integrated. The actual research examples are *combinations*, where proper *integration* is apparently more difficult to attain (cf Scheerens, 2013). Quasi-experimental designs, like time series designs, regression discontinuity designs, and the use of matching techniques are also discussed. To this otherwise very complete chapter perhaps some additional attention might have been paid to the introduction of econometric techniques, like "difference in difference" designs and the application of instrumental variables. And finally, to this reviewer's taste, the elephant in the room not addressed is the problem of the relative simplicity of even our most sophisticated quantitative models in matching the complexity of (mostly just implied) contextualized models of educational effectiveness.

Chapter 8: Qualitative methods in educational effectiveness and improvement research
Mike Hadfield and Christopher Chapman

The chapter provides an outlook on qualitative methods that is increasingly expanding: it is not just the use of qualitative data in educational effectiveness research, not even in "school improvement research", although it is acknowledged that qualitative methods are more likely in the context of school

improvement. It is not just about researcher and practitioner partnership but primarily about "professional learning" of acting professionals. It is not just about empirical evidence, but involves a critical perspective, where critical means that educational ends should be questioned for being worthwhile, so that what counts as "improvement" has to be redefined situationally. In "cutting-edge methods in school improvement" external facilitators increasingly depend on "some practitioners who are already involved in inquiry-based approaches to improvement". The term used for these inquiry based approaches is "boutique inquiry approaches"; a term that is not defined in the chapter but suggests a kind of shopping for methods (no practical illustrations are provided). Ambitions are high: "powerful learning experiences" , "in-depth" awareness and conducting causal analyses. When some indications are given on what this is all about some surprisingly traditional schemes pop-up; Dewey's' problem solving model (p. 216) and the realization that "the relationship between the teacher the student and the content" form the instructional core (p. 215). This chapter raises questions about demarcation between professional discourse and professional development on the one hand and scientific "disciplined inquiry" on the other and whether approaches that lack the ultimate benchmark of improved student achievement have a "legitimate" place in the domain that this Handbook seeks to cover.

Chapter 16: Critical and alternative perspectives on educational effectiveness and improvement research
Tony Townsend, John McBeath, and Ira Bogotch

The two central points of criticism discussed in this chapter are reductionism and the rational empiricist nature of the educational effectiveness paradigm.
Reductionism or insufficiently addressing the complexity of education can be dealt with from within main stream educational effectiveness research. The authors implicitly acknowledge this in referring to contributions, firmly rooted in the empiricist tradition, which have made progress in taking into considerations a broader range of educational outcomes (Creemers and Kyriakydes in their dynamic model) and more complex causal models (Hallinger and Heck in their research on indirect leadership effects). They also see the expansion from school effectiveness to educational effectiveness as a step in the right direction.
But further on in their text they qualify these methodological advances as "distractions" (p.406), call for a paradigm change inspired by critical social theory and express themselves critically of a "value-free stance". It is hard to see what implications such a different optic would have. When "effectiveness research" is defined as checking whether goals are attained, and what works best in doing so, goals are accepted as legitimate, ideally as the result of a democratic political process. Even if researchers would call themselves engaged or critical in this debate they would just be one voice among many. They could have a critical and also empirically based role by analyzing the consistency of goals with value positions and political programs, and carry out research that makes inventories on the support for certain priorities. But the authors do not seem to be looking for such "technical" solutions to the issue and appear to call for partiality: critical of government use of effect- and effectiveness studies and in support of teachers resisting external evaluations.

Part III: THEORY DEVELOPMENT

Chapter 6: Theory development in educational effectiveness research
Bert Creemers and Leonidas Kyriakides

The chapter describes disciplinary perspectives of EER; it discusses an economic, a sociological and a psychological perspective. The psychological perspective has been most influential in the sense of behavioristic and cognitivist learning theories. The chapter then discusses the development of integrated models of educational effectiveness, in which school level organizational factors were combined with teaching level factors and with system level contextual conditions. The second part of the chapter is dedicated to the Dynamic Model of Educational Effectiveness, by Creemers and Kyriakides (2008). The dynamic model builds on earlier integrated models, providing elaborations in the sense of a specific interest in studying developments over time, the consideration of non-linear relationships, complex interactions of factors within and between aggregation levels, a broad outlook on effectiveness criteria and specific measurement dimensions of effectiveness enhancing factors. The Dynamic model has generated an interesting set of empirical studies, which are generally confirming the factors that are part of the model. Direct links are made to evidence based school improvement, in which school level levers of classroom level teaching have a key role (professional development and school based evaluation). The Dynamic Model certainly deserves its central place in this chapter on theory development. The way the Dynamic Model was developed from an earlier model, indicated as the Comprehensive Model (Creemers, 1994) is interesting. Where the Comprehensive Model established a set of more general principles of educational effectiveness (consistency between factors at different levels, cohesion among units, constancy and stability over time, and control) and thus provided parsimony on generally applicable dimensions, the Dynamic Model added specificity and complexity. This development could be seen as an attempt to strengthen the validity of the model to cover a complex reality, but, in my opinion, would, at some point, have to re-address more general underlying factors. The primary aim of theory is not mirroring complexity but understanding on the basis of explanatory mechanisms.

Part IV: KNOWLEDGE BASE

Chapter 3: Effective school processes
David Reynolds, Charles Teddlie, Christopher Chapman, and Sam Stringfield

This chapter gives an overview of the results of school effectiveness research, in the sense of malleable school process variables that work. The chapter does this as well as many other reviews have. Despite of the fact that this chapter, as the previous one, suggests development over time, by distinguishing subsequent phases, there is remarkable stability in the identification of the school factors that "work" (leadership, achievement orientation and high expectations, disciplinary climate, parental involvement, monitoring of students' progress, cooperation and professional development). Such consistency over

time, by the way, ought to be considered as a strong point for the stability of the knowledge base. One factor that is missing out in the overview in this chapter is opportunity to learn, in the sense of the degree to which tested subject matter and skills have been covered in the school curriculum and teaching. The chapter does not discuss the evidence to a degree of depth, where effect sizes of these variables are presented; in spite of the fact that several "multi-factor" meta-analyses and many more "single factor" meta-analyses have been available for a while (Scheerens, 2016). These results indicate, among others, that frequently cited individual studies (like the Leadership study by Day et.al, 2010) may overrate average effect sizes. The chapter discusses many important themes that should be taken into consideration for a better interpretation of the current research outcomes and the improvement of future studies: the interaction between school level measures and classroom level processes, differential effects (across sub-groups of students), the importance of contextual characteristics and the processes of "ineffectiveness").

Chapter 4: The scientific properties of teacher effects/effective teaching processes
Daniel Muijs, David Reynolds and Leonidas Kyriakides

This chapter provides a well-structured overview of the development of teaching effectiveness research. It lays out an accumulating evidence base, with strong support for structured "direct" teaching approaches in basic skill's attainment, the identification of a few dimensions of "constructivism" that showed consistent effects (learning of learning strategies in particular) and subsequent integration of behaviorist and cognitivist elements in more recent research and model development. The chapter also looks at developments in brain-imaging methodologies, and considers this as an orientation of growing importance. Although the title of the chapter mentions both teacher effects and teaching effectiveness, the emphasis is completely on the latter. Research on more or less stable personal characteristics of teachers, like education and experience, personality characteristics and teacher knowledge is not explicitly addressed in the chapter. Although the former two types of characteristics have not been particularly successful in education production function research, the latter, teacher knowledge, and particularly teacher pedagogical content knowledge are very promising (Scheerens and Bloemeke, 2016).

Part VI: SCHOOL IMPROVEMENT

Chapter 5: School improvement and systemic reform
David Hopkins (the part on school improvement)

The chapter is structured according to a sequence of phases in the development of school improvement work. Phase 1 refers to the organization developmental approach, with M. Miles as the key figure (1960's- 1970's). Phase 2 (1980's) is described as "action research and individual action", in which concern for educational outcomes (A Nation at Risk) became an important driver for connection with the school effectiveness movement (Edmonds, 1979), with high expectation of school based review as a major manifestation of this concern. Phase 3 "Managing change and comprehensive approaches to

school reform "(1990's and onwards) featured site-based management and Comprehensive School Reform (CSR) programs, emphasizing on the one hand, school autonomy, but on the other hand external support for school improvement. The CSR initiatives and programs under the label of "Schools as High Reliability Organizations" can also be seen as efforts in evidence based reform, where the results of empirical school effectiveness research formed the substantive core of these programs. Phase 4, (2000 and onwards) describes new emphases on capacity building and leadership in school improvement. Educational leadership, new forms of "data driven" internal school evaluation, and cooperative structures, within and between schools received particular emphasis. The systematic overview that is offered by this chapter is very useful. An issue that is only partially addressed is evidence on the effects, of the various emphases in school improvement approaches. Effects of CSR programs are assessed as "patchy and variable", the fact that leadership effects are often very small is not mentioned and no mention is made about the very thin evidence about the effectiveness of teacher and school cooperation.

Chapter 13: Leadership development and issues of effectiveness
Jim O'Brien and Christine Forde

The chapter starts out with a strong statement about the "centrality of leadership" in ensuring effectiveness: "a principle now 'globalized' and enshrined in sets of policy initiatives in different national education systems" (326). In a Handbook on Educational Effectiveness *Research*, one might next have expected some support, corroboration or explicit challenge of this statement by empirical research results. Yet, the chapter meets this expectation only in a rather indirect way, as it is focused on the effectiveness of school leadership professional development programs by means of program evaluations. About these it is concluded that "Evaluations tend to be limited, focusing on the package of a particular CPD activity and rarely achieving insights into changes in classroom activity or student outcomes – which may also be said of the large-scale studies today" (347). The fact of the matter is that reviews and meta-analyses of school leadership effects on student outcomes tend to show modest to very small effects and challenge Leithwood's claim that "school leadership is second only to classroom instruction among all school related factors that contribute to what students learn at school", uncritically cited on page 330 of the chapter. There are many reasons to confirm the importance of school leadership development, but evidence based argumentation for its effect on student achievement is not one of them.

Chapter 14: Educational effectiveness and improvement research, and teachers and teaching
Louise Stoll, Lorna Earl, Stephen Anderson, and Kim Schildkamp

This chapter positions school improvement as implementation of knowledge on effective schooling and teaching. It provides a systematic overview of key strategies for school improvement: using data, enhancing teacher and teaching effectiveness, collaborative learning cultures and leadership oriented at teaching and learning. Data use is referred back to one of the recognized school effectiveness enhancing factors, "frequent monitoring of student achievement". The change in terminology to "data"

and “data use” has associations with the fact that schools currently have many data sources, and in particular applications (e.g the data team method, Lai & Schildkamp, 2013) emphasize that a broad set of input, process and outcome data is supposed to be considered. But data use belongs to a larger family of internal or school self-evaluation, formative assessment and “internal accountability”. In theory this family of improvement strategies has great potential, although getting it to actually work is to be seen as an improvement program in itself. Enhancing teaching and learning is described in terms of particular syntheses of knowledge on effective teaching on the one hand and supportive implementation processes on the other. Comprehensive School Reform programs are a case in point. The authors discuss several successful examples, but their overall conclusion is quite reserved; they see successful examples as rather exceptional. They question the paradigm of externally developed programs requiring fidelity in implementation. They argue that engagement of teachers should be acquired as well as more fundamental issues of conceptualization of changed practice. Support should come from collaborative practices, within and between schools, and from supportive leadership. All of these claims are illustrated by examples from a broad range of national settings and cultures.

Part VI: SYSTEM LEVEL EFFECTIVENESS AND REFORM

Chapter 5: School improvement and systemic reform
David Hopkins (the part on systemic reform)

Hopkins justly calls the several decades of international assessment research “the equivalent of SER sat system level”. Structural differences among educational systems (he compares the USA and Hong Kong) are seen as “given” contextual factors, which underline the complexity of making comparisons. An alternative interpretation is to realize that such structural conditions, like the degree of centralization/decentralization of a country and accountability regimes are “partly” malleable, where “partly” refers to the toughness and long lead time that should be expected in trying to change or adapt these arrangements. He then turns to actual system level improvement strategies and discusses “performance based approaches to educational reform” as studied by Leithwood, Jantzi and Mascall, 1999), the British National Numeracy and Literacy Strategies, and educational reforms in Finland as studied by Hargreaves, Halasz and Pont (2007). The first two approaches booked initial successes, which appeared not to be sustainable in the longer term. Hopkins does not elaborate on the analysis of the Finnish approach. However, it should be noted that the Finnish success story has widely divergent explanations. Shalberg’s (2011) pointing at enhanced autonomy and progressive teaching, while Sahlgren’s (2015) interpretation (which incorporates recent decline in Finnish scores) states that success is due to the traditional top down nature of Finnish educational policy and traditional teaching. From these experiences the author draws up a list of factors that characterize high-performing national and regional systems (p.141). This is to be seen as an original contribution to the debate with a strong “red line” of conditions that are credibly linked to enhancing the primary process of teaching and learning. The chapter finishes with a comprehensive list of key challenges to educational policy-making and research in which the focus on the primary process of teaching and learning is maintained, and in which technical approaches are paired with political, moral and cultural issues.

Chapter 9: Educational effectiveness research in new, emerging and traditional contexts
Sally Thomas, Leonidas Kyriakides and Tony Townsend

This chapter starts out with an introduction which documents the international acceptation of school effectiveness research as a key influence on how educational quality is understood and measured, and the centrality of (broadly defined) student outcomes in this endeavor (221). It then identifies the influence of "context" as a pervasive issue, both in the sense of different school environments, and in terms of different macro-level conditions influencing the way educational effectiveness research is framed. The chapter describes the development of educational effectiveness research in China, the Middle East and Africa. The results on China, though small in scale, given the vastness of the country and within country differences, indicate that effectiveness, monitoring and evaluation approaches receive interest and support, and that studies with comparable sophistication and with comparable outcomes to Western EER practices are possible. In some African countries issues of access to education and the availability of basic material conditions of schooling (buildings, textbooks, and trained teachers) still take precedence over optimization of school organization and teaching, Experiences in Northern Africa and the Middle East bear traces of considerable instability in some countries. Also cultural and religious factors may give rise to philosophies of education and prioritization of educational goals that differ considerably from emphases present in Western educational effectiveness models. In these parts of the world equity issues are considered as of equal or even greater importance than educational quality. The chapter confirms the relevance and applicability of the educational effectiveness research paradigm (including its key concern for student outcomes) in these new contexts, but would hypothesize considerable variation in substantive models, in the sense of input-process-context conditions that account for performance differences. System level conditions that are particularly relevant in these regions are economic prosperity and poverty, political and economic stability (including violent events like revolutions and wars) and cultural tradition.

Chapter 10: Comparative educational research
David Reynolds, Brian Caldwell and others

This chapter notes the tradition of comparative international assessment studies; however on the scene since 1960, but strongly boosted by OECD's PISA since 2000. Methodological limitations of these studies are rightly mentioned (not longitudinal, debatable control for contextual variables, limitations in the measurement of school level process variables). Interestingly, results and interpretations of these studies by OECD's Schleicher and McKinsey are cited without any reservation. The main body of the chapter describes the state of affairs in educational effectiveness research in Latin American countries, and Asia. This is very useful documentary material, but no attempt is made to draw some substantive conclusions from these studies. For example, an attempt at explaining the outstanding results of Cuba, remains somewhat on the surface (relatively high education budget) and makes no connection with systemic alignment versus loose coupling of educational systems as a theoretically relevant dimension.

Chapter 11: Educational effectiveness and improvement research and educational policy. The rise of performance-based reforms
David Reynolds, Anthony Kelly and Christopher Chapman

This chapter is a fascinating account of the development of performance based reforms, particularly in the UK, but with reference to other national contexts as well.
Starting out from a situation of local autonomy and relatively limited state intervention in the 1950's and 60's, in the 1970's various factors lead to a growing concern over educational outcomes being sufficiently served by progressive "pupil-centered" teaching methods, and a call for larger government involvement. This gradually led to what the authors describe as the "supply/demand side" performance paradigm. The supply side characterized by an emphasis on performance outcomes, developments in the realm of external school inspection and league tables, and the demand side catering for parent demands and school choice. The labor government of 1997 stimulated school performance being measured and published as "value-added" , increased the variety of school types (to enlarge choice options for parents) and initiated two "prescriptive" programs, known as the Literacy and Numeracy Strategies (p.286). Although these programs were positively evaluated and demonstrated achievement gains in both subject matter areas, the overall assessment became more nuanced and even showed disappointment as the outcomes did not appear to be sustainable over time. According to the authors this state of cognitive dissonance lead to increased attention on building the capacity of the teaching profession, and attention for potential conflict between externally induced accountability policies and professional motivation. Interestingly the authors conclude that the new emphasis on professionalism did not lead to abolishing the established performance-based approaches "in any countries" (289) and they see the next stage of educational reform as a continuation of performance based strategies, accompanied by measures to enhance the professional capacity of school staff. The chapter continues with a discussion of evidence of the effectiveness of performance based reform, which is assessed as moderately positive, and finishes up with recent developments. Among these, internal evaluation and formative assessment at school level are seen as potentials for school based improved. Attention is also drawn to important contextual developments, such as the information society, challenges for value education, as the role of the church in western societies has diminished, and possible implications of globalization such as an increased role of the profit sector in various educational areas. Finally two case studies from the UK illustrate an additional dimension in the current shaping of systemic educational reforms, namely diversification of school types, increased space for local initiatives with new meso-level governmental arrangements and support structures, such as networks and between school cooperation. In other words: major dynamics in structural adaptations and organizational development. But perhaps it is here that new educational reform is at risk of flying out of control. Not only is the evidence of the effectiveness of cooperation and organization development missing (Muijs, 2016, Scheerens, 2016) but the total absence of assessing costs poses an even more nagging question about efficiency.

Chapter 12: Educational effectiveness research and system reconstruction and change
Sam Stringfield and Anthony Mackay

This chapter provides a panoramic overview of the development of educational effectiveness and improvement over five decades, starting with the Coleman report (1966). The perspective in this chapter is more global and less UK focused than the previous chapter. They discern a gradual upscaling in research focus, from teaching effectiveness, to school and district effectiveness to national system

effectiveness. Each of these unit levels are indicated as complex systems, with central control versus loose coupling of sub-units as a major facet of this complexity. Higher level initiative and control of educational reform has a central place in the outlook of the authors, also in the face of functional decentralization and increased school autonomy in many systems. In this way they are prompted to see implementation problems as a major source of ineffectiveness. Strength of the empirical knowledge base differs with aggregation level in quantity and quality: strongest on teaching effectiveness, on a fair level of consensus with respect to school level, and much more uncertain as far as the national system level is concerned. On this latter point they provide nuanced commentary on recent reports by OECD and McKinsey, based on PISA results. Even if we had robust findings, the authors conclude, we still would not be sure of policy implementation of these results. And to this could be added that "underutilization" is just one side of this problem, the other being "false positives", in the form of policy measures that are sold as effective, without having been proven to be so convincingly.

Chapter 15: The challenges of globalization and the new policy paradigm for educational effectiveness and improvement research
Anthony Kelly and Paul Clarke

In this chapter globalization, neo-liberalism, social democracy , "performativity" , professionality and concern for sustainability are described as inputs to national and local educational policy. Globalization is expected to strengthen the role of market forces accompanied by state accountability policies. Together these developments are expected to further stimulate "performativity", which in its turn is understood as being at odds with professional interests of schools and teachers. In this chapter EEIR is not addressed as an academic discipline, but as a political position. Current EEIR is squarely put down as an instrument of the government. The desired alternative is presented in the last sentence of the chapter, which describes the challenge of EEIR as "to identify why, and measure how, teachers and head teachers resist certain aspects of … unfavorable agendas … (p.378), where these unfavorable agendas are shaped by a governmental focus on competition and performance, partly as a reaction to globalization. In other words EEIR should support teachers and head teachers and not the government. Trivial as it sounds but this is what the chapter has to offer as an alternative against the accursed "performativity", defined as the "positivist assumption that it is possible and always desirable to measure performance". There is repeated mentioning of EEIR needing to change its metrics; but it is never explained what this would actually mean, and the promised alternative, the "new policy paradigm for educational effectiveness" remains obscure. Given the common understanding of the effectiveness concept across most chapters of the Handbook, one could see what would remain if "performance measurement" were to lose its central position: just a standpoint in a political debate and goal displacement.

Part 7: CONCLUSION

Chapter 17: Conclusions
David Reynolds, Christopher Chapman, Paul Clarke, Daniel Muijs, Pam Sammons and Charles Teddlie

This final chapter starts out with a brief summary of major developments in the field; also when comparing to the contents of the previous, year 2000 version of the Handbook. They note methodological development, fuller integration of classroom level teaching facets, growth in addressing system level effectiveness, progress in theory development, expansion of international applications and an improved ("matured") treatment of school improvement. About the core knowledge attained on effectiveness enhancing factors at school and classroom level the authors conclude that "the major gains in our knowledge have been more about the nature of educational effectiveness… than about the courses of it". And further on they note "a conventional feel to the findings" as far as results on the school factors, at least in Anglo-Saxon countries, are concerned. The rest of the chapter discusses challenges and areas of interest for further development:

"*Fully embracing and understanding complexity*" (complexity across cultural, economic and political boundaries; connecting schools with local contexts and including a broader scope of outcomes);

"*Enhanced take up by practitioners and policy makers*" On this issue the authors think that the educational effectiveness knowledge base is underutilized, and as one of the reasons for this they mention small effects and results that are "inconvenient" to policy makers. They discuss a range of ways to improve visibility and communication with practitioners and policy-makers. The most far going measure being participatory work by researchers, policy makers and practitioners: " a field that naturally and seamlessly embraces and engages practitioners and policy makers in the core of its work" (p. 419).

"*Focusing on what could be in addition to what is*". Here different roles for effectiveness research on the one hand and school improvement on the other are recommended. Effectiveness research should take a critical stance and concentrate on evaluation of innovatory programs. With respect to improvement educational effectiveness experts could be directly involved or even lead development of improvement programs.

"*Becoming efficient as well as effective".* More interest in cost-benefit and cost-effectiveness questions is recommended and reference is made to economists' work on educational production functions.

"*The need to develop ecological relevant orientations"*. This is a section that propagates more attention to ecological issues in education across the globe. Sympathetic as this call is, it is not clear what the implications are for the furthering of educational effectiveness research and evidence based school improvement.

"*The need to study context specificity*". The undertone of this section is that educational effectiveness research has not done enough to take context specificity into consideration. However, as a scientific research discipline, educational effectiveness research is primarily interested in generalization and in this section this is almost described as an impairment. Next, a more systematic treatment on different ways in which "context" is to be defined in educational effectiveness research is missed (For example by distinguishing controlling for student background conditions, differential effectiveness across sub-groups of students, composition effects, local environmental conditions, national structural and cultural conditions, ingenious cultures, cultural traditions across the globe, and more generic context conditions as established from contingency theory).

"Researching in novel areas of education". The discourse is concentrated on applications in Higher Education, and focused on genetic influences on educational attainment. This is a theme that is of a

more general importance to educational effectiveness research, than just relevant to research in Higher Education (cf. various contributions by Gary Marks; e.g. Marks, 2015)
"Drawing on new external perspectives". External perspectives are illustrated by comparing traditions in accountability orientations between countries, particularly between England and Scotland. The issue is the influence of such school external perspectives on partnership between researchers, policy makers and educational practitioners, which sometimes also involve new brokerage roles.
The closing message of this concluding chapter is a call for "more integrated and flexible solutions", with a renewed focus on professional development, "a commitment to ownership over what works and why" and "a dedication to joined-up public service provision" as arrow heads. It is a call for academics to leave their ivory towers and to join networks for "collaborative inquiry-driven approaches" specifically aimed at improvement of conditions for marginalized and disadvantaged groups.

Discussion

Handbooks and state of the art reviews in a research discipline and area of application have a relevant function in organizing the field, taking stock of achievements and developments, and positioning itself with respect to relevant environments. The current Handbook serves these functions rather well. An added advantage is that a previous version of the Handbook can be used as one of the baselines for changed emphases and progress. System level reform, association with international comparative assessment studies, and a "mature" (to use the expression used in the Handbook) treatment of school improvement studies have received relatively more emphasis in the current Handbook.
When one considers the Handbook as an (however advanced) introduction to the field, it does well in defining basic concepts and maintaining consistency in the way the EEIR paradigm is manifested in most of the chapters. Although challenged in a few chapters on qualitative and "critical" approaches the large majority of the contributions remains true to a focus on measurable outcomes and an in-partial role of researchers. To the extent that upholding this focus might be more difficult for school improvement than for educational effectiveness research, chapters 5 and 14 point at viable solutions. These solutions are varied applications of evaluation research, including formative assessment and data feedback.
Despite this overly favorable impression there are a few issues that are debatable, in the sense of coverage and treatment. These refer to the *demarcation of the field, coverage and "blank spaces", fuzzy borders between research and action, critical questions about the knowledge base,* and how *complexity* and *theory* are dealt with.

The demarcation of "the field"
What is described as the field of educational effectiveness and improvement research in this Handbook should be seen as a subdomain of a larger area of outcome related educational research. Key markers of this larger field are that it takes place in educational settings, is outcome related (has student achievement and attainment as dependent variables) and looks for effective malleable conditions. The field as represented in this Handbook has an inter-disciplinary orientation, has ICSEI and the EARLI-

special interest group on educational effectiveness[1] as its main professional organizations and conferences, and *School Effectiveness and School Improvement* as its most important journal. It should be noted that both professional organizations and the journal have a clear international scope. The most visible segment of the larger area of outcome related educational research that has minimal coverage in the Handbook is the research in the US based SREE[2] network, which has strong input from economists and econometricians. This selectivity also implies an underrepresentation of recent US based research in the Handbook. Similarly, it is my impression that outcome oriented research on learning and instruction, with psychology as its disciplinary base, is a broad field that is only partially covered in the Handbook.

Coverage and "blank spaces"

Substantively, two important areas receive superficial treatment in the Handbook: teacher effectiveness and curriculum alignment. Even though the Handbook has a chapter on teacher/teaching effectiveness (Ch. 7),this is practically all about *teaching* behavior and does not address *teacher* effectiveness in the sense of relatively stable characteristics of teachers, such as level of initial training, experience and knowledge (e.g. the important concept of pedagogical content knowledge). Finally the curriculum, and related operationalization such as "opportunity to learn", "content covered", and instructional alignment are hardly mentioned in the Handbook; where meta-analyses and recent international studies indicate non-negligible effect sizes in this area (Scheerens, 2017).

Fuzzy borders between research and action

In a value laden context like education a careful and nuanced approach is required to demarcate the researcher's role. It is beyond the scope of this review to go into this in any detail. Yet, when effectiveness researchers are invited to side with teaching professionals in their protest against the government (chapter 15), and to embark on unspecified partnerships with educational practitioners to get involved with "boutique inquiry approaches" (chapter 8) it becomes questionable whether we are still within the realm of scientific research. Equally border crossing are suggestions in chapter 16 to take inspiration from critical social theory and abandon a value free stance in nothing less than a "paradigm shift". And, finally, sympathetic as it may be, the propagation of "ecologically relevant orientations" (chapters 8 and 17) does not belong in a Handbook on educational effectiveness and improvement *research*.

Critical questions about the knowledge base

The overall outlook in the Handbook is that educational effectiveness and improvement research has now a fairly established knowledge base. And it should be emphasize that the attempt to synthesize this knowledge is to be praised as a real contribution to both researchers and educational practitioners. The position of an established knowledge base is supported by the relatively strong consensus about the set of factors at school and teaching level that are considered to work, and are described in chapters 3 and 4. But there is much less consensus of effect sizes as established in meta-analyses. Chapter 4 refers to

[1] (ICSEI) International Conference for School Effectiveness and Improvement; (EARLI) European Association for Research on Learning and Instruction

[2] SREE Society for Research on Educational Effectiveness

the fact that Hattie (2008) reports much higher effect sizes than Seidel and Shavelson (2007). Moreover, important variation in estimated average effect sizes between meta-analyses covering the same factors is more of a general pattern (Scheerens, 2014). In addition, results of some of these meta-analyses show very small average effects for highly reputed factors as leadership and teacher cooperation. Divergence across meta-analyses and uncertainty of the meaning of average effect sizes provokes a **range of fundamental debates, of both a methodological and substantive nature:**

- Implications of the fact that a large share of studies in this field are correlational and not experimental
- Implications of the persistent lack of standardized instruments (with some favorable exceptions, rightly mentioned in Chapter 8)
- Questions about the methodology of the meta-analyses; including pressures to inflate treatment effects, as in publication bias and study characteristics that influence effect sizes (Cheung and Slavin, 2016)
- Recent work showing that context effects (student background characteristics in particular) explain more and malleable factors less of the total student level variance (e.g. Marks, 2015)
- More in depth analyses of "ineffectiveness" in the sense of explaining small effect sizes (Scheerens, 2016).

Perhaps the field is less "arrived" as the Handbook appears to suggest and, among others, this might call for a more prudent attitude with respect to the utilization issue.

Methodology: controlling or embracing complexity?

The state of the art methodology of educational effectiveness research, also involving further developed multi-level and structural equation modeling, as discussed in Chapter 8 of the Handbook, are just about fit to cover basic conceptual models, which include modification and mediation. Yet major challenges occur when contextual conditions, differential effects for sub-units and treatment heterogeneity are to be taken into consideration (e.g. Raudenbush, 2011, Hedges, 2016). The evidence based movement in the US and the emphasis on randomized experiments, as compared to the traditionally more common "correlational" studies, stimulates reflection on the kind of treatments in EER. In correlational studies specific effectiveness enhancing factors are included in attempts to model variance in schooling. This has often been indicated with the metaphor of "opening black boxes" and has yielded the well-known lists of effectiveness enhancing factors at school and classroom level. Judging from the contributions to the annual SREE conferences, experimental designs are often applied to "black box treatments", like specific programs, or instructional packages. From a practical or even commercial perspective information on whether a specific package works is useful, but from a scientific point of view this would only be the case if the program had a specific rationale and would be replicable. Fascinatingly creative research methodological solutions are being sought for controlling the complexity of educational effectiveness research but more work on theory and explanatory mechanisms (see the subsequent section) are important as well.

Rather than controlling, the more qualitative oriented contributors to the Handbook would like to "fully embrace" complexity. In one strike they seek to solve the utilization issue, by emphasizing participation, co-construction and ownership by practitioners. This is what the Handbook sometimes announces as a paradigm shift. Maybe there is a border area between research and professional work of teachers

involved in “reality checks” like formative assessment and data-use, as described in Chapter 14, and which is a very interesting area to be further developed. But the rather vague hints in the Handbook calling for a shift in the scientific paradigm of educational effectiveness research are not convincing. The participatory approach seems a pipe dream, not able to support causal interpretation, prone to bias and epistemologically stranding in “strong empiricism” (in the sense that it is impossible to describe everything to the smallest detail as the phrase ‘embracing complexity’ suggests).

A note on theory

“Research in education should support the development of explanatory and predictive theories of educational processes and mechanisms. Education research must answer questions about why, how, under what circumstances, and for whom, education practices and policies affect individual outcomes. Without an evidence-based theory of educational processes and mechanisms, pragmatic evidence of effectiveness may not be generalizable to new settings or different populations” (Citation from the introductory text of the program of the 2011 SREE meeting https://www.sree.org/conferences/2011/). I found the notion that theory is about applying explanatory mechanisms to make sense of empirical facts and drive research not strongly represented in the Handbook. This is very much related with the ambition to obtain generalizable knowledge. If the phenomena of educational effectiveness can be understood from more general principles, embedded in established theory, key features provide parsimonious handles for generating research and dissemination of results. Examples, next to the Dynamic Model, already referred to, are understanding the model of High Reliability Organizations as conforming to principles of the bureaucracy and seeing data-use and feedback procedures in terms of retro-active planning models.

All this being said the Handbook is a very useful report on the state of the art of educational effectiveness and improvement research and, as such, a key reference for future work in this domain by educational researchers and other education professionals.

References

Chapman, C., Muijs, D., Reynolds, D., Sammons, P., and Teddlie, C. (2016) *The Routledge international handbook of educational effectiveness and improvement: research, policy and practice.* London and New York: Routledge

Cheung, A.C.K. and Slavin, R.E. (2016) How methodological features affect effect sizes in education. *Educational Researcher* 45, (5) 283 -292

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

Edmonds, R. (1979) Effective schools for the urban poor. *Educational Leadership* 37 (1), 112-129

Hargreaves, A., Halasz, G., and Pont, B. (2007) *School leadership for systemic improvement in Finland.*Paris: OECD

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge

Hedges, L.V., (2016) Challenges in Building Usable Knowledge in Education. Opening Address SREE conference. Washington D.C. https://www.sree.org/conferences/2016s/pages/video.php

Lai, L.K. and Schildkamp, K. (2013) Data-based decision making in education. *School Leadership and Management,* 17, 9 -21

Leithwood, K. , Jantzi, D. and Mascall, B. (1999)Large-scale reform: what works? Ontario Institute for Studies in Education. University of Toronto.

Marks, G. N. (2015). The size, stability, and consistency of school effects: Evidence from Victoria. *School Effectiveness and School Improvement*, 26(3), 397–414.

Muijs, D. (2016) When does(n t) collaboration work? Evidence on school-to-school collaboration as as strategy for school improvement. Key-note address at he EARLI SIGs 16&23 conference in Oslo, 29- 30 September.

Raudensbush, S. (2011) Modeling Mediation: Causes, Markers, and Mechanisms. Opening Address SREE conference. Wahington D.C. https://www.sree.org/conferences/2011/video/

Sahlberg, P., (2011) *Finnish lessons.* New York: Teachers College Press.

Salhlgren, G.H. (2015) *Real Finnish lessons. The true story of an education super power.* Center for Policy Studies. Surrey

Scheerens, J. (2013) Lo scenario politico della valutazione educativa e il ruolo dei metodi quantitativi e qualitativi. 3rd International Symposium on Educational Research: New Realism and Educational Research. Universita degli Studi Roma 3. Rome, 7 and 8 June

Scheerens, J. (2014) *What is effective schooling? A review of current thought and practice.* Report prepared for the International Baccalaureate Organization. http://www.ibo.org/globalassets/publications/ib-research/continuum/what-is-effective-schooling-report-en.pdf

Scheerens, J. (2016) *Educational Effectiveness and Ineffectiveness. A critical review of the knowledge base.* Dordrecht, Heidelberg, New York, London: Springer

Scheerens, J. (2017) O*pportunity to learn, curriculum alignment and test preparation. A research review.* Dordrecht, Heidelberg, New York, London: Springer

Seidel, T., & Shavelson, R.J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research 77*(4), 454-499.

Teddlie, C., and Reynolds, D. (2000) *International handbook of school effectiveness research.* London: Falmer Press