The effects of school-based decision making on educational outcomes in low and middle income contexts: a systematic review

Roy Carr-Hill, UCL Institute of Education

Caine Rolleston, UCL Institute of Education

Tejendra Pherali, UCL Institute of Education

Rebecca Schendel, UCL Institute of Education

with

Edwina Peart, UCL Institute of Education

Emma Jones, UCL Institute of Education

Grantee Final Review

Accepted by 3ie: September 2015

Note to readers

This systematic review has not been funded but has been quality assured by 3ie. We are making it available to the public in this final review version.

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the systematic review was accepted. Any comments or queries should be directed to the corresponding author, Roy Carl-Hill, roy.carr_hill@yahoo.com

Suggested citation: Carr-Hill, R, Rolleston, C, Pherali, T and Schendel, R, 2015. *The effects of school-based decision making on educational outcomes in low and middle income contexts: a systematic review, 3ie Grantee Final Review.* London: International Initiative for Impact Evaluation (3ie)

Table of Contents

Plain language summary

Motivation: Although the rhetoric around decentralisation suggests that school-based management has a positive effect on educational outcomes, there is limited evidence from low income countries of this general relationship. Existing reviews on school-based decision-making have tended to focus on proximal outcomes and offer very little information about why school-based decision-making has positive or negative effects in different circumstances. This review aimed to address these gaps by answering the following questions: (1) What is the impact of school-based decision-making on educational outcomes in low- and middle-income countries (LMICs)? (2) What are the barriers to (and enablers of) effective models of school-based decision-making?

Approach*:* In order to answer these questions, we conducted a mixed-methods systematic review. All included studies had to: 1) be empirical; 2) focus on primary and secondary schools within LMICs; 3) investigate a *change in decision-making authority from a higher level of decision-making authority to the level of the school;* 4) provide data on *at least one educational outcome* (either proximal or final); and 5) rely on data collected since 1990. Studies included in reference to the first review question needed to include an appropriate counterfactual, but studies of various empirical designs were included in reference to the second. After a comprehensive search process, we identified twenty six impact studies that met the inclusion criteria. We used meta-analysis to report on the impact of school-based decision-making reforms on six educational outcomes: 1) student drop-out; 2) student repetition; 3) teacher attendance; and 4) student learning, as assessed via i) language test scores, ii) math test scores, iii) aggregate test scores (i.e. tests of more than one subject). We also examined potential sources of heterogeneous impacts across studies using moderator analysis focusing on level of decentralisation, country income group and study methodology. Finally, we conducted narrative synthesis of factors that appear to enable or hinder effective school-based decision-making.

Findings: Devolving decision-making to the level of the school appears to have a somewhat beneficial effect on drop-out in some contexts and on repetition when looking across studies. Effects on test-scores are more robust, being positive in the aggregate, and in analysis of middle income countries. Effects on teacher attendance are stronger in contexts of high decentralisation and of low income. School-based decision-making reforms appear to be less effective in communities with generally low levels of education where parents have low status relative to school personnel.

Implications: School-based decision-making reforms are less likely to be successful in highly disadvantaged communities. The involvement of school management committees in personnel decisions can play a role in improving teacher attendance, but this may not always translate into impacts on learning. Identification of effects on student learning may take longer than anticipated by evaluation timelines.

Executive Summary

Background

Although there have been significant improvements in recent debates, access to education - particularly for girls, poor children and children in conflict-affected areas – remains limited. There is also worrying evidence that many children enrolled in school are not learning. Recent estimates suggest that around 130 million children who have completed at least four years of school still cannot read, write or perform basic calculations (UNESCO, 2014, p. 191).

Many governments have attempted to address this situation, while also improving efficiency and reducing costs, by devolving decision-making authority to schools, as it is assumed that locating decision-making authority within schools will increase accountability, efficiency and responsiveness to local needs (Gertler et al., 2008). This devolution includes a wide variety of models and mechanisms, differing in terms of which decisions are devolved (and how many), to whom decision-making authority is given, and how the decentralisation process is implemented (i.e., through 'top-down' or 'bottom-up' processes). All models and mechanisms are presumed to increase responsiveness to local needs and accountability by bringing community members into direct contact with schools, and to increase efficiency by making financial decisions more transparent to communities, reducing corruption and incentivising investment in high quality teachers and materials.

Objectives

Although the rhetoric around decentralisation suggests that school-based management has a positive effect on educational outcomes, there is limited evidence from low income countries of this general relationship. Existing reviews on school-based decision-making have tended to focus on proximal outcomes (e.g. Guerrero et al., 2012, on teacher absenteeism), with very few considering the full range of relevant outcomes, including student learning. The more comprehensive reviews that do exist (Santibanez, 2007 and World Bank, 2007) are not formal systematic reviews, according to the criteria set by the Campbell Collaboration. They also need updating, as they (a) rely on literature that is now nearly ten years out of date and (b) focus almost exclusively on Central America, referencing almost no evidence from other low or middle income countries. Existing reviews on this topic also tell us very little about why school-based decision-making has positive or negative effects in different circumstances.

This review aims to address these gaps by answering the following questions: (1) What is the impact of school-based decision-making on educational outcomes in low- and middle-

income countries (LMICs) (Review Question 1)? (2) What are the barriers to (and enablers of) effective models of school-based decision-making (Review Question 2)?

For the purposes of the review, 'school-based decision making' was defined as any reform in which decision-making authority has been devolved to the level of the school. Within this broad definition, there are three main mechanisms discussed in the literature: (1) reforms that devolve decision-making around management to the school level; (2) reforms that devolve decision-making around funding to the school level; and (3) reforms that devolve decision-making around curriculum, pedagogy and other aspects of the classroom environment to the school level.

Methods

This review followed an explicit protocol following methodological guidance provided by the Campbell Collaboration and the EPPI-Centre at the UCL Institute of Education (Becker et al., undated; Gough et al., 2012; Hammerstrom, 2009; Shadish & Myers, 2004).

To be included in the review, all studies had to: 1) Be empirical in nature and focused on primary and secondary schools within LMICs; 2) Investigate a *change in decision-making authority from a higher level of decision-making authority to the level of the school* (excluding studies where the intervention was conceptualised, managed and implemented by an external decision-making agency, or aimed exclusively at improving the functioning of existing devolved decision-making structures); 3) Provide data on the relationship between school-based decision-making and *at least one educational outcome* (either proximal, e.g. attrition, equality of access, increased enrolment; or final, e.g. student learning, as captured by test scores, psychosocial and non-cognitive skills, etc.); and 4) Rely on data collected since 1990.

In order to be included in reference to Review Question 1, studies needed to be causal in nature, meaning we included: (1) Experimental designs using randomised or quasi-randomised assignment; (2) Quasi-experimental designs; and (3) Before-and-after studies which collect longitudinal data at baseline and endline, as well as those using cross-sectional endline data only, provided data are collected from a comparison group or where an appropriate method of analysis has been used to control for confounding. For Review Question 2, we included studies of any empirical design, so long as they provided additional data relating to those interventions featuring in the impact component of the synthesis.

Potentially relevant literature was identified through a five-stage search strategy, which comprised: 1) Identification of existing systematic reviews in related areas; 2) Targeted searches in a wide range of bibliographic databases and websites; 3) Hand searches of the eight most relevant journals relating to the topic; 4) Citation chasing; and 5) Contacting experts involved in the research area. A comprehensive list of search terms was developed in collaboration with information scientists at the EPPI-Centre. Search terms were also translated into French, Spanish and Portuguese for use in regionally-specific databases. All

identified literature was subjected to a two-stage screening process. Relevant studies were then appraised for robustness of evidence and methodological rigour prior to synthesis.

We identified 2,821 titles through our five-stage search. Of these, 100 met our eligibility criteria. Thirty of the 100 met the design criteria required for RQ1, but three were removed from the RQ1 synthesis, due to high risk of bias. A fourth study had to be excluded due to missing data. Twenty-six impact studies were, therefore included in the meta-analysis. These 26 studies investigate the impact of 17 individual interventions. Of the 73 non-causal studies subjected to quality appraisal, nine were identified which provided additional data on the included interventions.

In order to answer Review Question 1, we conducted meta-analysis, relying on the use of 'standardised mean difference' (SMD) calculations to compare effects across studies. In our meta-analysis, we were able to report on the impact of any school-based decision-making reform on six educational outcomes: 1) student drop-out; 2) student repetition; 3) teacher attendance; and 4) student learning, as assessed via i) language test scores, ii) math test scores, iii) aggregate test scores (i.e. tests of more than one subject). We also examined heterogeneity by investigating differences in impacts based on three moderating variables – level of decentralisation, income level, and type of evaluation design. Further, we discuss and synthesise sub-group effects discussed in the included studies themselves. Analysis in reference to Review Question 2 followed the principles of framework synthesis (Thomas et al., 2012), in order to identify the main barriers and enablers that appear to have influenced the impact of the interventions under review.

Results

Devolving decision-making to the level of the school is found to have a somewhat beneficial effect on drop-out – a pooled effect of reducing drop-out by 0.07 standard deviations (SDs). For repetition, the equivalent pooled effect is a reduction of 0.09 SDs. Effects on test-scores are larger and more robust. We find a positive and significant improvement of 0.21 SDs in aggregate test scores on average, and positive and significant improvements of around 0.07 SDs in scores on language and 0.08 on maths tests. Further analysis of test score results suggests that these results pertain to middle income countries, while we did to find statistically significant improvements in test scores in low income country settings, with the exception of one study in Kenya (now a middle income country). Evidence does not show that effects on teacher attendance are significant overall, but there is evidence that effects are stronger in contexts of high decentralisation.

Most of the included studies do not conduct any sub-group analysis relating to individual characteristics, such as gender and student background; those that do differ in their findings. However, there is some evidence to suggest that school-based decision-making reforms have a stronger impact on wealthier students with more educated parents. It appears that school-management reforms may be particularly impactful on children in younger grade levels.

School-based decision-making reforms appear to be less effective in disadvantaged communities, particularly if parents and community members have low levels of education and low status relative to school personnel. Devolution also appears to be ineffective when communities choose not to actively participate in decision-making processes. Small schools, however, may find school-based decision-making to be effective, particularly if community members establish a collaborative, rather than an adversarial, relationship with teachers.

Conclusions and implications for policy, practice and research

Overall, we can conclude that devolving decision-making authority to the school level can have a positive impact on educational outcomes, but that this is only likely in more advantaged contexts in which community members are largely literate and have sufficient status to participate as equals in the decision-making process.

Our findings carry a number of implications for policy and practice. First, it appears that school-based decision-making reforms in highly disadvantaged communities are less likely to be successful. Parental participation seems to be the key to the success of such reforms and this is linked to the real authority or status and cultural capital of community members. Second, the involvement of school management committees in personnel decisions appears to play a role in improving proximal outcomes, such as teacher attendance, but success is also likely to be linked to the overall teacher job market and the prospects of long-term employment. Third, the specifics of programme design appear to be crucial. Given the limited evidence available in this review, and the contextualised nature of that evidence, we cannot conclude with certainty that incorporating certain elements into school-based management reforms are generally beneficial. However, it does appear that the details of such supplementary elements may be important. The evidence also suggests that, at least in some contexts, impact on student learning may take longer than is often allowed within evaluation timelines. Where donors are involved, this also means that decentralisation reforms may require sustained donor commitment over the long term.

The review also suggests a number of fruitful directions for future research. Although a large number of titles were identified during our initial search, the small number of impact studies included in the meta-analysis represent limited geographic diversity and a small number of discrete interventions. There needs to be further robust analysis of the impact(s) of large-scale school-based decision-making reforms that have recently been implemented, as well as further analysis of the conditions that mitigate their impact. There is also a clear need to examine the potentially negative impacts of these reforms, given widespread adoption of such policies. Although this review has highlighted a number of potential enablers and barriers, the limited evidence base has prevented us from drawing any robust conclusions on the conditions necessary for positive impact. A future review of the same topic, drawing on broader qualitative evidence, would complement the findings of this study.

1. Background

1.1. DESCRIPTION OF THE PROBLEM

Education is internationally understood to be a fundamental human right that offers individuals the opportunity to live healthy and meaningful lives. Evidence from around the world also indicates that education is vital for economic and social development, as it contributes to economic growth and poverty reduction, sustains health and well-being, and lays the foundations for open and cohesive societies (UNESCO, 2014). In recognition of the vital importance of education, governments across the globe have made a substantial effort to expand and improve their education systems, as they strive to meet the Education for All goals, adopted by the international community in 1990. These efforts have borne remarkable results; it is estimated that the number of out-of-school children has halved over the last decade (ibid, p. 53). However, there are still serious barriers to overcome, particularly in terms of access, completion and learning (Krishnaratne et al., 2013). Access to education - particularly for girls, poor children and children in conflict-affected areas - remains a crucial issue. The 2013 Global Monitoring Reports claims that an estimated 57 million children are still out of school, over half of whom are in sub-Saharan Africa (UNESCO, 2014, p.53).[1] Furthermore, despite increases in enrolment numbers, there has been almost no change since 1999 in the percentage of students dropping out before the end of the primary cycle. The evidence also indicates that many children enrolled in school are not learning. Recent estimates suggest that around 130 million children who have completed at least four years of school still cannot read, write or perform basic calculations (UNESCO, 2014, p. 191).

1.2. DESCRIPTION OF THE INTERVENTION

Many governments have attempted to address this worrying situation, while also improving efficiency and reducing costs within the education sector, by decentralising decision-making processes. Decisions about curricula, finance, management, and teachers can all be taken at one or more of several administrative levels: centrally at the national or federal state level, by provinces/regions within a country, by districts or by schools. The devolution of decision-making authority to schools has been widely adopted as the preferred model by many

[1] Carr-Hill (2012) suggests that, because most of the estimates for low income countries are based on household surveys, this figure should actually be doubled. Household surveys omit the homeless by design, thereby excluding mobile, nomadic, or pastoralist populations. Moreover, in practice, household surveys typically under-represent those in fragile, disjointed households, slum populations and those in conflict-affected areas posing security risks.

international agencies, including the World Bank, the U.S. Agency for International Development (USAID) and the UK Department for International Development (DFID), as it is assumed that locating decision-making authority within schools will increase accountability, efficiency and responsiveness to local needs (Gertler et al., 2008). Often described as 'school-based' or 'community based' management, the devolution of decision-making authority to schools includes a wide variety of models and mechanisms. These differ in terms of which decisions are devolved (and how many), to whom decision-making authority is given, and how the decentralisation process is implemented (i.e., through 'top-down' or 'bottom-up' processes). School-based decision-making can be used to describe models in which decisions are taken by an individual principal or head teacher, by a professional management committee within a school, or by a management committee involving local community members. This last model may simply imply an increased role for parents in the management and activities of the school, or it may result in more active provision of training and materials to empower broader community involvement (Krishnaratne et al., 2013).

The devolved decisions can be financial (e.g. decisions about how resources should be allocated within a school; decisions about raising funds for particular activities within a school; etc.), managerial (e.g. human resource decisions, such as the monitoring of teacher performance and the power to hire and fire teachers; decisions relating to the management of school buildings and other infrastructure; etc.) or related to the curriculum and/or pedagogy (e.g. decisions related to the articulation of a school's curriculum; decisions about how elements of a national curriculum will be taught and assessed within a given school; etc.). In order to support the process of decision-making, many models also involve some means of providing information to community members on the performance of an individual school (or school district) relative to other schools (Barrera-Osorio & Linden, 2009). All of these models and mechanisms are considered to potentially increase accountability and responsiveness to local needs by bringing local community members into more direct contact with schools, and to increase efficiency by making financial decisions more transparent to communities, thereby reducing corruption and incentivising investment in high quality teachers and materials.

For the purposes of this review, 'school-based decision-making' has been defined as including any model in which at least some of the responsibility for making decisions about planning, management and/or the raising or allocation of resources is located within schools and their proximal institutions (e.g. community organisations), as opposed to government authorities at the central, regional or district level. The 'intervention' considered within this review, therefore, is *any reform in which decision-making authority is devolved to the level of the school.* Within this broad definition, there are three main mechanisms discussed in the literature: (1) reforms that devolve decision-making around *management* to the school level; (2) reforms that devolve decision-making around *funding* to the school level; and (3) reforms that devolve decision-making around c*urriculum, pedagogy and other aspects of the classroom environment* to the school level.

1.3. HOW THE INTERVENTION MIGHT WORK

School-based decision-making is widely promoted by donors in lower-income countries as a means for improving educational quality and is often taken up enthusiastically by national governments. Both generally articulate the ultimate outcome of school-based decision-making models as being a positive change in student outcomes (including but not restricted to learning outcomes). In addition to learning outcomes (most often measured through standardised tests of cognitive skills), there are many other possible student learning outcomes which may be valued by schools, donors and governments, such as improved student ability to demonstrate psychosocial and 'non-cognitive' skills. Changes in student aspirations, attitudes (such as increased appreciation of diverse perspectives) and behaviours (such as the adoption of safe sex practices) could also be considered important educational outcomes.

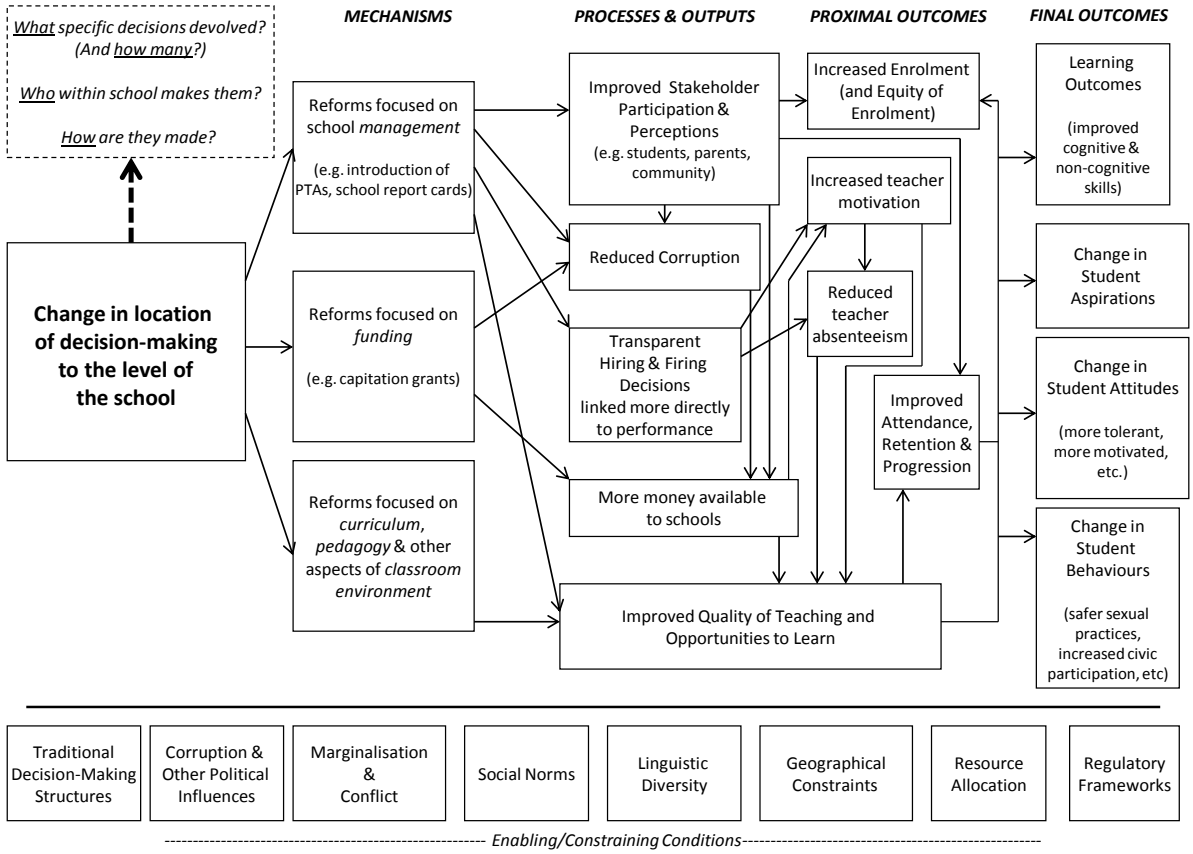
However, it is clear that devolving decision-making to the level of the school does not lead *directly* to such outcomes. Rather, school-based decision-making is likely to impact on outcomes via a number of causal pathways. Reforms that increase accountability and responsiveness to local needs are assumed to lead to positive stakeholder perceptions of (and engagement in) educational provision, which, in turn, is expected to increase enrolment, attendance and retention and to reduce corruption within schools. It is also presumed that increased accountability will encourage schools to make recruitment decisions on the basis of teacher performance, rather than mechanically relying on qualifications or allowing for nepotism to interfere. Such personnel practices, in turn, are seen to lead to reduced teacher absenteeism, increased teacher motivation and, ultimately, improvements in the quality of teaching within schools. It is also assumed that local communities will encourage schools to adopt more locally relevant curricula, which can then have a positive impact on the quality of teaching and student opportunities to learn. At the same time, decentralised funding mechanisms and other reforms aimed at increasing efficiency within schools, particularly when combined with efforts to increase community participation, are presumed to result in more resources being available to schools, another important factor in improving educational quality (Krishnaratne et al., 2013). Increased efficiency is, in turn, assumed to affect the unit costs of educational provision, potentially reducing costs or improving outcomes for a given cost, which may be particularly valued by governments in less well-resourced settings. School-based decision-making mechanisms, therefore, result in a number of proximal (or intermediate) outcomes, in addition to the final outcomes mentioned above. These proximal outcomes include increased enrolment, improved equality of access, improved attendance, improved retention, improved progression, and higher quality educational provision.

However, there is growing evidence that decentralisation reforms may actually have unintended and sometimes negative effects in certain political and economic circumstances (Banerjee et al., 2008; Bardhan & Mookherjee, 2000, 2005; Carr-Hill et al., 1999; Condy, 1998; Glassman et al., 2007; Pherali et al., 2011; Rocha Menocal & Sharma, 2008; Rose, 2003; Unterhalter, 2012). Decentralising decision-making may lead to elite capture at the

local level and/or further corruption within school systems, for example, or may limit educational opportunity for marginalised ethnic groups. There is some consensus in this literature that decentralisation is only likely to have a positive impact on outcomes when (a) there is clear government policy and/or regulations about the powers and role played by different agencies and stakeholders; (b) there are sufficient financial resources available within the system; and (c) there is some form of democratic culture (see De Grauwe et al., 2005; Lugaz et al., 2010; Pherali et al., 2011). Those vested with the authority to make decisions on behalf of the school must also have the capacity and knowledge to make such decisions, or their decisions are unlikely to have a positive impact on outcomes (World Bank, 2004). This body of evidence highlights the contingency of the effects of decentralisation, linked to important interactions between formal structures of decision-making and informal structures of power and authority within bureaucracies, communities and schools. Furthermore, each link in the causal chain rests on certain assumptions which must be met in order for a change in the location of decision-making to have the desired effect(s). For instance, the assertion that involving parents and community members in the hiring and firing of teachers (an 'accountability' mechanism employed in many contexts) will improve quality of teaching rests on the assumption that (a) parents and community members will be able to identify high quality teachers who should be retained and/or rewarded, (b) the incentives provided will positively impact student learning and (c) former more centralised systems were less than optimal with regard to teacher recruitment and accountability, leaving scope for improvement through reform. This is not always achieved. In some contexts, teacher incentive schemes have been found to have a negative impact on overall student learning, if, for instance, they create perverse incentives for teachers to block the enrolment of low-performing students in order to maintain high average test scores within their classrooms (Glewwe et al., 2003). The impact of school-based decision-making models is, therefore, likely to differ depending on a wide variety of implementation factors, relating to the objective of the reform, the particular decisions that are devolved, the individuals given decision-making authority and the nature of the decision-making process.

At the beginning of the review process, we constructed a conceptual framework that depicted our understanding of the causal pathways, contributing factors and underlying processes that could affect the impact of school-based decision-making on educational outcomes. This framework (depicted below as Figure 1) was used as a 'working hypothesis' (Oliver, Dickson & Newman, 2012, p. 68) to guide the articulation of our specific review questions and review methodology (as recommended by Anderson et al., 2011).

Figure 1: Conceptual framework



Source: authors.

1.4. WHY IT IS IMPORTANT TO DO THIS REVIEW

Although the rhetoric around decentralisation suggests that school-based management has a positive effect on educational outcomes, there is limited evidence from low income countries of this general relationship. In reality, much of the decentralisation literature focuses exclusively on the proximal outcomes of school-based decision-making (described above). This is likely due to the relative ease of measuring such outcomes, as well as the shorter time period generally required to identify impact on intermediate outcomes. Evidence from the U.S. suggests that there can be a time lag of up to 8 years between the implementation of a school-based management model and any observable impact on student test scores, although intermediate effects may be more rapidly identifiable (World Bank, 2007, p. 13). This may explain why studies with different time scales have found mixed evidence around the impact of school-based management models on student learning outcomes (e.g. Jimenez & Sawada, 1999; King & Ozler, 2005).

As a result of these trends within the empirical literature, existing reviews on school-based decision-making have also tended to focus on proximal outcomes (e.g. Guerrero et al., 2012, on teacher absenteeism; Petrosino et al., 2012, on student enrolment). There are very few that consider the full range of relevant outcomes, including student learning. Those that do

have tended to focus exclusively on one particular mechanism (e.g. Bruns et al., 2012, on accountability reforms), rather than considering the full range of school-based decision-making models. The comprehensive reviews that do exist (Santibanez, 2007 and World Bank, 2007) are not formal systematic reviews, according to the criteria set by the Campbell Collaboration. They also need updating, as they (a) rely on literature that is now nearly ten years out of date and (b) focus almost exclusively on Central America, referencing almost no evidence from other low or middle income countries. There is, therefore, a need for a current globally-comprehensive systematic review of the impact of school-based decision making on a wide range of educational outcomes. Existing reviews on this topic also tell us very little about why school-based decision-making has positive or negative effects in different circumstances, a gap which this review also aims to address.

School-based management is a key component of education reform across the world, and it is a particular focus of education activities sponsored by many of the core development agencies, including the World Bank, USAID and DFID. It is, therefore, crucial that we gain deeper understanding of how school-based decision-making affects a broad range of educational outcomes in both positive and negative ways and how such models can be strengthened and improved. It is our hope that the timing of this review will also help to increase the potential impact of the results, as it coincides with ongoing conversations within the development community around the most appropriate focus (and strategies) for the next round of international development goals post-2015 (see http://post2015.org/; http://www.beyond2015.org/; https://sustainabledevelopment.un.org/).

2. Objectives

This review aims to answer the following overarching review question: *What is the evidence around how decentralising decision-making to the school level affects educational outcomes in low and middle income contexts (LMICs)?*

This broad question has been broken down into two discrete sub-questions:

(1) What is the impact of school-based decision-making on educational outcomes in LMICs?
(2) What are the barriers to (and enablers of) effective models of school-based decision-making?

The primary objective of the study, therefore, is to gather, assess and synthesise the existing evidence around how the decentralisation of decision-making to schools affects a broad range of educational outcomes in LMICs (Review Question 1 above). We have addressed this objective by examining the results of causal studies (i.e. those with an appropriate counterfactual) that consider the impact of at least one model of school-based decision-making on any of the proximal or final outcomes depicted in the conceptual framework above. We also aimed to draw conclusions about why particular models of school-based management work in some lower-income country contexts (and not in others), in order to make determinations about the particular contextual and implementation factors which act as barriers to – or enablers of – impact (Review Question 2 above). This objective has been addressed by examining evidence collected through a broader range of studies, including but not limited to that obtained from the included studies referenced in response to Review Question 1.

3. Methods

This review followed an explicit protocol (Carr-Hill et al., 2014), which in turn followed methodological guidance provided by the Campbell Collaboration and the EPPI-Centre at the UCL Institute of Education (Becker et al., undated; Gough et al., 2012; Hammerstrom, 2009; Shadish & Myers, 2004).

As this review aimed to both aggregate the demonstrated effects of school-based decision-making on educational outcomes and draw conclusions around the conditions and circumstances that can affect outcomes, we elected to conduct a mixed methods review, following the guidelines developed by Snilstveit (2012) for 'effectiveness plus' systematic reviews in international development. As such, our conceptual framework was used throughout the review to guide the search strategy, decisions regarding the inclusion and exclusion of studies, coding, and synthesis. In keeping with 'effectiveness plus' review methodology, we have considered different kinds of evidence in relation to our two review sub-questions. As the first review question is an 'effectiveness' question, studies included for synthesis needed to have an appropriate comparator or control group (or to have employed an appropriate method of constructing a counterfactual or control for confounding during analysis). However, a broader range of evidence, including studies based on qualitative data, were reviewed in response to the second sub-question, as we felt that other methods would be particularly useful for clarifying which external conditions and/or implementation factors can substantially affect outcomes.

3.1. CRITERIA FOR INCLUSION AND EXCLUSION OF STUDIES IN THE REVIEW

In order to be included in the review, all studies had to meet the selection criteria listed below.

3.1.1. Types of participants and settings

We looked exclusively at evidence related to primary and secondary schools in LMICs. In order to be included, studies needed to be based in at least one context classified (at the start of a given intervention) as either 'low' or 'middle' income, according to the World Bank classification. We excluded evidence collected in LMICs located within Central and Eastern Europe (including Turkey) or the former USSR.

3.1.2. Types of interventions

In order to be included, studies needed to investigate empirically the results of a *change in decision-making authority from a higher level of decision-making authority to the level of the school.*

As we were specifically interested in the impact of a *change* in decision-making authority which shifts decision-making to the school-level, studies analysing the impact of interventions which are implemented in schools but which do not include any additional decision-making authority in schools were excluded (e.g. government or NGO school feeding programmes). Specifically, studies including school-level interventions were excluded if the intervention was conceptualised, managed and implemented by an external decision-making agency, such as an NGO. The rationale for exclusion is that while theoretically schools *could* make use of devolved decision-making powers to implement such interventions, for example with the support of a grant, the effects of interventions implemented by external agencies are unlikely to be generalizable to interventions implemented by schools, so that the evidence from such studies does not shed light on the impact of actual school-level decision-making.

Studies of interventions aimed exclusively at improving the functioning of devolved decision-making structures – but not introducing new decision-making authority – were also excluded (e.g. interventions aimed at strengthening the effectiveness of pre-existing village education committees, such as the report card initiative discussed in Banerjee et al. 2008). Such studies do not report the effects of a change in decision-making authority specifically so lie outside the scope of the review. Moreover, examining questions of the more effective use of school's existing authority and jurisdiction would extend to a very large range of studies concerning issues of school management, suited to a separate review. However, studies which examine alternative ways in which new decision-making authority is granted to schools or employed by schools are included.

We excluded studies investigating a change in decision-making authority to a level higher than the school (e.g. studies of decentralisation to the region or district level). Studies that investigated the effects of privatisation of schooling were excluded on a related basis. While new private schools are in some cases more autonomous, expansion in this sector, sometimes the result of deregulation of the private sector, does not itself represent a shift in the decision-making authority of existing schools. Further, even where existing schools are privatised and privatisation does in fact affect the school's decision making authority, we consider this change to be primarily a change in the whole *nature* of school financing and governance, rather than a change in decision-making *authority*, such that the results of these studies are not informative with regard to the potential effects of decentralisation of authority to schools specifically. While privatisation of schooling may affect the outcomes of interest in this review, this is likely to occur via a range of mechanisms including effects on the composition of schools and on their accountability to parents, which will not be separable from changes in school-level decision making since they occur simultaneously.

We excluded studies of centralisation or recentralisation (reducing school-level decision-making authority) given that the scope of the review is on the impacts of a shift *towards* school-based decision making (i.e. decentralisation) and that this is the question of primary policy interest. Accordingly, studies which did not focus on a shift in decision-making authority towards the school were not included at the initial search stage. Evidence on the impacts of centralisation or recentralisation may be considered complementary to this review while it falls outside of the review remit.

Further, studies focusing on decision-making at levels lower than the school were also excluded. These include demand-side interventions (e.g. conditional cash transfers) intended to influence decisions made at the household, family or child-level.

This broad conceptualisation of school-based decision-making includes a number of discrete interventions, such as the establishment of school management committees and the distribution of school capitation grants. Given this potential diversity, we did not develop an exhaustive list of intervention models *a priori*. Rather, any study exploring an intervention meeting this definition of school-based decision-making was included.

3.1.3. Types of outcome measures

Included studies needed to investigate empirically the connection between school-based decision-making and *at least one educational outcome* (either proximal, e.g. attrition, equality of access, increased enrolment; or final, e.g. student learning, as captured by test scores, psychosocial and non-cognitive skills, etc.). Studies reporting stakeholder perceptions of a change in outcomes were excluded, as were studies exclusively reporting on processes or outputs (e.g. changes in the frequency of community participation).

Studies of any follow-up duration and studies with multiple follow-ups were included.

3.1.4. Types of study designs

All included studies needed to be empirical in nature. Normative, conceptual and/or descriptive sources were excluded.

In order to be included for synthesis in relation to **Review Question 1**, studies needed to rely on an explicit comparison or adopt an appropriate empirical strategy to identify causal effects. We used a two-stage approach to determine study eligibility. In the first stage, studies were considered potentially eligible for inclusion if they compared groups *not experiencing* school-based decision-making reforms with those *experiencing* school-based decision-making reforms or if they compared groups experiencing *different* school-based decision-making reforms (e.g. funding reforms versus school management reforms). Eligible study designs were:

1. Experimental designs using randomised or quasi-randomised assignment to the reform/intervention (e.g. randomised control trials)
2. Quasi-experimental designs, including studies in which:

    a. Assignment is based on known allocation rules including a cut-off rule on a continuous or ordinal policy variable (e.g. regression discontinuity design)
    b. Assignment is due to a natural experiment (e.g. exogenous geographical/political variation)
    c. Assignment is based on other selection mechanisms (e.g. self-selection by participating schools)
3. Before-and-after studies which collect longitudinal data at baseline and endline, as well as those using cross-sectional endline data only, provided data are collected from a comparison group or where an appropriate method of analysis has been used to:
    a. Match/create equivalent groups (e.g. statistical matching methods, such as propensity score matching and covariate matching); or
    b. Control for confounding in multivariate analysis (e.g. difference-in-differences and fixed effects regression, instrumental variables approaches, and regression analysis).

Any comparison needed to be contemporaneous (i.e., the interventions must have been implemented during the same time period - and, in comparisons between a reform group and a non-reform group, data needed to reflect the same time period) in order to be included. All of the included studies needed to analyse data at the level of the child or at the level of the school or community. Studies analysing comparison groups at sub-national or country level were excluded.

In the second stage, we determined whether studies would be included for synthesis in relation to Review Question 1 according to risk of bias assessment. Studies needed to be assessed as being either 'low' or 'medium' risk of bias (as outlined in Section 3.4.3) in order to be included. Studies deemed as being at high risk of bias were excluded from consideration in reference to Review Question 1. This included:

a) Studies where the study design was of questionable causal validity, such as those where comparison groups were not matched on observables, differences in covariates were not accounted for in multivariate analysis, or where there were serious threats to the validity of the statistical procedure used to deal with attribution;
b) Studies in which there was clear evidence of spillovers or contamination to comparison groups from the same communities; and
c) Studies in which reporting biases were evident.

However, studies in this category were not excluded entirely from the review. Rather, they were reclassified as potentially includable in reference to Review Question 2.

The eligibility criteria for **Review Question 2** included a broader range of empirical study designs, given the likelihood that non-causal studies would provide important data relating to implementation and contextual factors. Studies included in reference to Review Question 2, therefore, represented a range of designs, including:

1. Process evaluations and/or project completion reports of any of the school-based decision-making interventions evaluated in reference to the first review question
2. Other empirical studies (employing quantitative, qualitative or mixed methods of analysis) which provided data on either:
    1) factors found to affect the implementation of one of the school-based decision-making interventions evaluated in reference to the first review question, or
    2) conditions or circumstances found to affect the impact of one of the included interventions on the specified outcome(s).

Comparison groups were not a prerequisite for inclusion in relation to the second review question. However, in order to be included, studies needed to meet the standards of transparency, appropriateness, rigour, validity, reliability and cogency set out in the DFID 'How to note' on 'Assessing the Strength of Evidence' (2014). Studies classified as being of 'low' quality according to these criteria were excluded from the review.

Studies eligible for Review Question 2 provided evidence from specific programmes included in Review Question 1. Studies which provided evidence for specific interventions which were not included in Review Question 1 were excluded.

3.1.5. Other exclusion criteria

Date of Data Collection: Studies in which all data were collected prior to 1990 were excluded.

Language: Studies written in English, French, Spanish and Portuguese were eligible for inclusion in the review. Studies written in other languages were excluded, unless English translations were available, as we did not have any further linguistic ability represented within the review team.

Publication Status: We included both published (e.g. journal articles, books, conference papers and institutional grey literature, including reports and process evaluations) and unpublished (e.g. dissertations, theses and unpublished empirical studies showing null and/or negative results) literature.

3.1.6. Connections between causal and non-causal evidence

At the protocol stage, we anticipated identifying very few causal studies meeting the design criteria outlined above. As a result, we assumed that we would be able to say very little in reference to Review Question 1, so we intended to focus our attentions instead on synthesising the available non-causal literature. However, as we were ultimately able to identify a relatively large number of impact evaluations, it was necessary to change our strategy regarding the use of non-causal literature in the review. Instead of examining a broad diversity of studies in reference to the second review question, we elected to focus the qualitative component of our synthesis on those interventions that feature in the impact component of the synthesis, i.e. we limit our qualitative analysis to studies of the school-

based decision-making reforms examined in the impact studies. Following our initial statistical synthesis, we therefore reviewed the list of studies retained as potentially useful in reference to Review Question 2, and any study not investigating one of the specific interventions included in the meta-analysis was excluded prior to qualitative synthesis.

3.2. SEARCH STRATEGY FOR IDENTIFICATION OF RELEVANT STUDIES

Our search strategy involved five primary methods for identifying potentially relevant literature:

1. Identification of existing systematic reviews in related areas that might yield relevant references for inclusion in the review
2. Targeted searches in a wide range of bibliographic databases and websites likely to contain information relevant to the review
3. Hand-searching of relevant journals
4. Citation chasing
5. Contacting experts involved in research on school-based management

Of these five methods, the first three were completed at the start of the review process (July and August 2014; precise dates are included in Appendix 8.2). The final two methods were completed once we had determined an initial included studies list, following the screening, coding and quality appraisal phases of the review (January 2015).

Review of existing reviews

Existing systematic reviews were first identified through the 3ie Database of Systematic Reviews, the EPPI-Centre Database of Education Research, and the Campbell Collaboration Library. The reference lists for all potentially relevant reviews were then screened for any potentially includable studies. In total, we identified six reviews to screen. (A list is included as part of the reference list for this report).

Electronic searches of bibliographic databases and websites

We then conducted detailed electronic searches, with the support of our colleagues at the EPPI Centre, in a number of bibliographic databases and websites. (A detailed list is included as Appendix 8.1).[2]

Hand searches of relevant journals

We also completed hand searches for potentially relevant articles in the following academic journals: Compare, Comparative Education Review, International Journal of Educational Development, Journal of Development Economics, Economics of Education Review, Education Economics, World Development, World Bank Economic Review, and World Bank Research Observer.

Citation chasing

[2] As existing systematic reviews (e.g. Petrosino et al, 2012) have indicated a lack of relevant studies on education decentralisation in developing countries published prior to 2000, we limited our electronic searches to studies published in or after 2000. We did set any such data boundary for our other search methods (e.g. review of reviews).

Once we had determined a final list of studies for quality appraisal, we screened the reference lists of all included studies in order to identify any additional key sources that were missed during the initial search. We were unable to complete any forward citation chasing, due to time constraints.

Contacting the "informal college" of researchers in this area

We also reached out to a small list of experts who are known to have published widely on school-based management, in order to determine if there might be potentially relevant completed studies that are not yet published. Details are included in Appendix 8.3.

3.3. KEYWORD STRATEGIES FOR DATABASES AND WEBSITES

Our search strategy rested on two main 'concepts', each of which consisted of a large number of potential search terms:

- Concept 1: School-based decision-making models and mechanisms
- Concept 2: Low or Middle Income Countries

The list of search terms involved in Concept 1 was developed through an iterative process. First, members of the review team proposed a list of models, mechanisms and common phrases which have dominated the literature on school-based management in recent years. A test search was then conducted in ERIC and the IIEP decentralisation database, using this initial list of terms, plus some controlled terms for 'primary education' and LMICs and the date restriction 'published since 2000'. The test search yielded 170 records in the IIEP database and 152 records in ERIC. A repeated search in ERIC, without the primary school terms, yielded 483 records. A sample of 350 of these records, plus all of the records generated by the first two searches, were then hand-screened by the review team to generate further search terms for inclusion in the final search strategy.

Relying on the expertise of the EPPI Centre, we assembled a list of controlled terms which tend to be used in the main electronic databases in reference to Concept 2.

Search strategy for electronic databases

Our final search strategy for electronic databases comprised both free-form and controlled terms for both concepts. As controlled terms vary by database, a list of stem terms was developed which was then adapted to each database's individual thesaurus. The full search strategies are included as Appendix 8.2 to this report.

Search strategy for websites and online catalogues

The search strategy for websites and online catalogues was based on the main strategy (used in the electronic databases). However, as most websites and catalogues do not allow Boolean searching, it was deemed infeasible to conduct separate searches for each discrete term in the electronic search strategy. Instead, a list of 23 discrete search terms, representing Concept 1 of the overall search strategy, was developed for use in the website searching.

These search terms were entered independently into each website's search engine,[3] and a detailed record of the results for each website was stored in a shared Excel file.

We also translated this list of core search terms into French, Spanish and Portuguese. When conducting searches on websites deemed likely to include sources in multiple languages (e.g. Latin American Journals Online), additional searches were run using the translated terms.

The list of the website search terms is included in Appendix 8.2.

3.4. SCREENING OF STUDIES

3.4.1. Screening for relevance

Once the initial search was completed, all potential titles and abstracts were imported into EPPI-Reviewer, a specialist software package designed to assist with systematic reviews, and a duplicate check was completed.[4]

We then completed two screening phases: (1) Screening on Title and Abstract, and (2) Screening on Full Text.

During both screening phases, studies were reviewed and assessed against the review's inclusion/exclusion criteria (outlined above). Given the large number of identified studies, it was not possible to double-screen every study. Instead, we conducted a moderation exercise at the start of each phase of screening, in order to allow for a discussion of decisions between individual team members and to resolve any inconsistencies. We also double-screened a random sample of 10 percent of the total studies during each phase.

Screening on title and abstract was completed by three members of research team, using a pre-determined list of codes (included in Appendix 8.4). Initially, the coders only achieved an 89 percent agreement rate, but analysis of the discrepancies revealed that there was 100 percent agreement for all but one code ('Exclude Not School-Based Decision-Making'). The problematic code was subsequently disaggregated into three categories ('Not Education', 'Decentralisation to other level', and 'Not SBDM'), and all titles with this code were recoded. A 10 percent sample of these (re-coded) titles yielded a 95 percent agreement rate.

Screening on full text was completed by the same three team members, using another pre-determined list of codes (also included in Appendix 8.4). During this stage, the 10 percent sample yielded a 94 percent agreement rate between coders.

[3] For some smaller websites (e.g. Inter-American Development Bank Evaluation Reports database), it was feasible to conduct searches using only the word 'education'.
[4] EPPI-Reviewer maintains a detailed search log of every decision made during the importing, screening and coding phases, allowing for future replication of the review process.

3.4.2. Initial coding

All studies retained at the end of the second screening phase were then coded on a number of descriptive dimensions, as suggested by the conceptual framework. (The initial code list is included in Appendix 8.4.) Double-coding was not possible due to time constraints, but a second moderation exercise was conducted with all participating team members prior to initial coding.

3.4.3. Assessment of methodological quality and risk of bias

All included studies were then appraised for robustness of evidence and methodological rigour.

Review Question 1

Those studies using methods appropriate for consideration in reference to Review Question 1 (i.e. all impact studies) were designated as being of either 'low', 'medium' or 'high' risk of bias, using the coding criteria outlined in Appendix 8.4. All of the 'effectiveness' studies were double-coded by two members of the review team before final classifications were confirmed. Any disagreements were resolved through discussion until a consensus was reached.

In order to be classified as a 'low risk of bias' study, a study needed to:

a) Demonstrate clear measurement of and control for confounding, including selection bias, and have no suspected sources of unobserved confounding;
b) Adequately describe the reform/intervention and comparison groups;
c) Have low risk of spillovers or contamination; and,
d) Demonstrate low risk of reporting biases and other sources of bias.

Studies were classified as at 'medium risk of bias' if either:

a) There were moderate threats to the validity of the attribution methodology (arising from issues with the implementation of the methodology), or
b) There were either likely risks of spillovers or contamination (arising from inadequate description of the intervention or comparison groups) or possibilities for interaction between groups (e.g. drawn from the same community), or
c) There were possible reporting biases.

All other studies were classified as 'high risk of bias studies'. This category, therefore, included:

a) Studies where the study design was of questionable causal validity, such as those where comparison groups are not matched on observables, differences in covariates are not accounted for in multivariate analysis, or where there are serious threats to the validity of the statistical procedure used to deal with attribution; or

b) Where there was clear evidence of spillovers or contamination to comparison groups from the same communities; or
c) Where reporting biases were evident.

High risk of bias studies were automatically excluded from synthesis in reference to the first review question and reclassified as potentially relevant for the second review question. Medium and low risk of bias studies were retained for synthesis.

It should be noted that these ratings are subjective and were based entirely on what was reported in the study documents. However, our independent assessments of the studies were broadly similar (we had 80% initial agreement across the nearly 50 studies). This would suggest that we were generally evaluating the threats to validity in a similar fashion.

Review Question 2

Studies which could only be retained in reference to the second review question (including any impact studies classified as high risk of bias) were coded for quality appraisal using a separate quality appraisal code list, also included in Appendix 8.4.[5] These non-casual studies were then classified as being of 'high', 'medium' or 'low' quality.

'High' quality studies needed to have received a 'High Quality' code for each of the dimensions assessed. 'Medium' quality studies needed to receive 'High Quality' designations for all transparency indicators, for all indicators related to the appropriateness of the research design, for all validity indicators and for evidence of supported conclusions but may have received a designation of 'Unclear' for some of the methodological indicators (e.g. details of data collection or analysis). Any study receiving at least one 'Low Quality' code was classified as 'low' quality.

Low quality studies were excluded prior to synthesis. High and medium quality studies were retained for synthesis in reference to the second review question.

A random sample of 10 percent of the Review Question 2 studies were double-coded to check for reliability between the three reviewers involved in the quality appraisal of the non-casual studies. A 94 percent agreement rate was achieved between the three coders.

3.5. DATA EXTRACTION

For each included study, we then extracted data regarding the study setting, participants, methods, details of the 'intervention', comparison conditions (if relevant), outcomes, and risk of bias/quality classification.

[5] The phrase 'risk of bias' can be problematic when discussing qualitative studies. As a result, the term 'quality' has been used in reference to this second group of studies.

For all impact studies (i.e. those relevant for inclusion in reference to the first review question), we also extracted any reported effect sizes (including the direction of the effect and any reported sub-group effects), confidence intervals and computation procedures.

Due to time constraints, data extraction was initially completed by one member of the review team. However, during synthesis, each study was read by a minimum of two reviewers, and all extracted data was double-checked by an alternate reviewer.

3.6. CRITERIA FOR DETERMINATION OF INDEPENDENT FINDINGS

A number of the included studies provide impact estimates on multiple outcomes (e.g. student learning outcomes and student drop-out rates) or on multiple dimensions of the same outcome-type (e.g. analysis of impact on learning outcomes, assessed through tests in science, math and literacy). Some studies report multiple estimates for the same outcome using different methodologies or specifications; others also provide estimates for more than one time period. The studies represent a broad range of intervention mechanisms and models.

Studies were first separated by intervention type and outcome/domain, so that pooled impact estimates could be produced separately for each intervention/outcome pair. In order to ensure that pooled impact estimates for each intervention type and outcome/domain were constructed from statistically independent findings, only independent estimates of effects were included, on the following basis:

- Where a study reports effect sizes relating to a particular intervention on more than one outcome/domain, we included these estimates separately in the relevant pooled impact estimate.

- Where a study reports more than one effect size for a particular intervention on an outcome/domain, for example based on different model specifications or different achievement tests used to assess the same domain, we included only one estimate, except in the case that a study is implemented across more than one non-overlapping and independent sample (being effectively independent studies), when one effect is included for each sample. The choice of effect involved up to two judgements: first, we selected the most robust methodology, with the lowest likelihood of risk of bias; second, we selected the most 'intensive treatment' (e.g. the longest exposure to the intervention or the most extensive form of decentralisation, in experiments with multiple treatment arms).[6]

- For each independent sample, only one estimate is included when effects are reported for more than one time-period, being the effect assessed as having the lowest risk of bias in attributing impact, or where the risk of bias is equal, for the most recent time-period.

[6] This decision was methodologically necessary in order to conduct the meta-analysis, as we could only include one effect per study. However, we recognise that evidence of differential effects over time is also policy relevant, so we consider the effect of time-lag in the heterogeneity analysis below.

- Where estimates of effects for the same intervention and sample are reported at more than one level – for example using individual pupil-level outcomes and outcomes aggregated at class or school-level, we include individual level results only to reflect the larger sample size, provided that the 'risk of bias' associated with the method employed is not greater than for the estimates at aggregate-level.

- If more than one paper analyses and reports the results of the same intervention/programme using similar or different methods and specifications but employing same or a similar sample (leading to dependent results) we treat these papers in a way equivalent to a single study reporting multiple effect sizes (outlined above).

Given the limited number of studies retained for final synthesis, it was not possible to provide separate pooled estimates for sub-groups, especially because the studies rarely report separate estimates for a common set of sub-groups.

3.7. STATISTICAL PROCEDURES AND CONVENTIONS

3.7.1. Calculation of effect sizes

Our preferred estimate of effect-sizes for meta-analysis is the 'standardised mean difference' (SMD) in outcomes between intervention and control groups (or comparison groups for non-experimental studies). This statistic provides an estimate of the change in outcomes due to the intervention in terms of standard deviations of the outcome of interest and is therefore comparable across studies, subject to certain assumptions. It is not possible in every case to calculate the SMD, however, particularly for studies which do not report standard deviations of the outcome variable and/or the number of observations in the study or the statistics required to compute or estimate the standard deviation or other required statistic. However, we have employed appropriate methods to generate comparable effect-sizes (as below) wherever possible, which permit comparison of effect sizes.

Reported data have been employed to compute standardised mean differences (Cohen's *d*) for continuous outcomes using the formula below for experimental studies, where the numerator is the difference in means between control and treatment groups (or post-treatment difference in a matching study) and the denominator is the pooled standard deviation across both groups.

$$d = \frac{\bar{x}_t - \bar{x}_c}{s_{pooled}}$$

For studies reporting regression results, we calculate SMD as follows,

$$d = \frac{\beta}{s_{pooled}}$$

where the numerator represents the regression co-efficient of interest, or the 'average treatment effect on the treated' in a matching study.

The pooled standard deviation is calculated as

$$s_p = \sqrt{\frac{(n_t - 1) * s_t^2 + (n_t - 1) * s_c^2}{n_t + n_c - 2}}$$

employing the sample sizes for treatment and control groups and the standard deviations of the outcomes for each group, or alternatively, for regression studies employing the standard deviation of the outcome at baseline:

$$s_p = \frac{\sqrt{((SD_y^2) * (n_t + n_c - 2) - (\frac{\beta^2 * (n_t * n_c)}{n_t + n_c})}}{n_t + n_c}$$

We make statistical adjustments required for small sample sizes in all cases (the effect is indiscernible for larger samples) using the following correction (multiplied by the SMD):

$$1 - \frac{3}{4 * (n_t + n_c - 2) - 1}$$

The standard error of the SMD is calculated as follows:

$$SE = \sqrt{\frac{n_t + n_c}{n_c * n_c} + \frac{SMD^2}{2 * (n_t + n_c)}}$$

We use the SMD and its standard error to calculate confidence intervals for effect sizes (see Keef and Roberts, 2004; Borenstein et al., 2009) and for meta-analysis using Stata's *metan* command.

In some cases, studies report effects on outcomes using definitions which result in effects of opposing signs having the same interpretation – for example while the outcome variable 'drop-out' is more commonly reported, occasionally studies report 'retention' which is the complement of drop-out. In such cases, we have adjusted the reported effects to be consistent – reporting drop-out as the outcome in all cases, for example, so that a negative effect is always desirable and that effects are directly comparable.

In some cases, information required for the direct calculation of standardised mean differences is not reported. Where other appropriate data are available, we have employed appropriate formulae to compute effect sizes from statistics reported (such as t, z or F statistics, p values and standard errors) using the Campbell Collaboration online effect size calculator (http://www.campbellcollaboration.org/resources/effect_size_input.php). Full information is included in Supplement 2.

We analyse the likelihood of 'unit of analysis error' (see Higgins and Green, 2011) by examining whether studies have employed appropriate statistical methods to account for data clustering, such as the use of cluster fixed effects and robust standard errors. Such error can occur, for example, in studies investigating a decentralisation intervention where decision-making power is shifted from districts to schools, which use a measure of impact based on pupil-level test scores in selected schools in districts in receipt of the intervention, as compared to pupils in selected schools in control districts. This is because the unit at which the intervention is implemented (district) differs from the unit of analysis (pupils clustered in schools). As pupils within clusters are likely to be more homogenous than across clusters, pupil-level observations are not fully independent. Such data 'clustering' at school and district level can be accounted for in the analysis to ensure standard errors and confidence intervals reflect the fact that treatment allocation is at cluster rather than individual level. Our analysis finds that in all studies where clustering of standard errors was required to avoid unit of analysis error, this had been done by the authors and was reflected in the study results.

Supplement 2 presents the effect size and variance calculations for all studies, along with any notes regarding the effect size calculations.

3.7.2. Meta-analysis

We began the synthesis process by creating a summary table of all included effectiveness studies (see Supplement 1). Given that some studies include multiple treatment arms involving different intervention models, it became quickly apparent that there were very few consistent intervention-outcome pairs in the sample.

As a result, we begin our analysis by reporting the impact of *any* school-based decision-making reform on the six educational outcomes for which sufficient data could be identified to calculate the SMD for more than one study: 1) student drop-out; 2) student repetition; 3) teacher attendance; and 4) student learning, as assessed via i) language test scores, ii) math test scores, iii) aggregate test scores (i.e. tests including more than one subject). We do not report aggregate test scores where more than one of the scores contained in the aggregate is already reported separately. Due to data limitations, other outcomes are discussed narratively but these effects are not pooled or presented visually via a forest plot.

We then examine the relationship between three moderating variables and these main effects:

1) *The school-based decision-making mechanism.*
   As nearly every study presents a different version of school-based decision-making, it was not possible to conduct detailed analysis around specific intervention models, but it was possible to classify the interventions into a broad typology of school-based decision-making and to consider any differential effects on outcomes. This typology is outlined in Section 4.2.

2) *World Bank income classification category*.
Hanushek et al. (2011) have argued that the impact of school autonomy depends on the level of development of the country implementing the reform. We test this hypothesis by analysing the impact of school-based decision-making models implemented in low income, lower middle income or upper middle income countries.

3) Type of evaluation design.
Finally, we investigate whether there is any difference in the results of studies that make some attempt at randomisation versus those using quasi-experimental approaches.

We also conduct robustness checks by examining how effect sizes vary between studies classified as 'low' or 'medium' risk of bias. In order to check for any potential publication bias in our results, we also produce funnel plots for each of the study outcomes and conduct the Egger et al. (1997) test for asymmetry in the case of each outcome. This test examines the relationship between effect sizes and standard errors in a linear regression framework, using inverse variance weights. Following Duval and Tweedie (2000), we also conduct a 'trim and fill' analysis for each set of estimates by outcome. This non-parametric method adjusts the meta-analysis for the number and outcomes of theoretical missing studies and attempts to correct the estimate of the pooled effect size for funnel plot asymmetry.

These moderators and methods were selected *a priori*. Two of the three moderators were chosen based on our pre-existing knowledge of the decentralisation literature; we were aware of multiple studies indicating that effects may vary depending on the model of school-based decision-making and on the level of development of the country in question (see, for example, Barrera-Osario et al., 2009; Hanushek et al., 2011; Santibanez, 2007). Type of evaluation design was chosen as the third moderator – and we decided to check for robustness, using risk of bias classifications, and to conduct tests of publication bias – because all three methods are standard practice in many systematic reviews (see, for example, Petrosino et al., 2012).

3.8. TREATMENT OF QUALITATIVE STUDIES

All of the included studies (both those included in the impact analysis and those retained as potentially useful supplementary sources) were coded on a number of dimensions pertaining to implementation and context, following the final coding list included in Appendix 8.4. These data were then analysed and aggregated, following the principles of framework synthesis (Thomas et al., 2012), in order to identify the main barriers and enablers that appear to have influenced the impact of the interventions under review.

As we have insufficient data to statistically test the specific impact of any of these factors on differences in effects (i.e. by conducting further moderating variable analyses on the forest plots), we combine the two components of our analysis by creating a revised conceptual framework, using a narrative synthesis approach along the causal chain (as suggested by Noyes & Lewin, 2011).

4. Results

4.1. FLOW OF STUDIES

Our initial search yielded **2,817 titles** (135 from systematic reviews, 2,141 from databases and 541 from website and hand searches). Of these, **1,541** were excluded during the first phase of screening on title & abstract. We were able to retrieve **1,186** of the remaining studies, of which **96** met our eligibility criteria. An additional four studies were identified through reference searching and expert checking.

Of these 100, **30** could be classified as 'impact evaluation' studies, as they met the design criteria required for inclusion in reference to *Review Question 1*. These studies were appraised for risk of bias, following the procedures outlined in Section 3.4.3. The remaining 70 were classified as non-causal studies and subjected to quality appraisal, following the procedures outlined in Section 3.4.3.

Following risk of bias assessment, three of the 30 impact studies were reclassified as non-causal studies of potential relevance for Review Question 2, as the risk of bias was judged to be too high for them to be included in reference to Review Question 1. In two of the three studies (Paes de Barros & Mendonca, 1998; de Umanzor et al., 1997), we identified a substantial risk of confounding factors influencing the impact estimates, while there was a high risk of bias due to attrition in the final study (Cueto et al., 2008). Other risks were also identified, including risk of motivation bias and clustering, in one of the three studies (de Umanzor et al., 1997). Full results of the risk of bias analysis are included as Appendix 8.5.

One additional study (Carnoy et al., 2008) had to be dropped from the final synthesis because of missing data.[7] **Twenty-six impact studies** were therefore included in the meta-analysis.

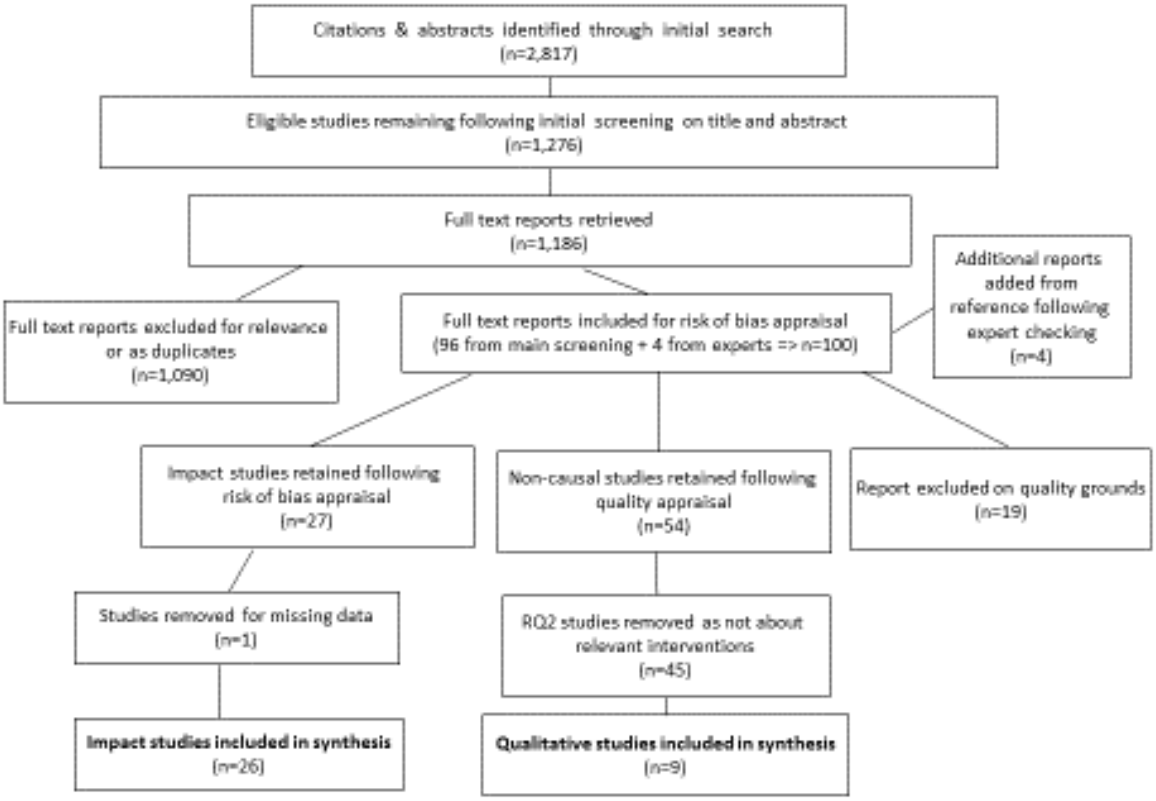
Of the 73 non-causal studies subjected to quality appraisal (i.e. the 70 non-causal studies, plus the three impact studies reclassified as only includable in reference to Review Question 2), **19** were classified as "Low Quality" and excluded from the review. A detailed outline of the reasons for exclusion of these 19 studies can be found in Appendix 8.6.

[7] The author was contacted to request the missing data, but no response was received.

As discussed in Section 3.1, the list of non-causal studies was further reduced by removing all studies about interventions not captured in the impact analysis. This final exclusion process resulted in a list of **nine non-causal studies** for synthesis relating to *Review Question 2*.

The pipeline of studies is illustrated in Figure 2. Lists of the included impact and non-causal studies are included as Supplement 1 and Supplement 3.

Figure 2: Pipeline of studies



4.2. INTERVENTIONS

In total, the 26 causal studies investigate the impact of 17 individual interventions. To complicate the analysis further, many of the studies involve multiple 'treatment' arms, each reflecting a slightly different variation of school-based decision-making. As each of these variants is likely to affect the overall impact, we begin by presenting a brief description of the 17 interventions referenced in the subsequent meta-analysis. Table 1 presents the most salient characteristics of the named interventions.

Table 1: Table of intervention characteristics

Name of Intervention	Country	Description	Relevant Impact Studies	Linked Non-Causal Studies
EDUCO	El Salvador	EDUCO, established in 1991, is a national programme that gives communities autonomy over most educational decisions. Under the EDUCO model, community education associations – in which parents are the majority – are responsible for administering and managing the school, including hiring, firing and paying teachers. Community education association members are elected by their peers and receive training on various aspects of school management. Community education association members must be literate and they are elected by their peers. They also receive training prior to assuming their duties	Jimenez & Sawada (1999); Jimenez & Sawada (2003); Sawada & Ragatz (2005)	de Umanzor et al. (1997)
PROHECO	Honduras	The EDUCO programme spawned a number of similar initiatives in Central America, including PROHECO in Honduras. Much like EDUCO, PROHECO schools are managed by parental councils, which have responsibility for a broad range of management duties, including hiring, firing and paying teachers.	Di Gropello & Marshall (2005)	N/A
Autonomous Schools Program	Nicaragua	In the early 1990s, the Nicaraguan government established 'consultative councils' in all public schools, in order to stimulate greater participation of teachers and parents in school decisions. Councils consisted of head teachers, teachers, parents and students. In 1993, the consultative councils at a small sub-sample of public secondary schools were transformed into School Management Councils in 1993 and given legal status and autonomy over the majority of school decisions. This pilot programme eventually expended into primary education in 1995. The councils of the newly-created autonomous schools, in which parents held the voting majority, had the ability to hire and fire teachers and the responsibility to maintain their infrastructure and academic quality. They also had control over monthly fiscal transfers that paid for teacher salaries, benefits and basic maintenance, and they had the right to charge and retain fees. The Ministry of Education retained control over staff promotion, teacher certification and the national curriculum.	King & Ozler (2005)	Fuller & Rivarola (1998); Gershberg & Meade (2005)
School Based Management	Indonesia	School-based management was established in Indonesia in 2003. SBM grants principals, teachers, and other local community-based members with autonomy over academic operations of schools. A grant program accompanied the reform, which provided a per-student amount to all schools that could be disbursed according to local priorities. In 2006, recognising that school committees were largely not realising the autonomy granted to them through the reform, a field experiment was implemented by the World Bank to test four measures aimed at helping committees to fulfil their management roles.	Pradhan et al. (2011)	Bandur (2008); Bjork (2003); Vernez et al. (2012)

Name of Intervention	Country	Description	Relevant Impact Studies	Linked Non-Causal Studies
AGEMAD	Madagascar	The AGEMAD reform sought to improve the efficiency and effectiveness of the education sector in Madagascar by specifying roles and responsibilities and introducing new monitoring tools at each level of the school management hierarchy. At the school level, the intervention focused on the provision of new administrative tools for teachers (e.g. lesson planning forms), the introduction of school report cards, and the organization of school meetings with school staff, parents and members of the community (intended to increase parental and community involvement in monitoring). An RCT was designed to test the impact of three different implementation designs: 1) a cascade model in which district officials were trained to implement the reform through the district; 2) a modified cascade model in which both district and sub-district officials were trained the implement the reform; and 3) an intensive model in which district officials, sub-district officials and individual schools were trained directly.	Glewwe & Maïga (2011); Lassibille et al. (2010)	N/A
Quality Schools Program (PEC)	Mexico	PEC was introduced in 2001 and seeks to increase community participation in school-based decision-making, reducing administrative burden for participating schools and providing technical support to participating schools. The programme is guided by national regulations of the federal government but administered by state departments. The federal government provides match funding to encourage state participation in the funding of PEC. In order for a school to qualify for PEC, school directors, teachers, and parents need to identify a school's problems and needs and design a school improvement plan. PEC schools qualify for annual programme grants of up to about $5,000 and also receive $2 for each dollar that the school raises from the municipal government or private sector. The grant amount depends on the socioeconomic status of the community, the educational needs identified in the school improvement plan and the characteristics of the community population. Communities must spend 80% of their grant in the first four years; funds must be spent on teacher training, interventions for at-risk students, educational materials/teaching equipment or infrastructure. Training is provided to school principals and directors of school-management committees.	Bando (2010); Murnane et al. (2006); Skoufias & Shapiro (2006)	Reimers & Cardenas (2007)
Support to School Management (AGE)	Mexico	AGE, a precursor to PEC, was implemented in the late 1990s as part of a broader school reform that aims to improve service delivery and education quality in highly deprived parts of Mexico. AGE provides a small amount of financial support ($500 -$700 per year depending on the school size) to parents associations who have autonomy in using the funds for school improvement. Parents receive training about the role of parent association, the use of school funds and how to participate in a range of activities that involve effective management of the school. The use of funds is restricted and cannot be used to fund salaries.	Gertler et al. (2012)	N/A

Name of Intervention	Country	Description	Relevant Impact Studies	Linked Non-Causal Studies
Program to Strengthen and Invest Directly in Schools (PEC-FIDE)	Mexico	PEC-FIDE was a spin-off of PEC, implemented in six Mexican states in 2008. Schools that had participated in PEC were also eligible for PEC-FIDE, but it was not possible for schools to receive funds from both programmes simultaneously. PEC-FIDE was similar to PEC, in that schools received grants in exchange for collaborative school planning and decision-making. The amount of the grant depended on school enrolment but generally averaged around $4,500. Funds could be spent on training, interventions for at-risk students, materials/equipment and infrastructure. School councils - comprising head-teachers, teacher representations and parent representatives - were responsible for drafting School Improvement Plans and received training prior to receipt of the grant. Crucially, schools do not opt in to PED-FIDE; they are assigned to the programme by the state government, depending on programme targets	Santibanez et al. (2014)	N/A
Third Elementary Education Project (TEEP)	Philippines	TEEP, implemented from 2000 to 2006 by the Philippine Department of Education, targeted the most deprived public primary and elementary schools in the Philippines. The act legalising the reform (Republic Act 9155) vested decision-making authority in the office of the school head, not in the broader community. The Act also grants managerial autonomy, not financial freedom nor autonomy over personnel decisions. Under TEEP, schools received cash grants for maintenance and operating expenses, based on the enrolment of the school. Schools were also allowed and encouraged to raise their own funds from their communities. TEEP was a well-resourced programme that combined physical and soft components with institutional reform. The programme invested in physical buildings and textbooks, provided training to teachers and principals, and facilitated partnership between the school and community.	Khattri et al. (2010); Yamauchi & Liu (2012)	N/A
School-Based Management	Philippines	Prior to the implementation of TEEP, there was a national law in the Philippines that granted principals autonomy over academic, administrative and financial affairs in their schools. Although the law encouraged the creation of school management committees, there was no mandate to create such committees, so they were only created if individual principals so desired.	San Antonio (2008)	N/A

Name of Intervention	Country	Description	Relevant Impact Studies	Linked Non-Causal Studies
BESRA	Philippines	Building on the success of TEEP, in 2006, the Philippine government mainstreamed SBM by including it as an element of the system-wide Basic Education Reform Agenda (BESRA). BESRA was built around five key reform thrusts relating to teacher development, social support for schools, early childhood development, private sector involvement in education and general improvement of educational governance. The SBM component involved the establishment (or capacity building for existing) school governing councils, the preparation of school improvement plans, and an increased level of resources managed and controlled at the school level. As part of BESRA, principals and other school staff received training. BESRA was scaled up to schools that were outside the original TEEP catchment area through the use of a partnership model under which non-TEEP schools were partnered with neighbouring TEEP divisions in order to introduce SBM. Unlike TEEP, BESRA did not involve any additional package of investments.	World Bank (2013); Yamauchi (2014)	N/A
Programme for School Improvement	Sri Lanka	PSI was designed to increase involvement of the school community including parents, teachers and past pupils in the management of school. The programme emphasized development of a school improvement plan, efficient utilization of resources, and improved cooperation between schools and communities in order to enhance quality of curricular and co-curricular activities. It also prioritised staff training to address the school needs and improve relationship between schools and communities. Under PSI, School Development Committees became responsible for managing schools. A Report Card Programme (SRCP) was implemented simultaneously, on a relatively small scale, in order to inform the school community of their school's performance. Report cards were completed by school personnel and distributed to parents and School Development Committee members.	World Bank (2011)	N/A
Rural Education Program	Colombia	The Rural Education Program empowers municipal operating units (comprising local officials and members of the education sector) to assess needs and choose educational interventions for rural communities. Schools in the project are given the authority to implement/monitor their chosen educational intervention and are also provided with a "basket" of educational goods and teacher training.	Rodriguez et al. (2010)	N/A
Whole School Development	Gambia	The WSD program provided training for head teachers, teachers and representatives of students and parents, in addition to a capitation grant. Grants were controlled by school management committees and could only be spent on teaching and learning activities.	Blimpo & Evans (2011)	N/A

Name of Intervention	Country	Description	Relevant Impact Studies	Linked Non-Causal Studies
School Based Management pilot programme	Niger	This pilot programme in Niger provided capitation grants to schools. No restrictions were placed on the use of the funds, except that parent associations were given complete authority over their use. Training was provided to committee members prior to disbursement of the grants.	Beasley & Huillery (2014)	N/A
Extra Teacher Program	Kenya	Parent teacher associations in Kenya traditionally used money raised through school fees to hire short-term contract teachers. However, when the introduction of Universal Primary Education eliminated fees, PTAs no longer had funding available for teacher recruitment. ETP was designed to reinstate the possibility of contract teacher contracts by providing funds to a random sample of school management committees in Western Kenya. Under the program, SMCs had the authority to hire and monitor contract teachers. A random subsample of schools in the study were provided additional training for SMC members as a supplementary intervention which was found to reduce the likelihood of reduced effort by non-contract teachers. The program was subsequently scaled up to the national level.	Bold et al. (2013); Duflo et al. (2012)	N/A
Evaluation of a participatory report card intervention	Uganda	An evaluation was designed to test the relative impact of two kinds of school report card: a standardised report card, designed by the Ministry of Education, and a participatory report card, designed by individual school management committees. Committee members were trained in both treatment groups, but only those in the participatory arm were given the freedom to design their own instrument.	Barr et al. (2012)	N/A

The diversity of specific intervention types rendered it impossible to conduct meta-analysis of a clear set of standardised intervention-outcome pairs; instead, we elected to create a typology of broad intervention types to use during synthesis, based on typologies of school-based management models included in Barrera-Osorio et al. (2009) and Santibanez (2007).

The typology was created by coding each study on a range of dimensions, based on elements of our initial conceptual framework. A full code list is included in Appendix 8.4. Studies with multiple treatment arms were given a full set of codes for each differentiated treatment model. The codes were then converted into ordinal or binary variables and added to the data set in Stata.

Once the data were aggregated, we were able to identify three broad intervention types, which could then be used in subsequent analysis:

High Decentralisation

The first category of school-based decision-making interventions comprises all models in which the school (and/or the local community) has decision-making authority over nearly all aspects of school management. Most importantly, in order to be classified as 'high decentralisation', the school – or school management committee – under investigation needed to have authority over both financial and personnel decisions (e.g. the authority to hire/fire teachers and the authority to pay salaries). Four interventions were classified as 'high decentralisation' (EDUCO, Nicaragua's Autonomous Schools programme, PROHECO, and the most intensive version of Kenya's Extra Teacher Program).

Medium Decentralisation

To be classified as 'medium decentralisation', a school – or the school management committee – needed to have authority over some management decisions. However, schools in this classification would not have authority over personnel decisions. Twelve interventions were classified as 'medium decentralisation' (all three variants of Mexico's school-based management reform – AGE, PEC and PEC-FIDE; all three variants of the school-based management reforms implemented in the Philippines, including TEEP; PSI in Sri Lanka; Gambia's Whole School Development programme; AGEMAD in Madagascar; school-based management reform in Indonesia; and the two unnamed school-based management interventions implemented in Niger and Uganda).

Low Decentralisation

'Low decentralisation' models do not involve much devolved decision-making authority. This classification include models in which schools have the power to make curricular decisions and/or decisions about infrastructure and buildings. No schools in this classification have authority over financial decisions. One intervention was classified as 'low decentralisation' (the Rural Education Program in Colombia).

4.3. DESCRIPTIVE STATISTICS

This section describes the general characteristics of the 35 impact and non-causal studies included for synthesis.

4.3.1. Impact studies

Although the final sample of impact studies is relatively small (n=26), it represents a diversity of geographic contexts. The region most heavily represented is Latin America (n=12), with Mexico (n=5), El Salvador (n=3) and Nicaragua (n=2) being the most common individual countries. This is unsurprising, given that Latin American countries were amongst the first lower income contexts to attempt to decentralise their education systems. Other Latin American countries featuring in our sample include Colombia and Honduras. Seven of the studies investigate school-based decision-making in sub-Saharan African contexts (specifically Kenya, Madagascar, Gambia, Niger and Uganda). No African country featured in more than two studies. Finally, seven studies analyse South or Southeast Asian contexts, with the Philippines being the most frequent (n=5). Other Asian countries include Indonesia and Sri Lanka.

The studies are also quite diverse in terms of income classification. Of the 26 impact studies, eight were based on low income contexts, 13 in lower middle income contexts and five in upper middle income contexts.[8]

Most of the studies investigate interventions targeted at primary schools (n=23, 88%). One study considers an intervention at the secondary level, while the remaining two studies consider outcomes at both primary and secondary level.

Nine of the studies (32%) use randomisation to assign participants to groups, while the remaining 17 (65%) use quasi-experimental procedures. Although the included studies represent a range of publication dates (from 1999 to 2014), all of the studies using random allocation have been published since 2008.

The risk of bias assessment (see Appendix 8.5) indicated that eight studies (27%) could be classified as of low risk of bias overall. All of these studies were assessed as having used randomised assignment appropriately and we were not able to identify any sources of bias relating to factors such as method of allocation, attrition, contamination, motivation bias or biases in analysis reporting. Most other studies (63%), including three RCTs, were classified as having medium risk of bias, usually due to risks of confounding and/or contamination of comparison groups. As mentioned above, three studies (10%) were assessed as having high risk of bias and were excluded from the meta-analysis.

Only six of the studies (23%) were published as articles in academic journals; the majority (N=16, 62%) are World Bank reports or working papers published by economic think tanks.

[8] Income classifications reflect the World Bank's income classification system. Classifications were linked to the start date of the intervention under investigation, rather than the current classification.

Three of the included studies were published as chapters in one World Bank publication. One is an unpublished PhD thesis. The implication of this is that about two-thirds of our included studies are reports which may never have been through an external peer review process.

A full list of the characteristics of the 26 impact studies can be found in Supplement 1.

4.3.2. Non-causal studies

We also consider evidence from nine non-causal studies. Of these, two are multi-country studies (Gunnarsson et al., 2008; Hanushek et al., 2011). The remaining seven relate to four of the interventions investigated in the impact studies: Indonesia's national school-based management reform (3 studies); Nicaragua's Autonomous Schools programme (2 studies); EDUCO (1 study); and PEC (1 study). A full list of the characteristics of the non-causal studies can be found in Supplement 3. The assessment of study quality in each of the included non-causal studies is presented in Appendix 8.6.

4.4. INTERPRETING THE META-ANALYSIS FINDINGS

We estimated the pooled effect size *d* across studies for each outcome for which sufficient data could be identified from more than one study (i.e. maths score, language score, aggregate test score, drop-out, repetition and teacher absence), using a random effects model with inverse variance weights. Standardized mean differences (Cohen's *d*) are scaled naturally so that: if there was a beneficial impact for an intervention, then the *d* was positive for any one of the test scores and for teacher attendance and negative for drop-out and repetition, and if the effect for the intervention was identical for the treatment group and the control group (e.g. 5% drop-out rate in both groups), then the *d* was zero. To give an example, an effect size estimate of .10 reflects one-tenth standard deviation improvement for treatment participants compared to control participants.

However, it is often unclear if such an effect has any substantive meaning beyond the study context. As discussed in Petrosino et al. (2012), Rosenthal and Rubin (1982) suggest converting a standardized mean difference to a percentage improvement of the treatment group compared to the control group. Using this technique (and assuming, for example, a baseline drop-out rate of about 10% across treatment and control), a standardized mean difference of -.10 could be interpreted as about 1 percent improvement in the intervention group. Whether or not such an effect is policy relevant depends largely on the context, the cost of the intervention, and other factors.

Moreover, certain outcomes, such as drop-out and repetition may be defined and measured differently in different country contexts; equally, teacher absence has been measured differently in the different studies, and of course the tests used to generate the test score are different in potentially important but unknown ways. One important caveat with regard to interpretation of test-score data is that changes in test scores measured in standard deviations are in fact relative measures, so comparisons across different tests are not direct

comparisons on the same underlying metric, so are only indicative. For example, it may be easier to generate a one standard deviation change in reading among a group of early readers than among a group of proficient readers and the interpretation of a one standard deviation change depends upon the sample and population concerned. Such differences are considered where appropriate as part of the discussion of heterogeneity of effects.

We conducted the meta-analysis on 27, instead of 26, effect sizes, for two reasons. First, three of the studies (King & Ozler, 2005; Parker, 2005; Santibanez et al., 2014) were found to include estimates for two discrete samples. As these separate estimates do not violate the assumption of independence of samples, we included them separately in the meta-analysis. Second, in two instances, we found that two studies had identical samples to another study in the final list (Lassibille et al., 2010, and Glewwe & Maïga, 2011, regarding the AGEMAD programme in Madagascar; Jimenez & Sawada, 1999, and Sawada & Ragatz, 2005, regarding the EDUCO programme in El Salvador). As the inclusion of the estimates from both studies would have violated the assumption of independent samples, we selected the estimates from the more robust study. The estimates from Jimenez & Sawada (1999) were therefore excluded from the meta-analysis, although the qualitative results have been included in the heterogeneity analysis in Section 4.9. In the case of Glewwe & Maïga (2011), the results are excluded because, while we consider this study equally robust by comparison with Lassibille et al. (2010), it reports results for only one outcome (aggregate test scores) also reported in the latter study which, in addition, reports a range of other outcomes.

For each analysis of overall intervention effects, we have calculated heterogeneity statistic in the form of the I-squared, reported for each forest plot. This provides an indication of how well the pooled effect represents the sample of studies in the analysis. As expected, given the variation in samples, interventions, countries, and design methods, the variability in effect size across studies is often large. Some of these heterogeneity effects are discussed in Section 4.9. Given the wide range of potential sources of heterogeneity, especially the differences in the nature of the interventions, we do not interpret the heterogeneity statistics specifically in quantitative terms, although we do use moderator analysis to explore possible reasons for heterogeneity.

4.5. OVERALL INTERVENTION EFFECTS

In this section, we report the effect of locating decision-making within schools on student learning and other proximal outcomes.

Although the included studies reference a range of outcomes, it was only possible to identify the necessary data for calculating pooled effect sizes across more than one study for six outcomes: drop-out, repetition, teacher attendance, and student learning in relation to math, language and aggregate test scores. For these outcomes, we report the pooled effect (a weighted average effect using random effects analysis, weighted using the inverse variance method) of locating decision-making within schools. Forest plots are provided in each case, which include data on the time elapsed between baseline and endline data collection

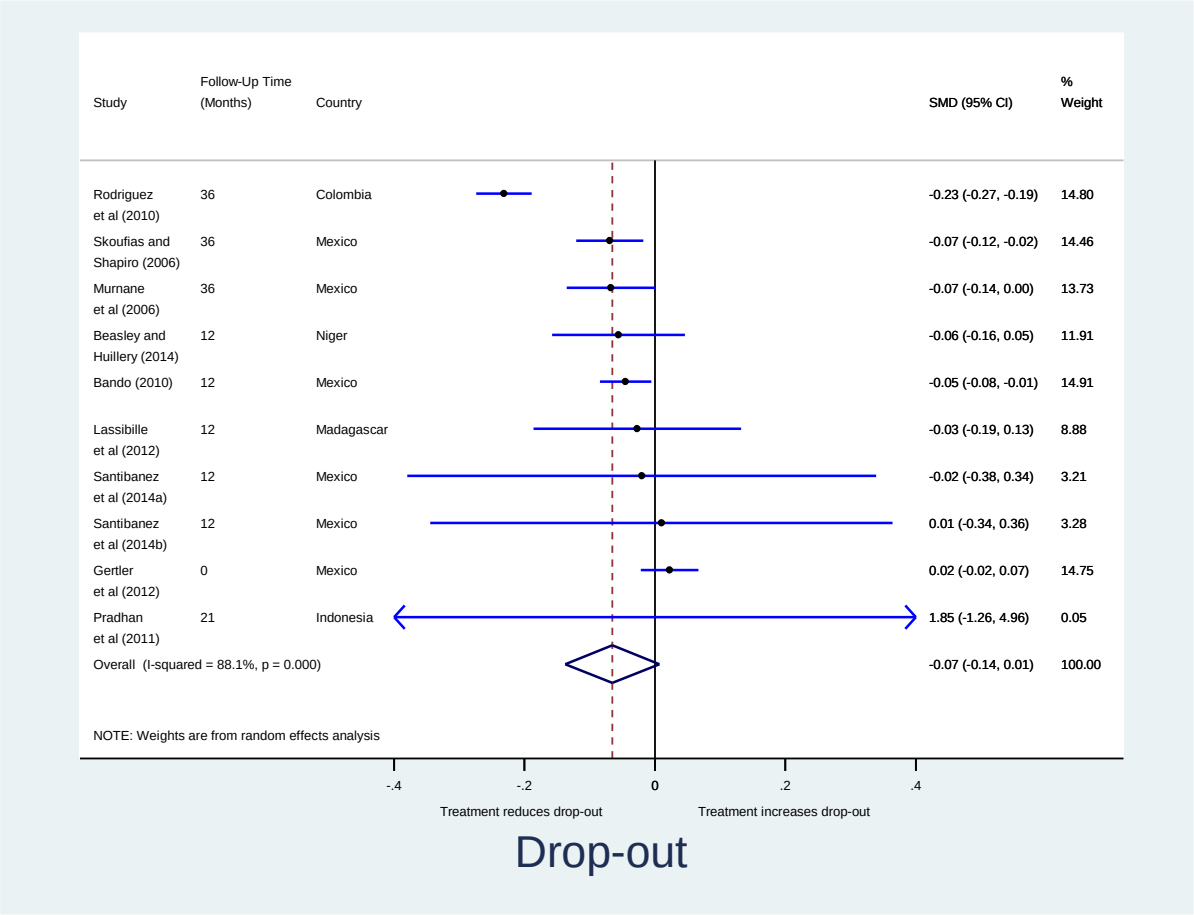
(labelled follow-up time) and the weighting of each study in the calculation of the pooled effect size. Confidence intervals shown are for the 95 percent confidence level (95% CI). Studies that include more than one independent sample are labelled separately, as in the case of Santibanez et al. (2014a) and Santibanez et al. (2014b); details of the sub-samples are provided in Supplement Table 2. Additional outcomes are discussed narratively in Section 4.5.5.

4.5.1. Student drop-out

Figure 3 presents the results for ten studies that measure the impact of a school-based decision making intervention on **school-level student drop-out rates**. Seven of the ten estimates are from Latin America; there is no obvious pattern by date of publication. All except two of the ten estimates are negative and two are statistically significant (in Colombia and Mexico). None is positive and significant.[9] Taking into account the confidence intervals, the overall estimate is negative at -0.07 SMD, but not statistically significant at 95 percent confidence (95% CI = -0.14, 0.01). However, there is significant heterogeneity in the findings across studies (I-squared = 88%) and evidence in some contexts does suggest statistically significant reductions in drop-outs. Rodriguez et al. (2010) provide the largest negative estimate from Colombia (-0.23 SMD; 95% CI = -0.27,-0.19). As a negative result is the desired result for this outcome, this suggests a beneficial impact on drop-out in some circumstances.

[9] Note that a negative result is the desired result for this outcome

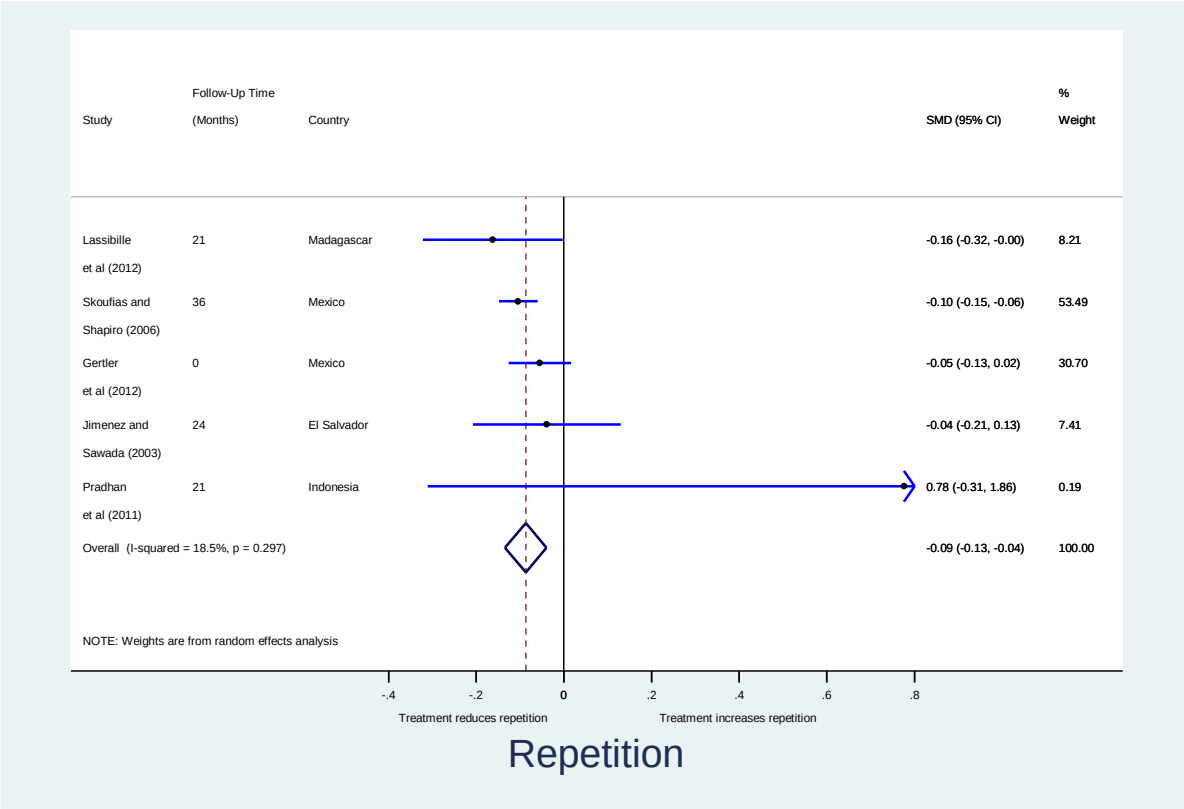
Figure 3: Main effects on student drop-out (n=10)



4.5.2. Repetition

Figure 4 reports results from five studies that measure the impact of a school-based decision making intervention on **school-level repetition rates**. Three of the five estimates are from Latin America, one is from Madagascar and one from Indonesia; there is no obvious pattern by date. Taking into account the confidence intervals, the overall estimate is negative and significant at -0.09 SMD (95%CI = -0.13, -0.04); and all but one of the individual study estimates are negative, while only two in Madagascar and Mexico are significant at the 95 percent level. The analysis of heterogeneity does not suggest it is significant across studies (I-squared = 18%), suggesting the findings are consistent across contexts. Due to the limited number of studies, we do not conduct further analysis of heterogeneity. As a negative result is the desired result for this outcome, this suggests a beneficial impact on repetition.

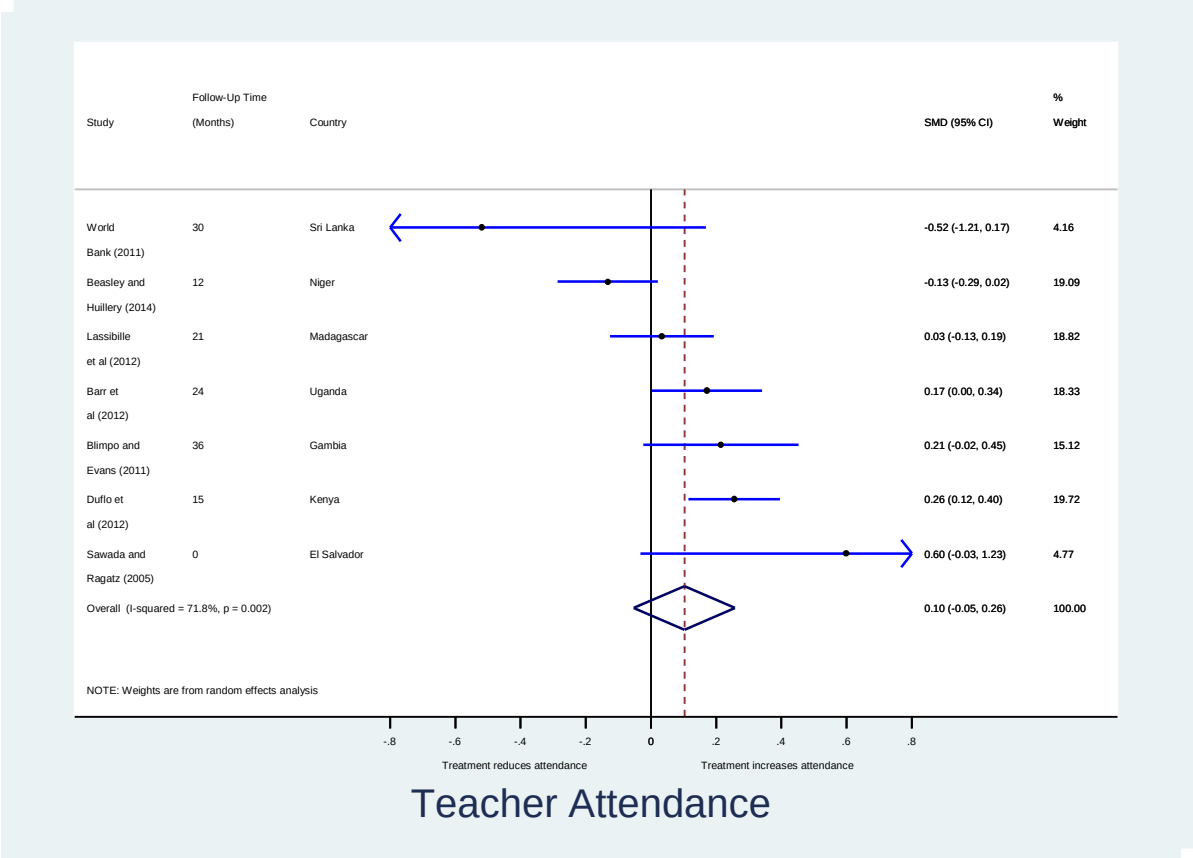
Figure 4: Main effects on repetition (n=5)



4.5.3. Teacher attendance

Figure 5 reports results from seven studies that measure the impact of a school-based decision making intervention on **teacher attendance**. Five estimates are from Africa and one each is from Latin America and Asia. There is no obvious pattern by date. Taking into account the confidence intervals, the overall estimate is positive at 0.1 SMD but is not statistically significant (95% CI = -0.05, 0.26). Analysis suggests there is significant heterogeneity in the estimates (I-squared = 72%), which is explored further in section 4.6. Indeed, two studies in Kenya and Uganda found significantly positive effects on teacher attendance.

Figure 5: Main effects on teacher attendance (n=7)



4.5.4. Student learning

Figure 6 presents the first set of results relating to student learning. The studies employ samples from a variety of school grades, indicated in Supplement Table 2. Here, we report results from sixteen studies that measure the impact of a school-based decision making intervention on **student math test scores**. The 19 estimates come from a range of contexts (Africa, Asia and Latin America); there is no obvious pattern by date. Only one estimate is negative and significant, while five, from a variety of contexts, are positive and significant – SMD exceeds 0.2 in Sri Lanka, Kenya and the Philippines. Taking into account the confidence intervals, the overall estimate is positive and significant at 0.08 SMD (95% CI = 0.02, 0.13). Significant heterogeneity in effects (I-squared = 69%) suggests that further moderator analysis is needed to explain differences between studies (as discussed in in section 4.6).

Figure 6: Main effects on maths test score (n=19)

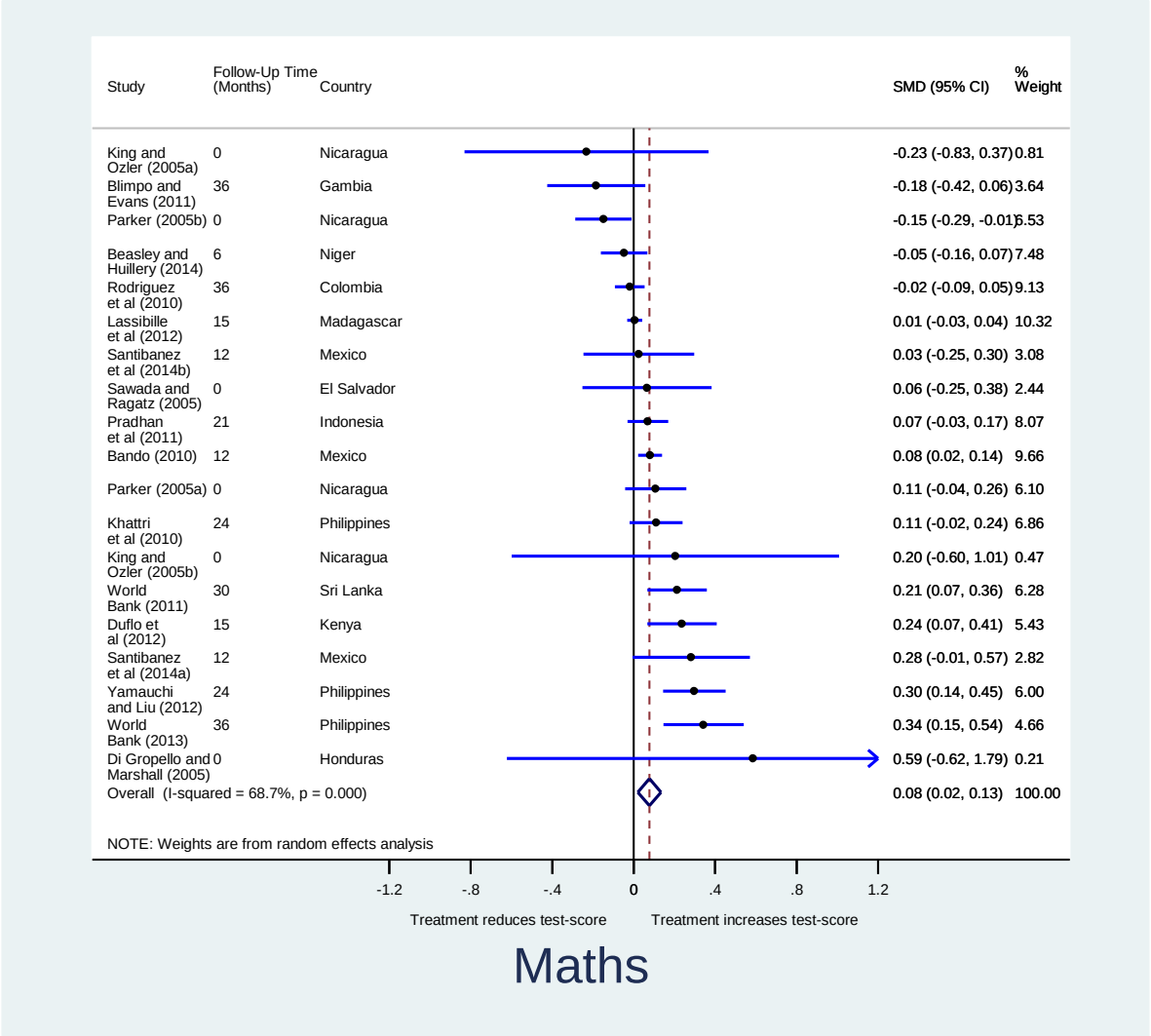


Figure 7 reports results from 14 studies that measured the impact of a school-based decision making intervention on **student language test scores**. Some studies report test data for more than one language. The languages tested are shown in Supplement Table 2, which are usually the language of instruction in school, where available. The 17 estimates come from Asia, Africa and Latin America; there is no obvious pattern by date. Taking into account the confidence intervals, the overall estimate is positive and significant at 0.07 SMD (95% CI = 0.02, 0.13); six of the 17 estimates are positive and significant, with SMD exceeding 0.2 in Indonesia, Kenya, Sri Lanka and one Mexico study, while none is negative and significant. The analysis suggests significant residual heterogeneity (I-squared = 62%), which is explored further in moderator analysis below (section 4.6).

Figure 7: Main effects on language test score (n=17)

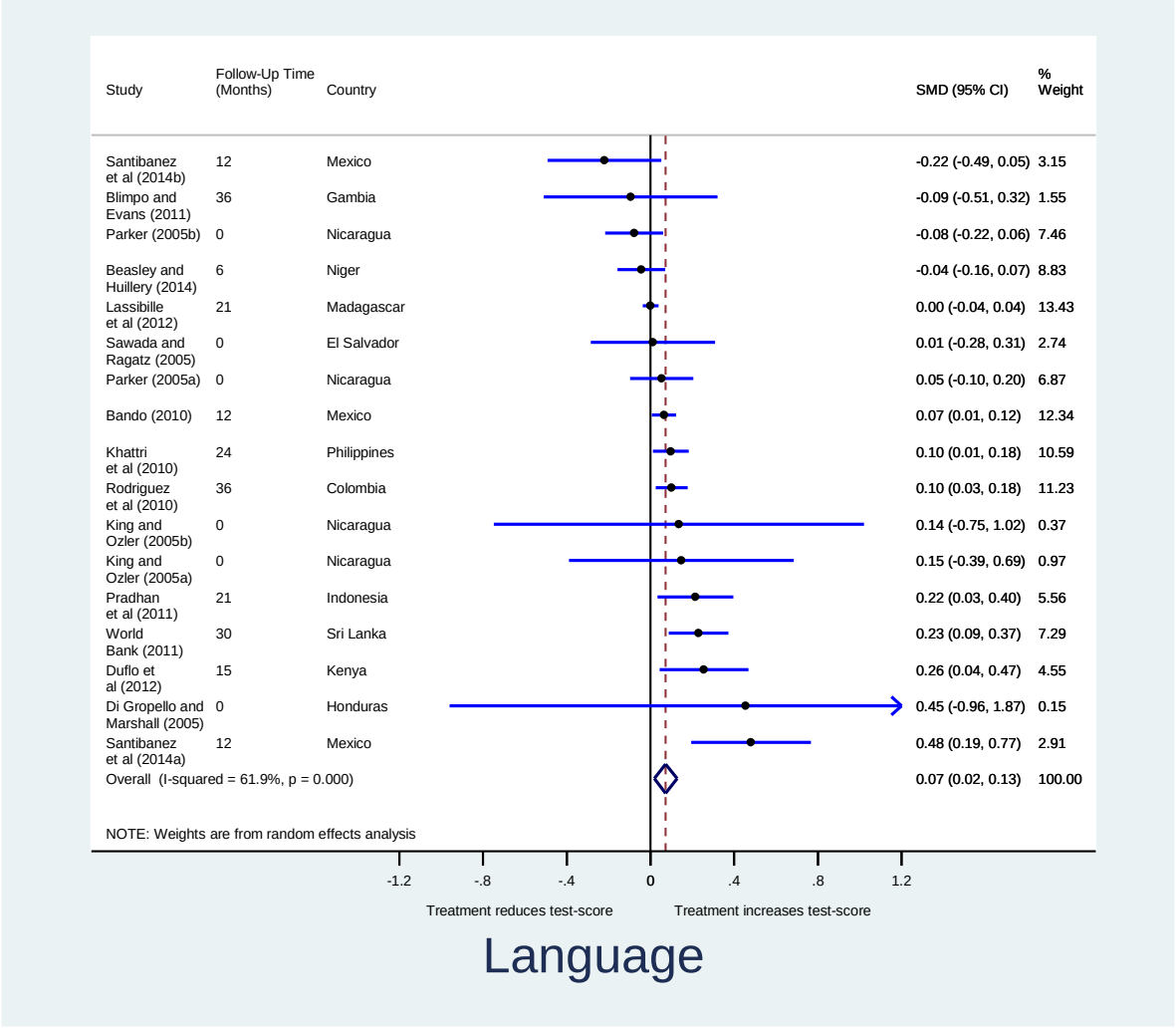
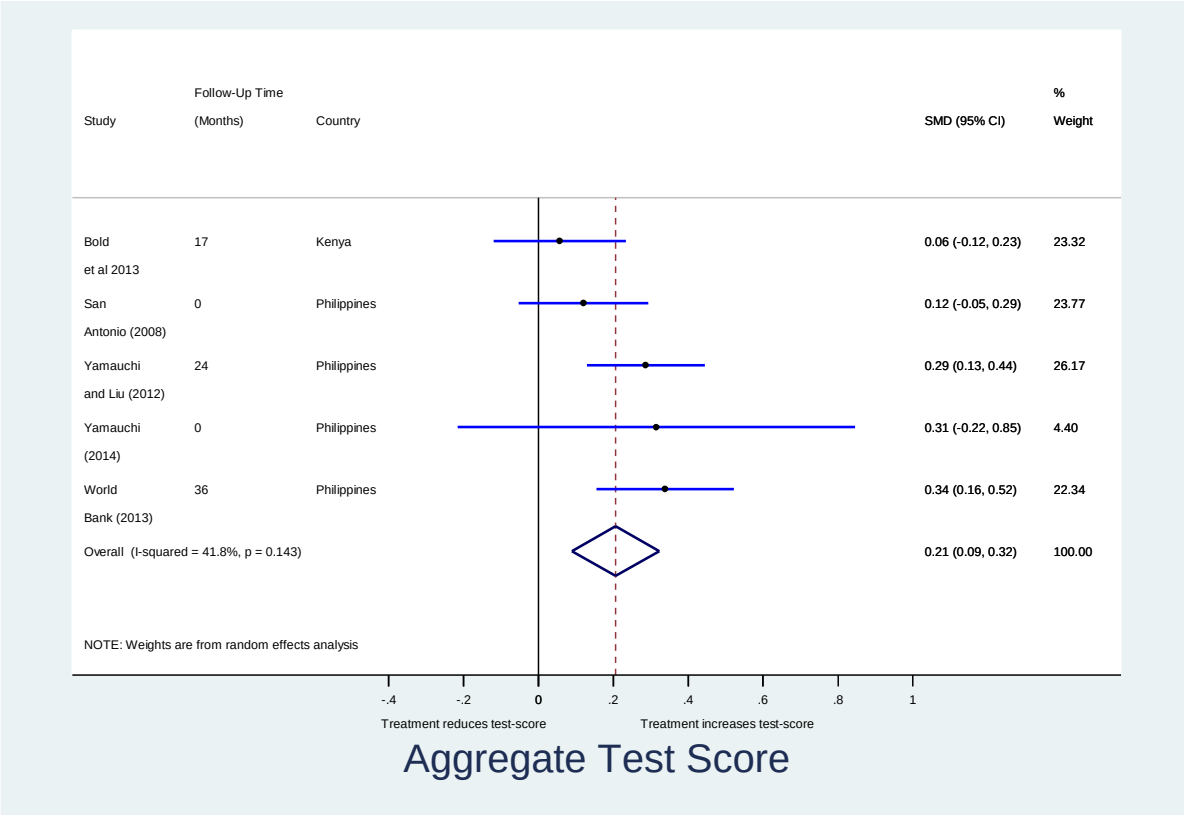


Figure 8 reports results from five studies that measured the impact of a school-based decision making intervention on aggregated student test scores.[10] The five estimates come from two countries (one from Kenya and four from the Philippines, all of which use the same test data); there is no obvious pattern by date. Two are positive and significant (both in the Philippines) with SMD around 0.3, and none is negative and significant. Taking into account the confidence intervals, the overall estimate is positive and significant at 0.21 SMD (95% CI = 0.09, 0.32). There is some residual heterogeneity (I-squared = 42%) although not significant. Due to the limited number of studies, we do not conduct further analysis of heterogeneity for this outcome.

[10] Aggregated tests are a multi-subject tests. The National Achievement Test in the Philippines comprises math, English, Filipino, science, and social science. The test used in Bold et al. (2013) covers only math and English.

Figure 8: Main effects on aggregate test score (n=5)



4.5.5. Other outcomes

In addition to the six outcomes discussed above, the included studies also report effects on student attendance, student failure and student progression. However, none of the studies include sufficient data to allow for the calculation of standardised mean differences in relation to these additional outcomes. We therefore present the results relating to these outcomes narratively.

Student absenteeism and attendance

Six of the studies consider impact on student absenteeism or attendance (Barr et al., 2012; Blimpo & Evans, 2011; Di Gropello & Marshall, 2005; Jimenez & Sawada, 1999; Lassibille et al., 2010; and Sawada & Ragatz, 2005).

Two of the studies measure absenteeism by collecting data on student attendance on the day of an unannounced visit to a school. Both of these suggest a positive effect on attendance. Barr et al. (2012) estimate that the additional impact of using a participatory process for developing and using a school report card ranged from 8 to 10 percent (with different statistical specifications), while Blimpo and Evans (2012; Table 13, p. 42) estimate that the Whole School Development intervention reduced student absenteeism by about 5 percentage points from a base of about 23 percent.

Another two studies define absenteeism as the number of days absent in the previous month. Both look exclusively at students in the 3rd grade. These studies are less positive in their assessment of impact on absenteeism. Jimenez and Sawada (2003; p437) found that a student in an EDUCO school was less likely to be absent after holding constant household, school, and participation characteristics. However, they found possible evidence of a Hawthorne effect on this outcome as differentiation by year found that the EDUCO effect was stronger for newer EDUCO schools. Sawada and Ragatz (2005; p. 297) identify no difference between EDUCO and traditional schools in overall mean of absence.

In addition to these pairs, two other studies investigate absenteeism in unique ways. Di Gropello and Marshall (2005), who use a student reported ordinal measure of attendance, find no evidence that PROHECO schools succeeded in reducing student absences. Lassibille et al. (2010; Table 3, p. 318), meanwhile, measure attendance across a given school during the month prior to a visit. Their study does appear to identify some effect of school-based decision-making on attendance, as they identify an increase in attendance of approximately 4 percentage points over the control, in schools which benefited from interventions at the school level. No significant effect was identified within the districts implementing only the sub-district- and district-level version of the intervention.

Student failure

Five studies investigate impact on student failure rates (Bando, 2010; Gertler et al., 2012; Murnane et al., 2006; Rodriguez et al., 2009; Skoufias and Shapiro, 2006). However, in none of these studies is failure precisely defined, in terms of which subjects are included in the assessment of a student's failure at the end of a year. Although it is probable that, in Latin America, these will include Spanish, Mathematics and Science, we do not know the relative weights given to each subject.

Closer inspection suggests that only two of the studies are likely to have used equivalent definitions (Murnane et al., 2006; and Skoufias and Shapiro, 2006). Both of these studies investigate the PEC programme in Mexico, and both define failure as the number of students who did not pass a given grade in a given school year as a proportion of the total number of students who were enrolled at the end of that year.

On the surface, the studies identify contrasting results, as Skoufias and Shapiro (2006) found that participation in PEC reduced failure rates by 0.24 percentage points, while Murnane et al. (2006) found no statistically significant impact of PEC participation on student failure rates. However, these findings should not be compared in isolation, as Murnane et al. go on to identify a number of reasons why their null finding could actually be considered evidence of a positive effect. Unlike Skoufias and Shapiro, Murnane et al. attempted to explicitly consider differences in trends prior to the implementation of the PEC intervention. Their analysis of these prior trends identified a significant difference in failure rates between schools that did and did not ultimately join the PEC programme. Given these prior differences, they suggest that their null finding regarding impact on failure could

actually be perceived as evidence of success of the programme, as one could argue that it was a significant accomplishment for PEC schools not to lose ground relative to non-PEC schools in student failure rates. Furthermore, the same authors also identified a positive impact on drop-out within PEC schools. The implication of such a finding is that PEC schools were more successful in retaining many students who may have been relatively low-achieving, which would have an inevitable impact on overall failure rates.

Bando (2010) also investigates the PEC programme, but she uses census data in her analysis. Although the census definition of failure is not explicitly specified in her study, it must differ from the definition used by the other studies discussed above, as they identify an overall failure rate of approximately 5 percent, whereas Bando identifies an average failure rate of roughly 20 percent. Bando's results suggest a positive association with failure rates; she also indicates that the effect on failure rates strengthens over time.

Two other studies consider student failure. Gertler, Patrinos and Rubio-Codina (2012) – also in Mexico but in reference to AGE, the precursor of PEC - show a significant reduction in grade failure, a finding which is robust to checks on pre-intervention trends between treatment and comparison schools. Rodriguez et al. (2009; p.420) also find a significant effect on failure, as they identify a reduction of an additional 1.4 percentage points in the PER schools as compared to the control schools.

Student progression and continuation

Two studies investigate impact on student progression and/or continuation (Barr et al., 2012; Jimenez & Sawada, 2003), and these offer discrepant findings. Barr et al. (2012) found no impact on the probability of continued enrolment, as a result of the participatory scorecard intervention. However, in their analysis, Jimenez and Sawada (2003) identify an association between being in an EDUCO school and a greater probability of continuing in school.

4.6. EXAMINATION OF HETEROGENEITY: MODERATOR ANALYSIS

In this section, we present analyses for three moderating variables which are likely to affect the impact of school-based decision-making reforms: the level of decentralisation (high, medium or low); the country income level; the type of evaluation method used (with or without randomised assignment). In each sub-section, we present separate forest plots for the four outcomes with a sufficient number of estimates to allow for disaggregation and where statistical tests suggested heterogeneity was significant (i.e. drop-out, teacher attendance, math test score, and language test score).[11] In many cases, our moderators demonstrate the differences in effects, and hence reduce the residual heterogeneity across studies. For the most part, however, we are unable to draw conclusions concerning heterogeneity of treatment effects by moderating variable owing to the relatively small

[11] As noted above, the statistical analysis for two outcomes (repetition and aggregate test score) which had small numbers of available observations suggested that heterogeneity across studies was not significant.

number of studies in each group and the potential effects of correlated sources of heterogeneity – for example when moderating by income level, differences in study quality and intervention type also affect results in the various categories. Nonetheless, we draw out indicative patterns while remaining cautious in our interpretation.

4.6.1. Broad intervention type

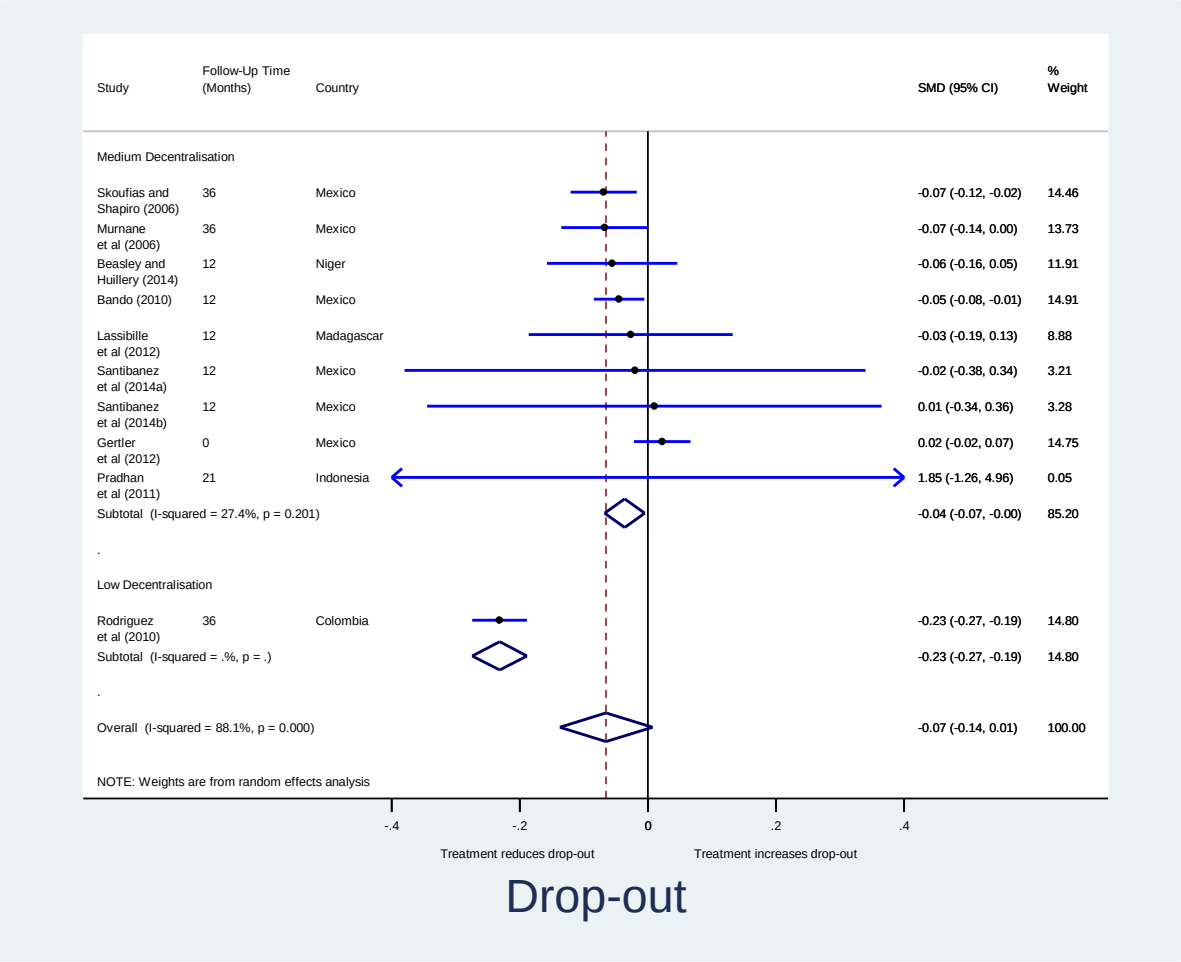
This section presents the results by outcome when broken down by broad intervention type (as discussed in Section 4.2).

Drop-out

We are not able to draw conclusions in relation to drop-out (Figure 9), except to say that a negative and significant effect of the interventions on drop-out is found separately for medium decentralisation contexts specifically (-0.04 SMD; 95% CI = -0.07, -0.00).[12] There is only one estimate for low decentralisation contexts. It is noteworthy that, when we conduct the analysis by degree of decentralisation, the residual heterogeneity (as measured by I-squared) for medium decentralisation is statistically insignificant, while the pooled effect size is statistically significant. When pooled together, the overall effect size is not significant, while there is significant residual heterogeneity (Figure 3).

[12] A negative finding is beneficial for this outcome.

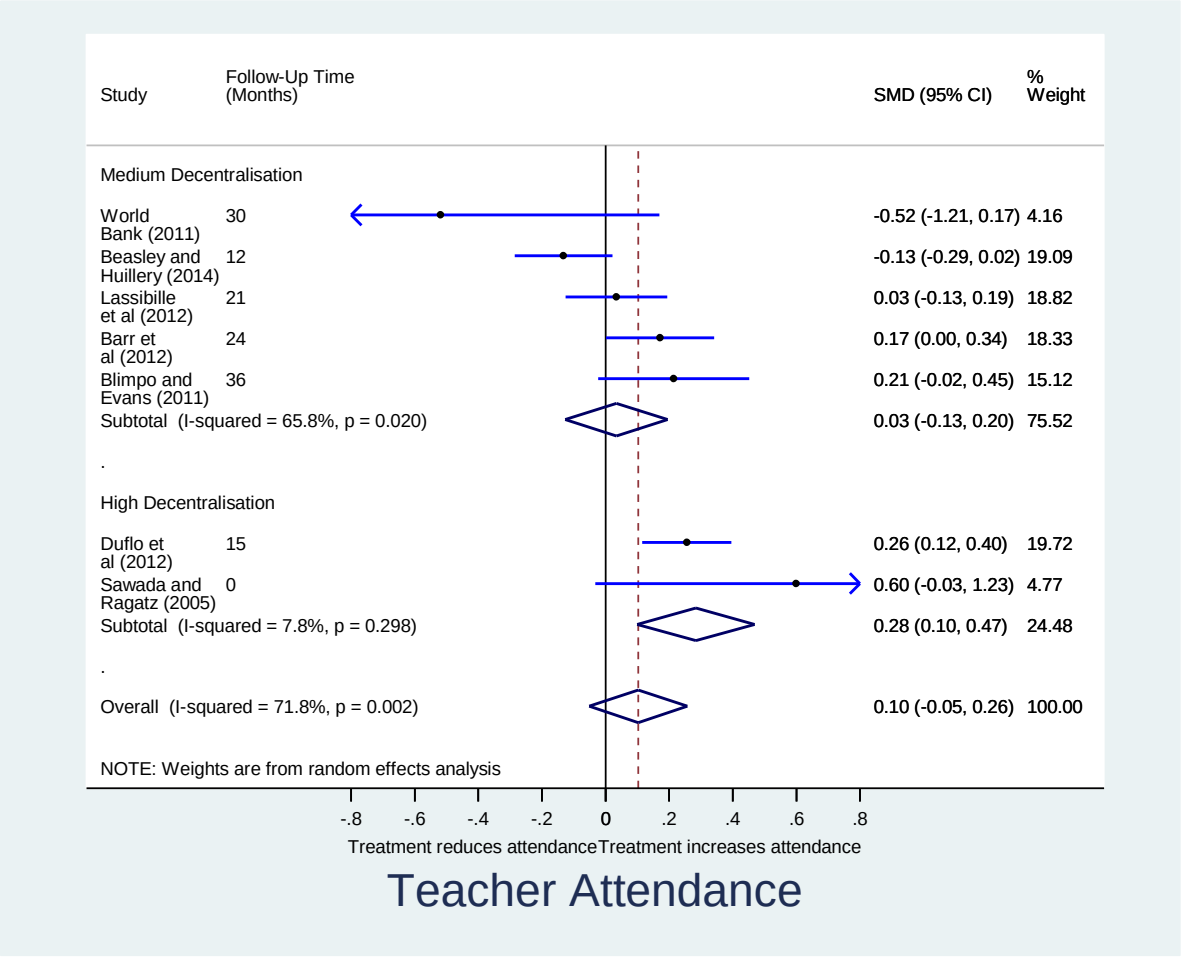
Figure 9: Effects on student drop-out by level of decentralisation (n=10)



Teacher Attendance

With regard to teacher attendance (Figure 10), while the number of studies is small, we find a strong and significant positive effect for high decentralisation studies (0.28 SMD; 95% CI = 0.10, 0.47), although this group comprises only two studies, recalling that high decentralisation includes recruitment and other personnel powers being devolved to the school. There is no evidence overall for effects on teacher attendance for medium decentralisation interventions when treated separately (0.03; 95% CI = -0.13, 0.20).

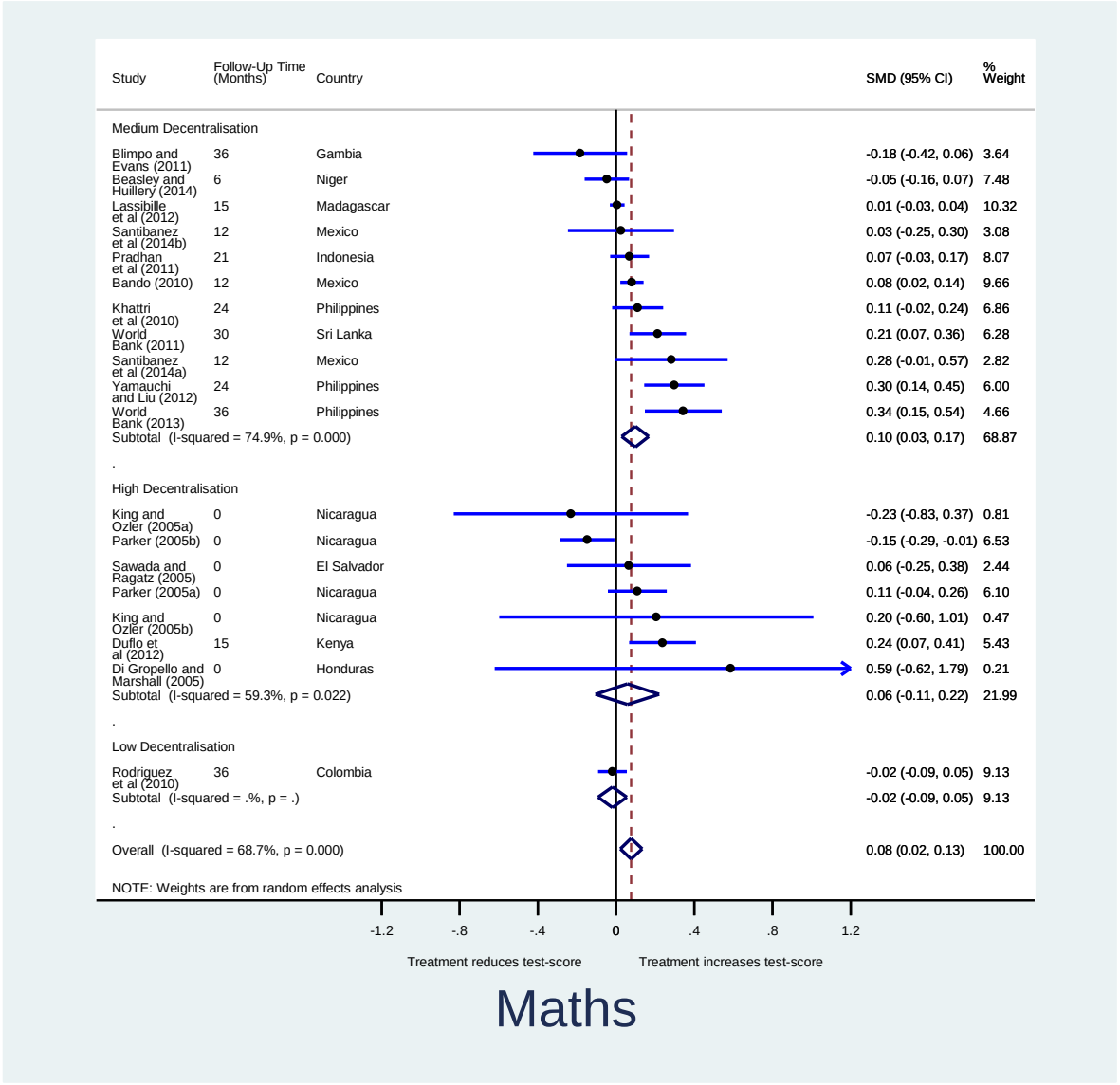
Figure 10: Effects on teacher attendance by level of decentralisation (n=7)



Student Learning

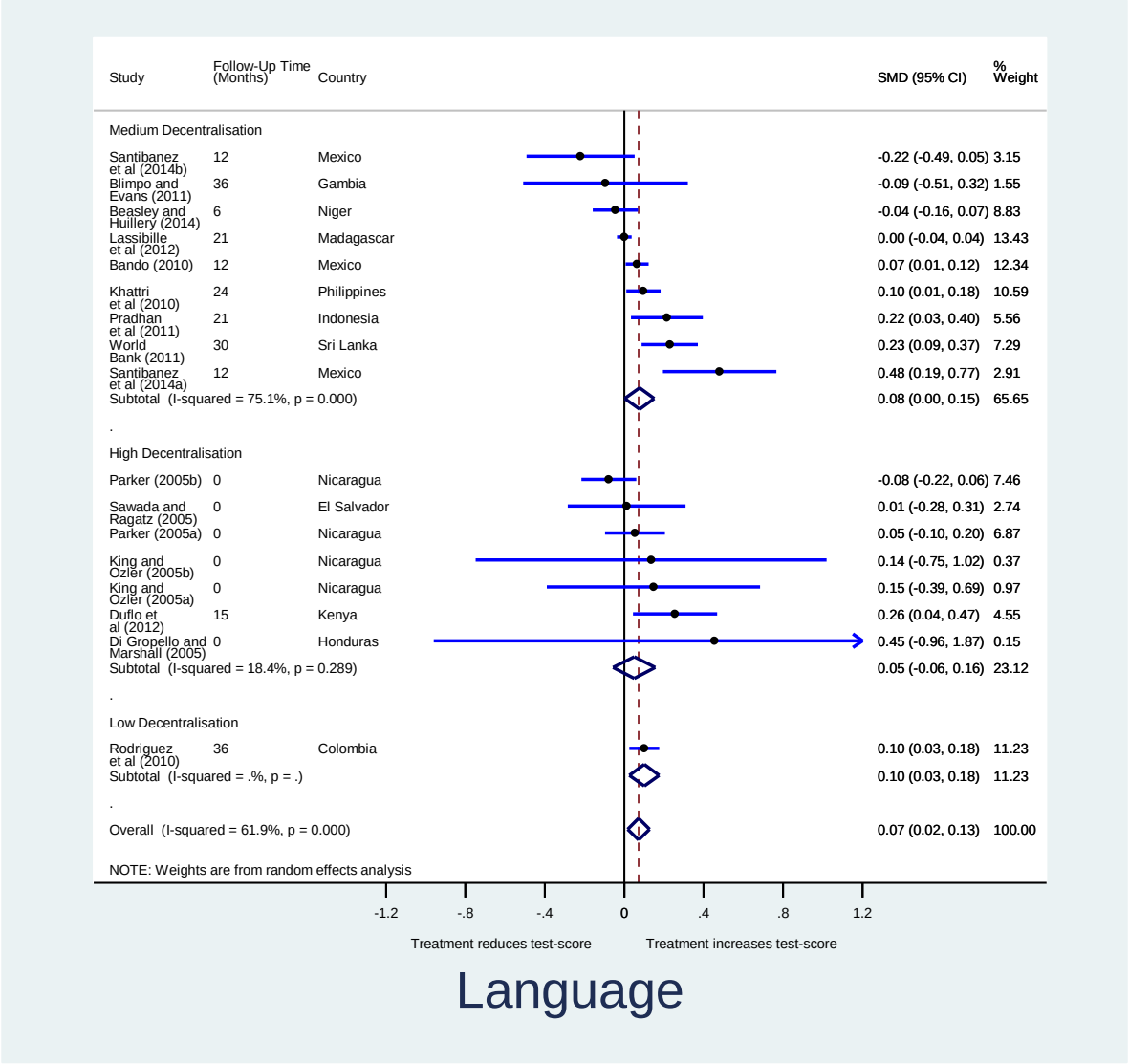
With regard to mathematics test scores, a positive pooled effect of 0.10 SMD (95% CI = 0.03, 0.17) is found for medium decentralisation interventions only when treated separately (Figure 11). However, there is residual heterogeneity in the effect sizes across studies within this category. The pattern among high decentralisation contexts is more mixed, without a significant pooled effect (SMD = 0.06; 95% CI = -0.11, 0.22), although one individual study estimate in Kenya is significantly positive (Duflo et al., 2012, which may be considered a particularly intensive treatment). There is only one study in a low decentralisation context, with no significant effect.

Figure 11: Effects on maths test score by level of decentralisation (n=19)



A very similar pattern is found for language test-scores (Figure 12). In medium decentralisation contexts, the pooled effect was estimated as 0.08 SMD (95% CI = 0.00, 0.15), although the residual heterogeneity suggests particularly large effects in some studies. In high decentralisation contexts, there is no evidence of an effect (0.05 SMD; 95% CI –0.06, 0.16) and the analysis of heterogeneity suggests that this finding is fairly consistent across studies. In addition, the one study of a low decentralisation intervention also shows a positive and significant result.

Figure 12: Effects on language test score by level of decentralisation (n=17)



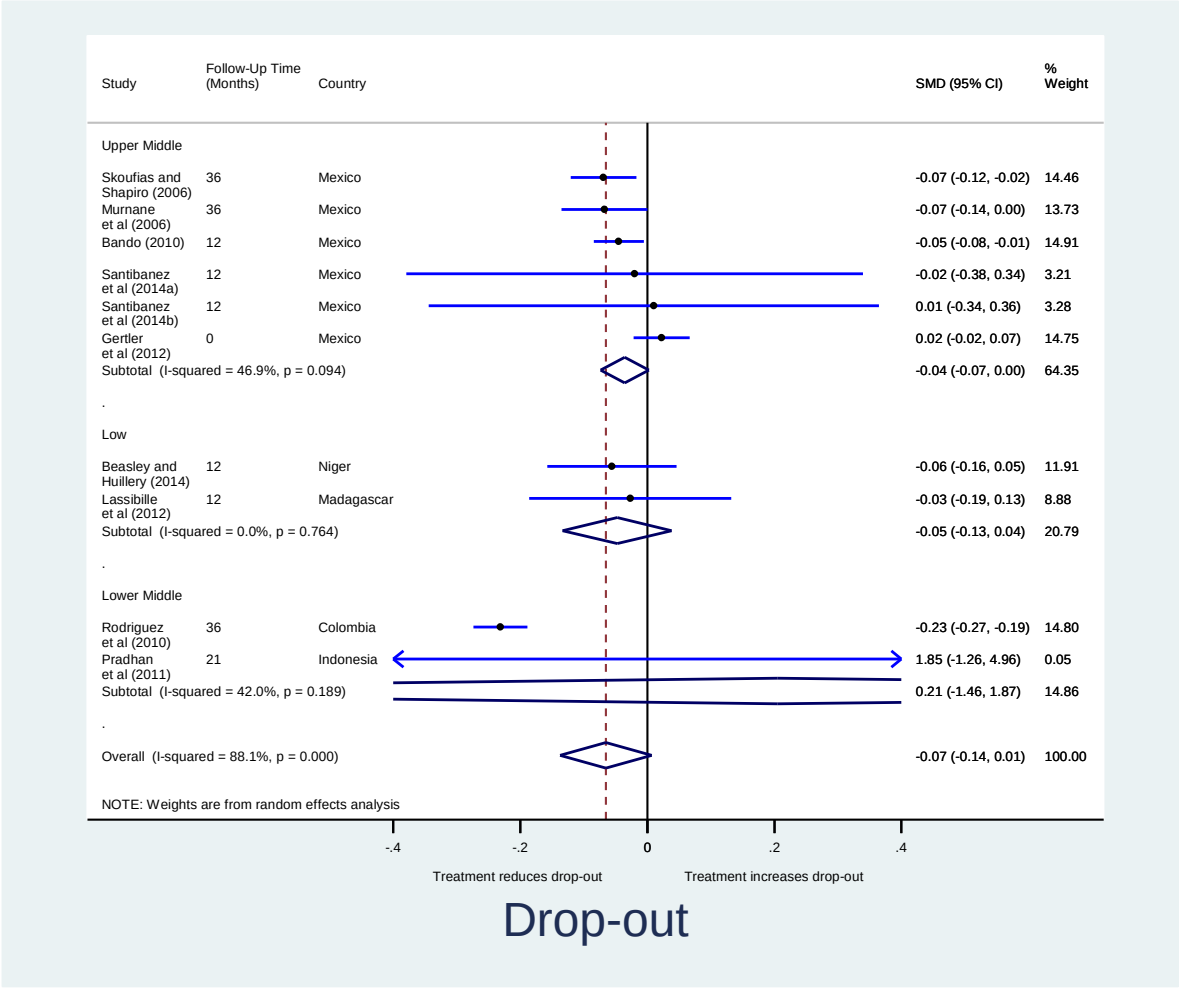
4.6.2. World Bank income classification category

This section presents the results by outcome when broken down by income level at the time of intervention.

Drop-out

In relation to the first outcome, we find no evidence that effects on drop-out differ significantly by income group, although we do find that they are negative and significant overall for the upper middle income group (0.04 SMD; 95% CI = -0.07, 0.00) (Figure 13).

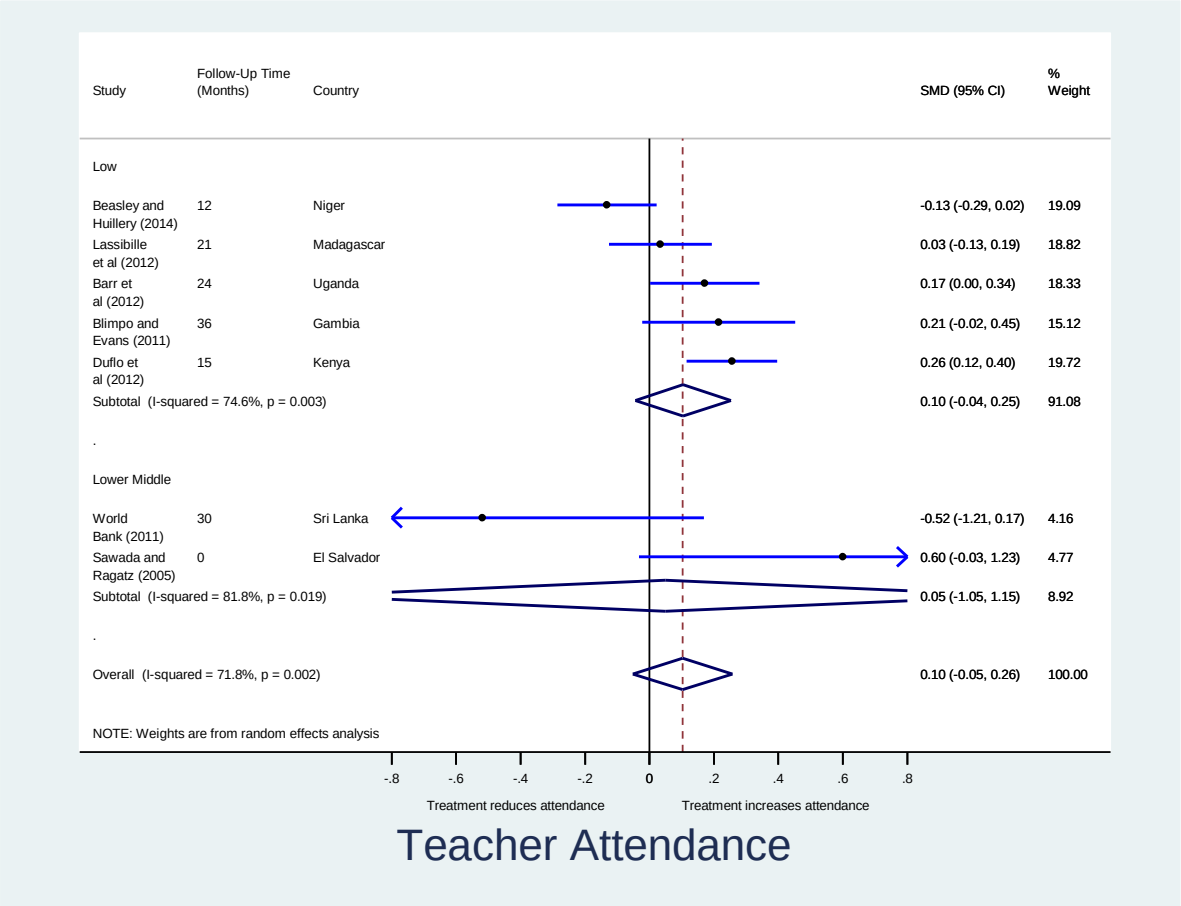
Figure 13: Effects on student drop-out by income level (n=10)



Teacher Attendance

Results for teacher attendance are dominated by studies from low income countries (Figure 14), where issues relating to teacher attendance may be particularly acute but no evidence is found for differences in effects by income group or for significant effects in each income group considered separately.

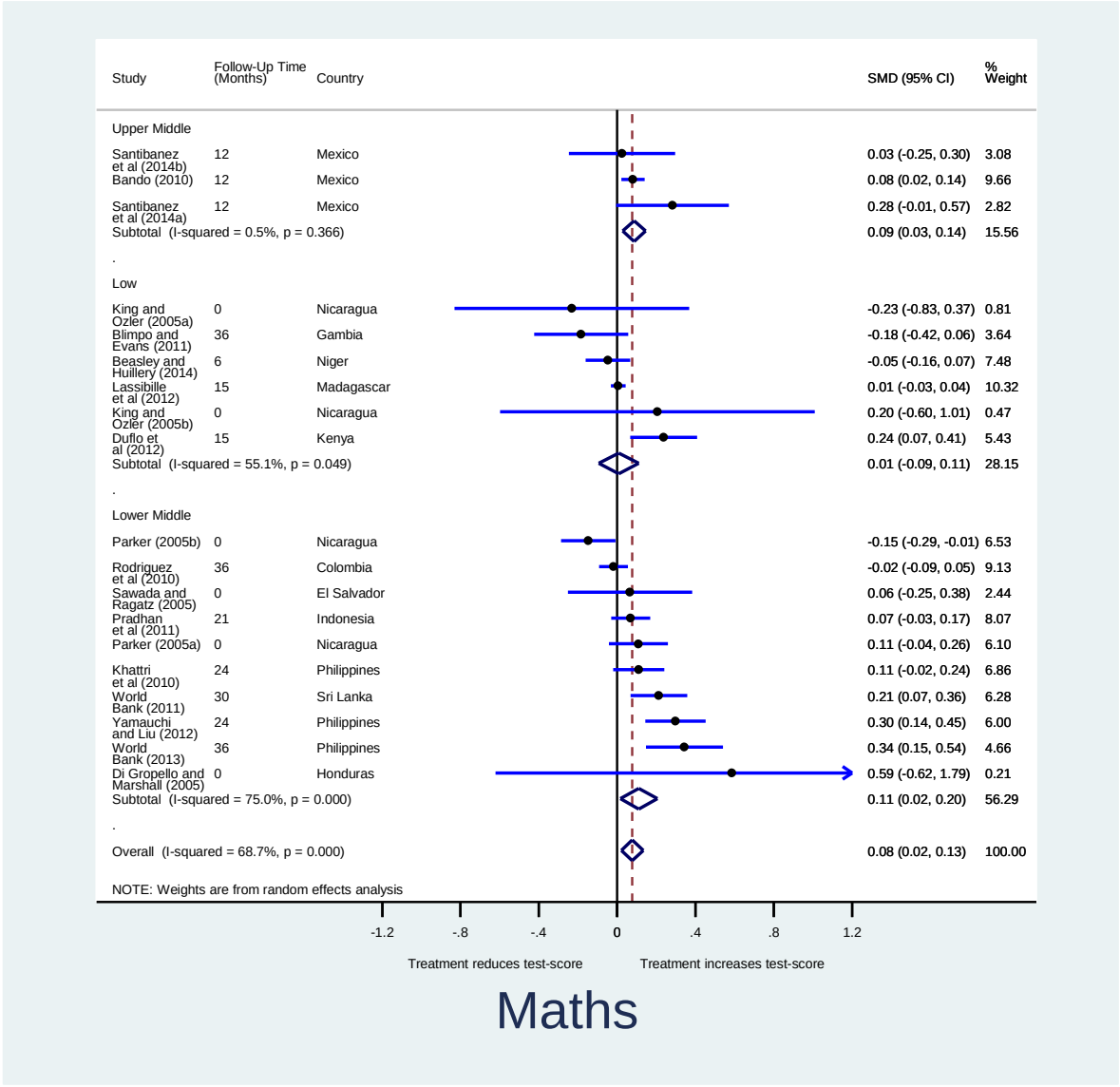
Figure 14: Effects on teacher attendance by income level (n=7)



Student Learning

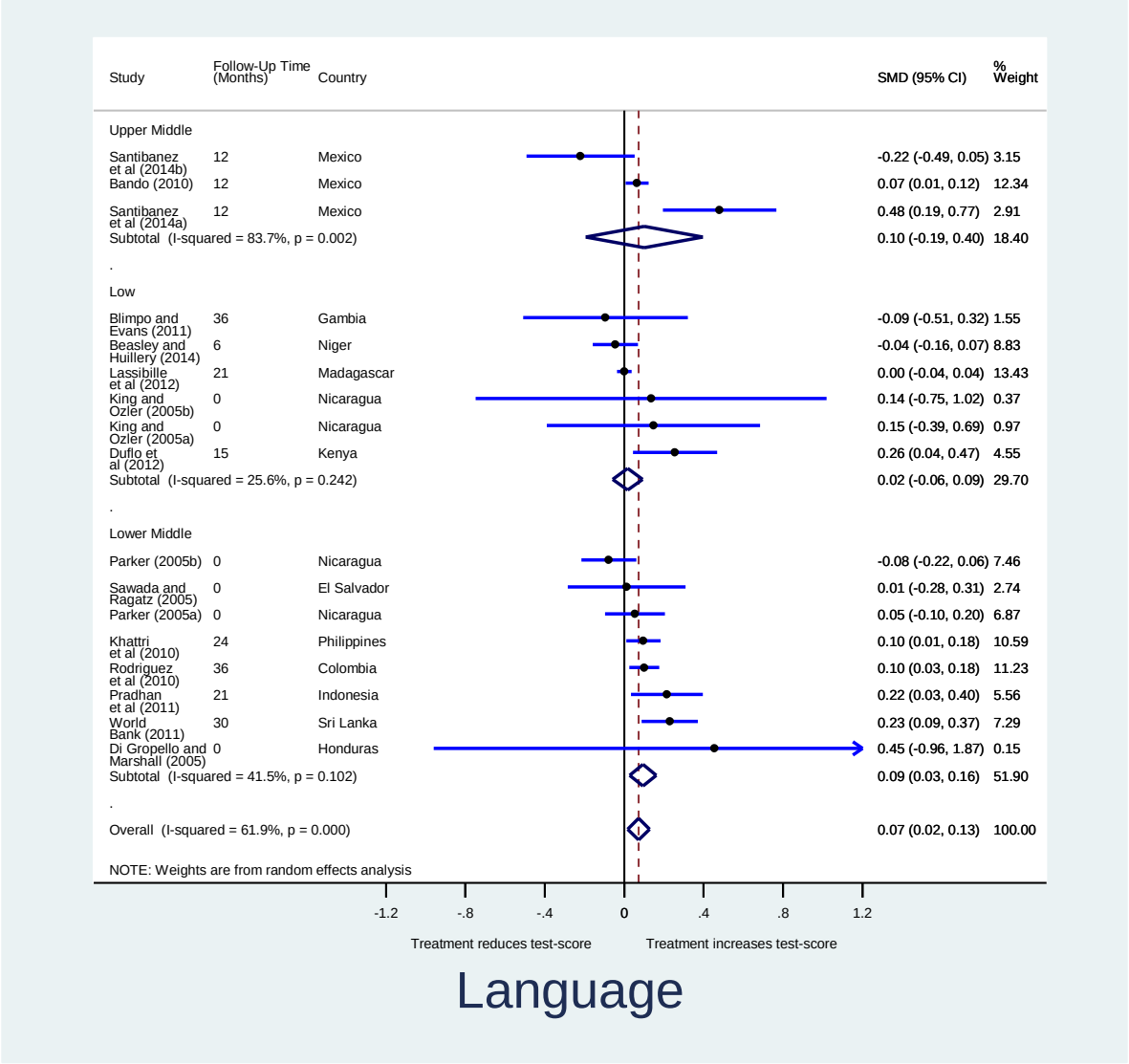
Concerning mathematics (Figure 15), the overall positive effect of the interventions on test-scores is found to be driven by the results of studies conducted in middle income countries, both upper-middle (0.09 SMD; 95% CI = 0.03, 0.14) and lower-middle (0.11 SMD; 95% CI = 0.02, 0.20). The effects are significant for both middle income countries separately. There is no evidence for significant effects overall on student learning in low income countries (0.01 SMD; 95% CI = -0.09, 0.11).

Figure 15: Effects on maths test score by income level (n=18)



This pattern is reflected somewhat with regard to test scores in language (Figure 16), while the overall positive pooled effect is driven by the results for lower-middle income countries only (0.09 SMD; 95% CI = 0.03, 0.16). Only three studies are available for upper-middle income countries, however, while the pattern of no significant effect for low income countries may be considered comparable to that for mathematics. For both outcomes (i.e. math and language), the findings in Kenya Duflo et al. (2012) are an exception to the pattern for low income countries; as noted above, these findings relate to an intervention which may be considered a particularly intensive treatment.

Figure 16: Effects on language test score by income level (n=16)



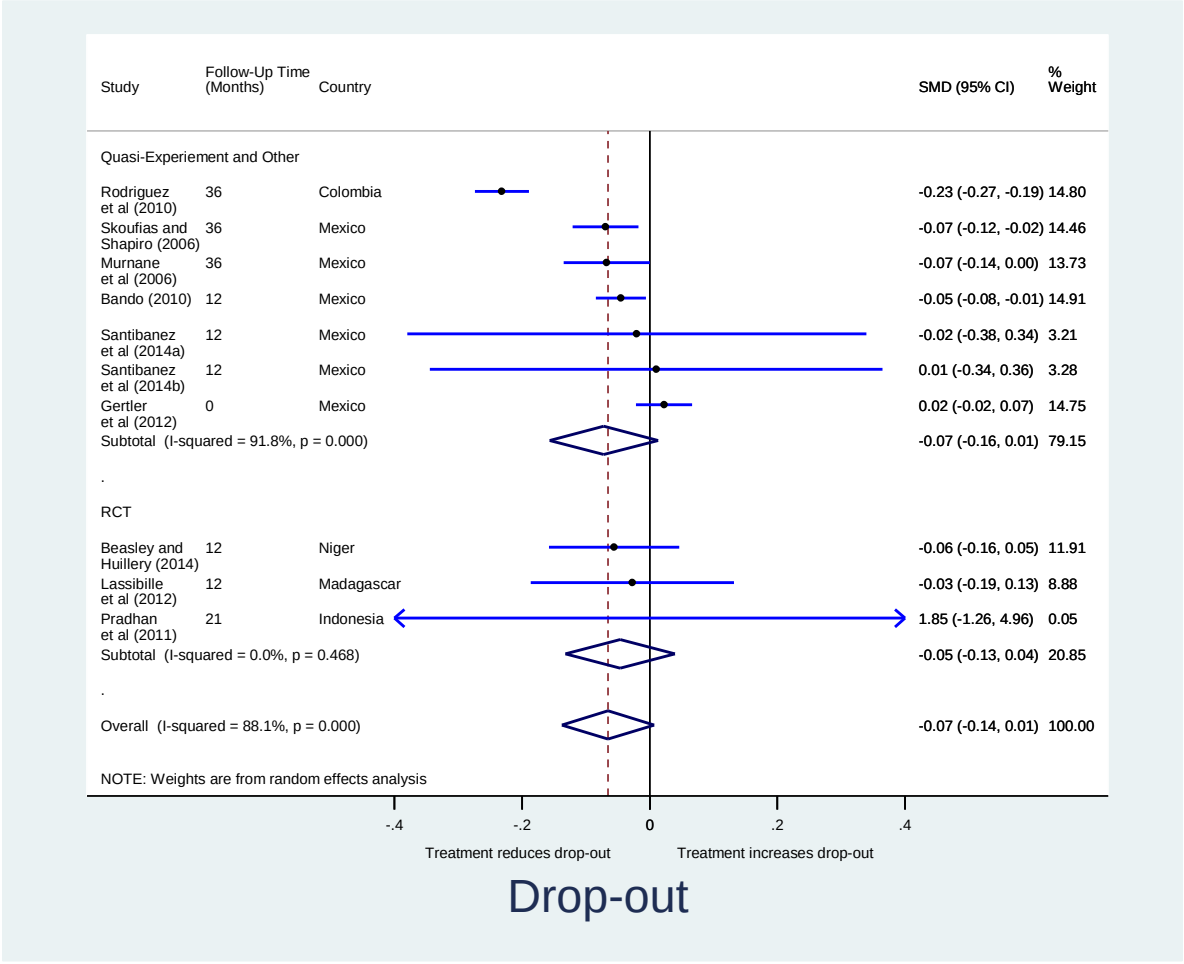
4.6.3. Type of evaluation design

This section presents the results by outcome when broken down by type of evaluation design (i.e. designs utilising randomisation versus non-randomised approaches). Within each group there is considerable diversity with respect to the actual design and methodology employed. Moreover, more recent reforms and interventions are more likely to have been evaluated using RCTs. On the basis that such interventions may in fact require several years to yield results, there may be a relationship between evaluation design, time-lag between the start of the intervention and the evaluation, and the results in terms of impact.

Drop-out

Regarding drop-out, the results for RCTs and quasi-experimental studies are somewhat similar overall, with a weakly negative – but, in part due to the small sample size, statistically insignificant – pooled effect being found for both groups of studies. No individual RCTs reported statistically significant effects on student drop-out.

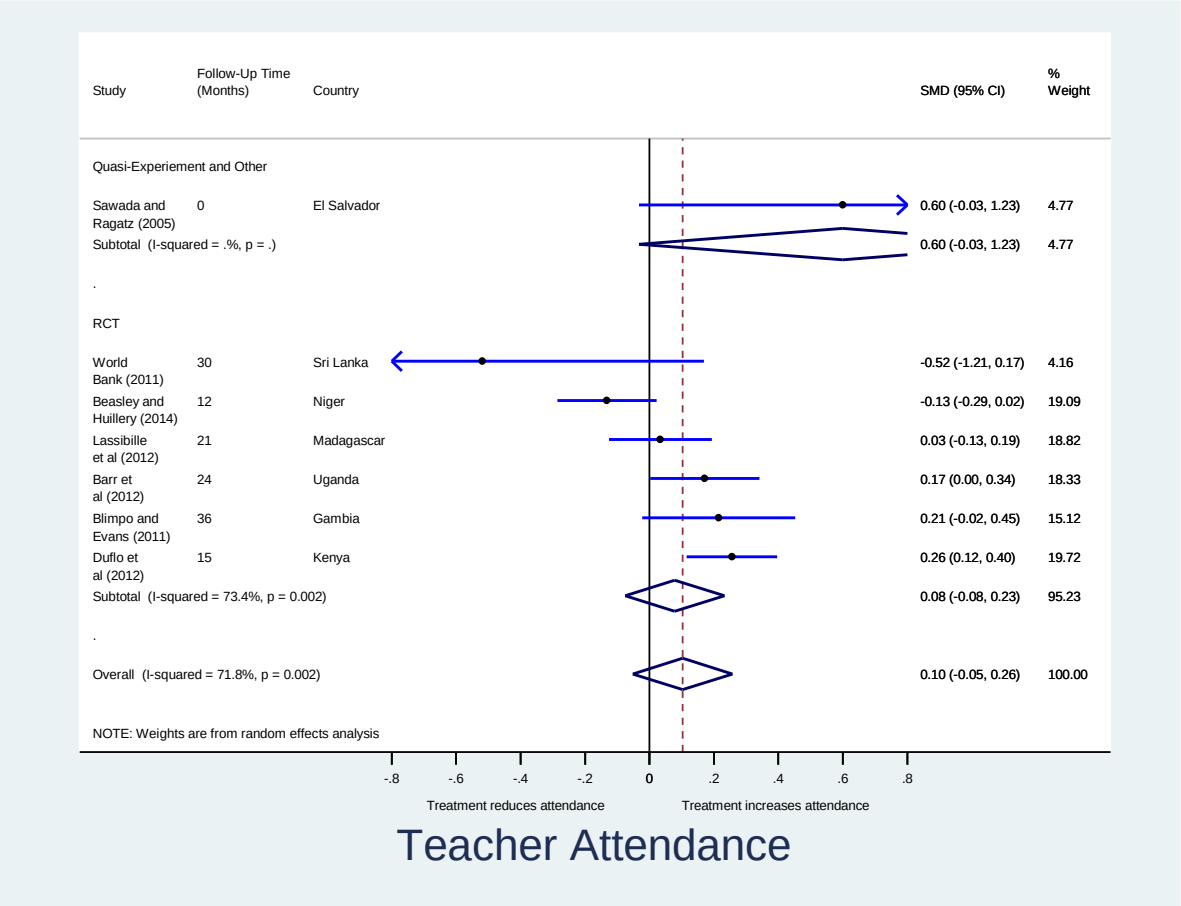
Figure 17: Effects on student drop-out by evaluation design (n=10)



Teacher attendance

All studies of teacher attendance are RCTs with one exception (Sawada and Ragatz, 2005) and the pooled result for this set of studies is consistent with the overall pooled result, suggesting a positive but statistically insignificant effect of decentralisation on teacher attendance (0.08 SMD; 95% CI = -0.05,0.26). Statistically significant findings were, however, reported in two individual RCTs, conducted in Kenya and Uganda.

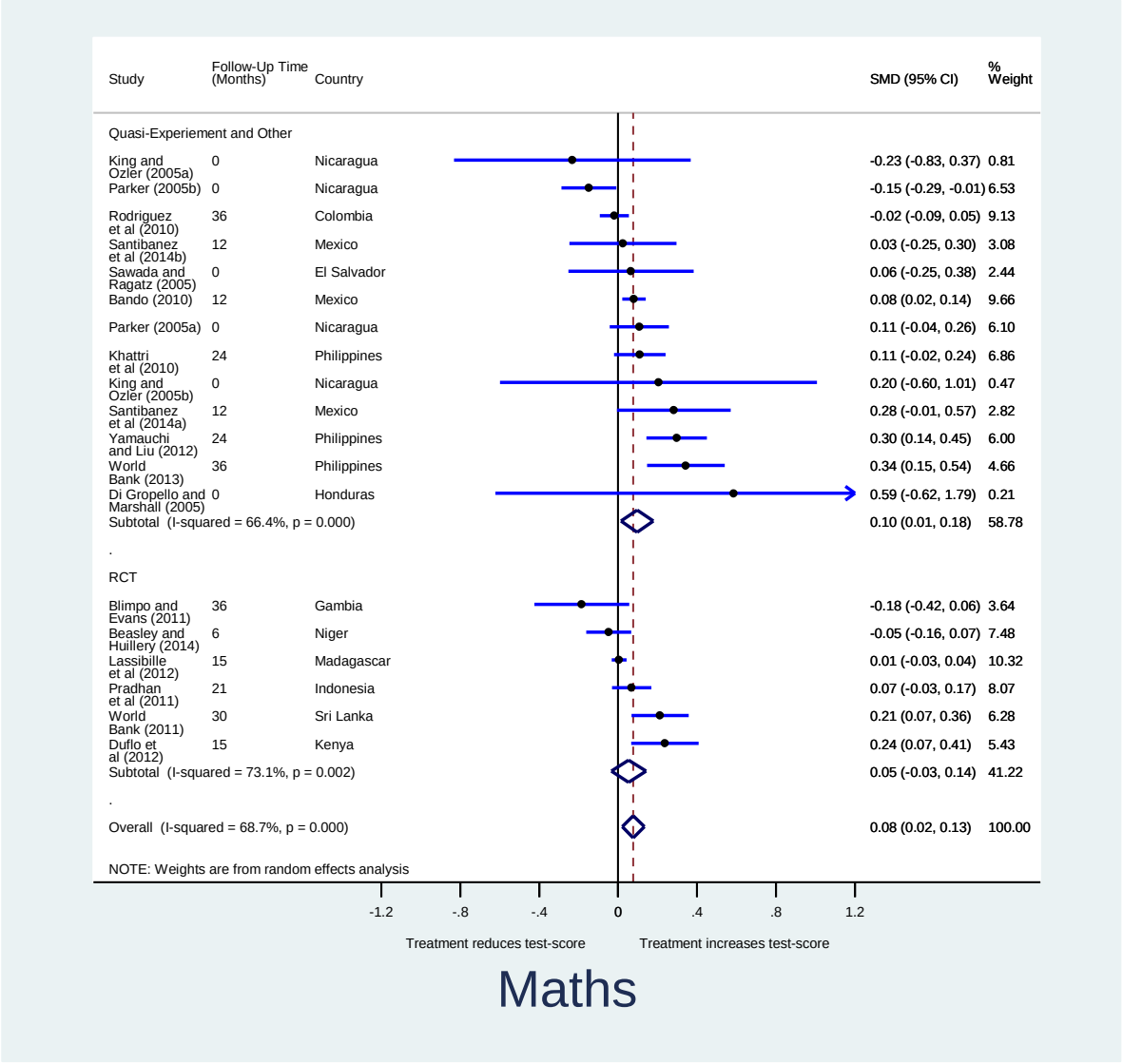
Figure 18: Effects on teacher attendance by evaluation design (n=7)



Student Learning

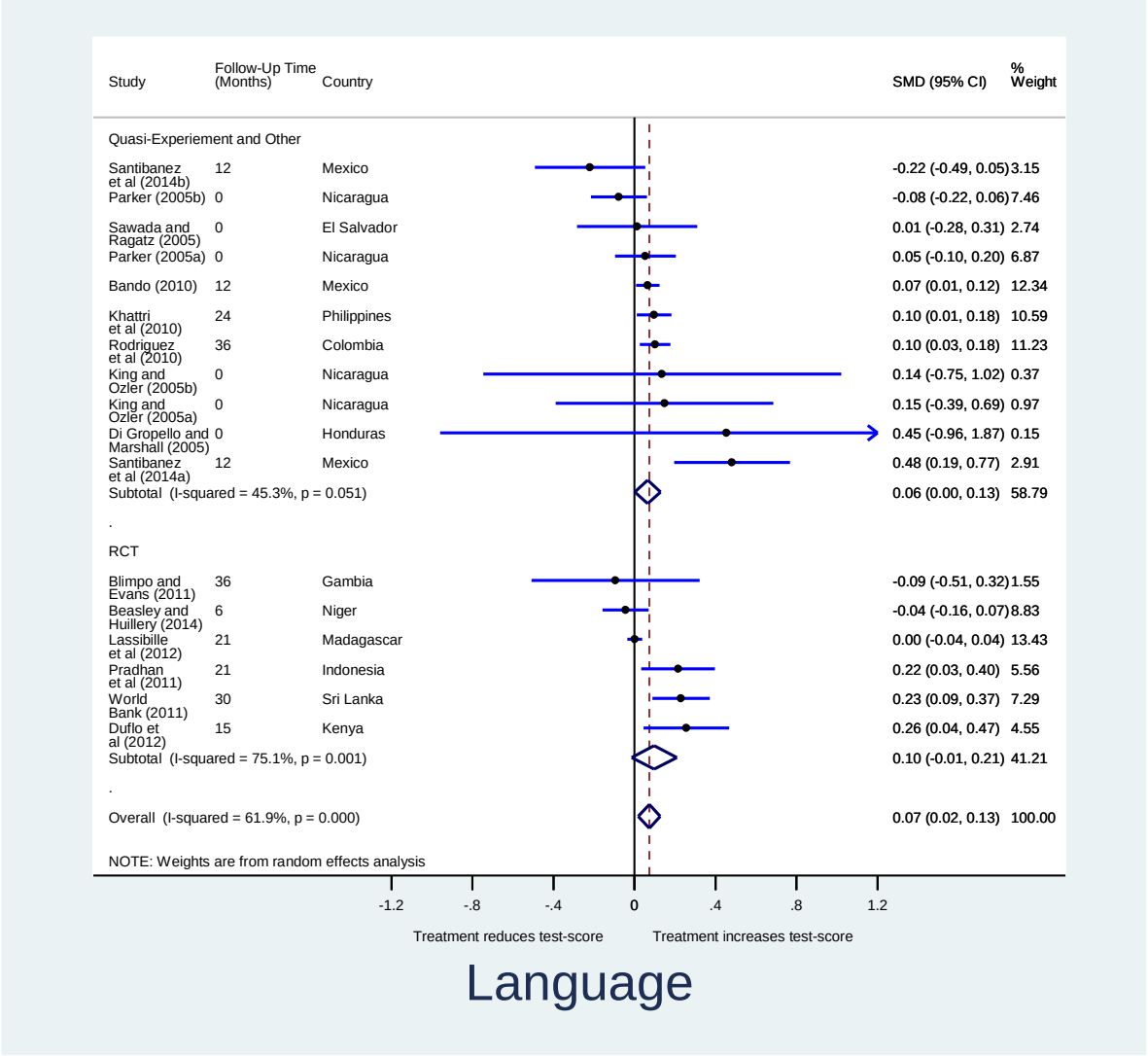
For mathematics, the significant positive pooled effect is found for quasi-experimental studies treated separately (0.10 SMD; 95% CI = 0.01, 0.18). The results from the sample of RCTs suggests smaller and statistically insignificant effects at the 95% confidence level (0.05 SMD; 95% CI = -0.03, 0.14), although two RCTs (in Kenya and Sri Lanka) do estimate significantly positive findings.

Figure 19: Effects on maths test score by evaluation design (n=19)



The pattern for language scores is very similar to that for mathematics. While the separate result for RCTs overall is marginally statistically insignificant (0.10 SMD; 95% CI = -0.02, 0.21), there are three RCTs which do estimate statistically significant effects on language tests in Indonesia, Kenya and Sri Lanka.

Figure 20: Effects on language test score by evaluation design (n=17)



4.6.4. Summary

Summarising the results of the meta-analysis, we find that overall the decentralisation interventions included in the study show somewhat negative effects on drop-out and repetition. Effects on test-scores are more robust overall, being positive and significant on aggregate in all cases, particularly in middle income countries. While pooled effects on teacher attendance are not significant overall, there is some evidence that these effects are stronger in contexts of high decentralisation and low income. There are examples of statistically significant findings for RCTs – in particular the study in Kenya by Duflo et al. (2012). However, pooled effects for RCTs are often weaker. It is important to note that these studies frequently, but not always, are assessed as being of low risk of bias. The next section further examines the robustness of the findings to bias.

4.7. ANALYSIS OF BIAS IN THE INCLUDED STUDIES

In this section, we examine whether the results differ depending on our rating of each study as being either 'low' or 'medium' risk of bias and conduct an analysis of publication bias.

4.7.1. Risk of bias sensitivity analysis

For the most part, we do not find notable differences in effect size point estimates between studies classified as medium and low risk of bias, although it is worth noting that the sample size for low risk of bias studies is relatively small. Hence we find a difference in statistical significance (medium risk of bias studies tending to show statistically significant findings, low risk of bias studies tending not to). We do find that the pooled effect for low risk of bias studies on drop-out is negative and significant when this group is treated separately (-0.05 SMD; 95% CI = -0.08, -0.01) (Figure 21). This is not the case for the other outcomes – maths (Figure 22), language (Figure 23) and teacher attendance (Figure 24) – where findings from low risk of bias studies are generally marginally insignificant, likely owing to small sample size in the cases of mathematics and language.

Figure 21: Effects on student drop-out by risk of bias assessment (n=10)

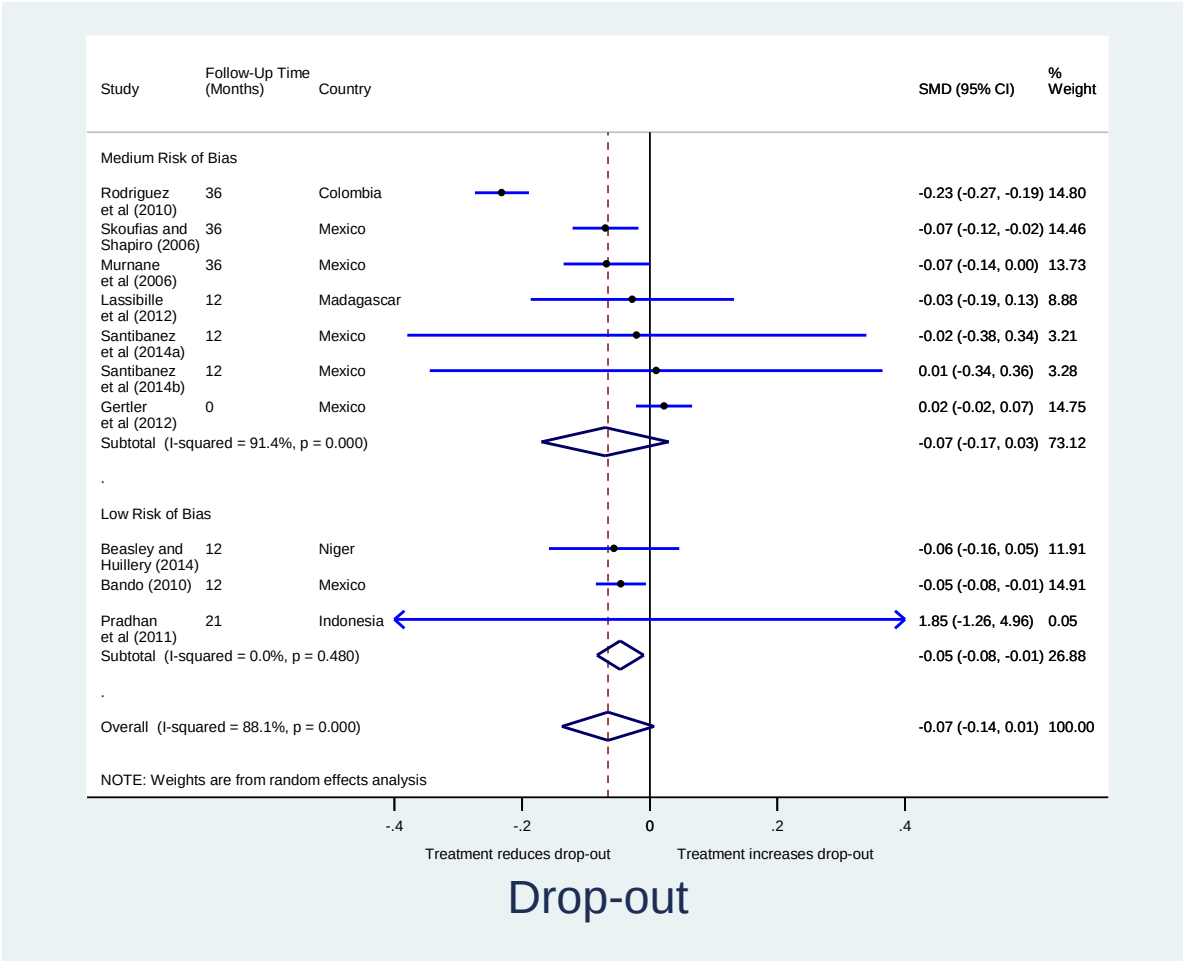


Figure 22: Effects on maths test score by risk of bias assessment (n=19)

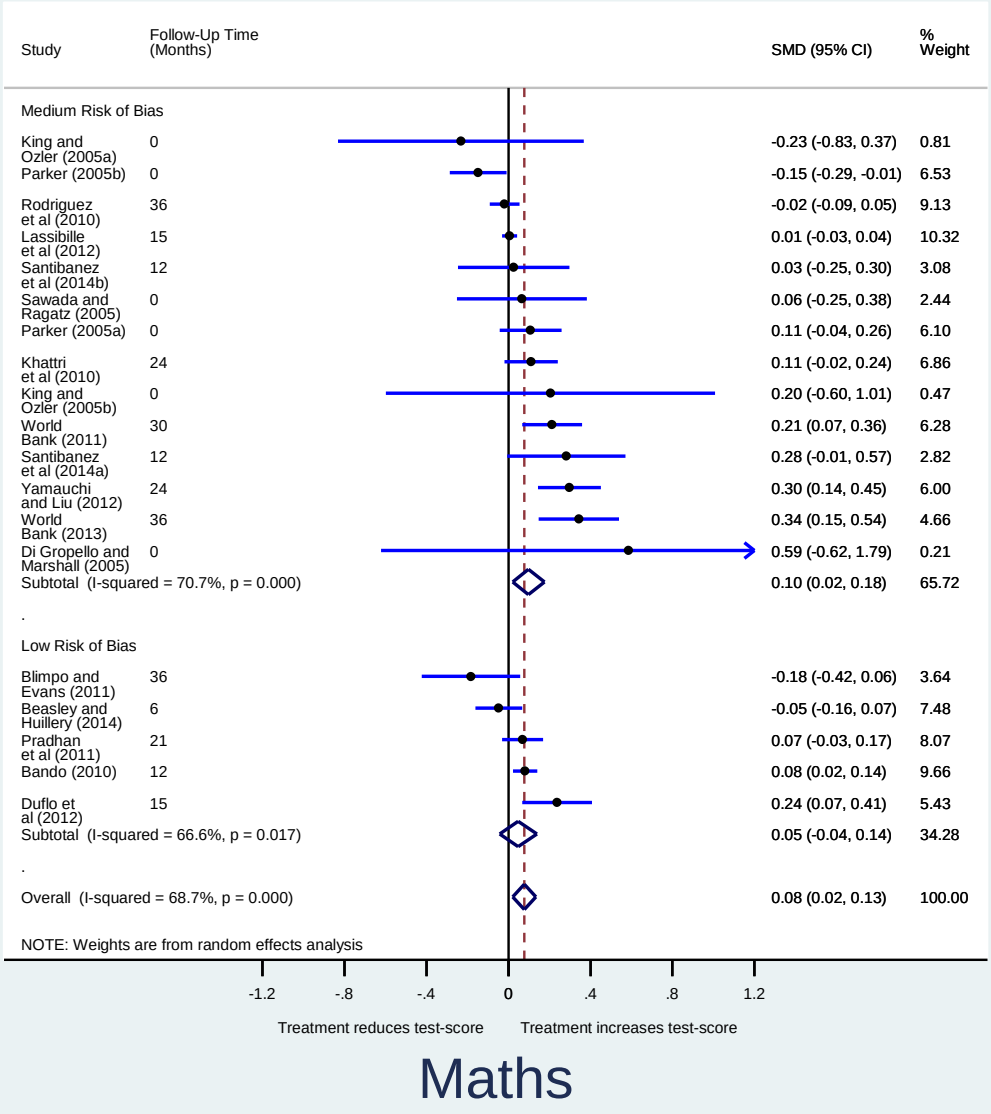


Figure 23: Effects on language test score by risk of bias assessment (n=17)

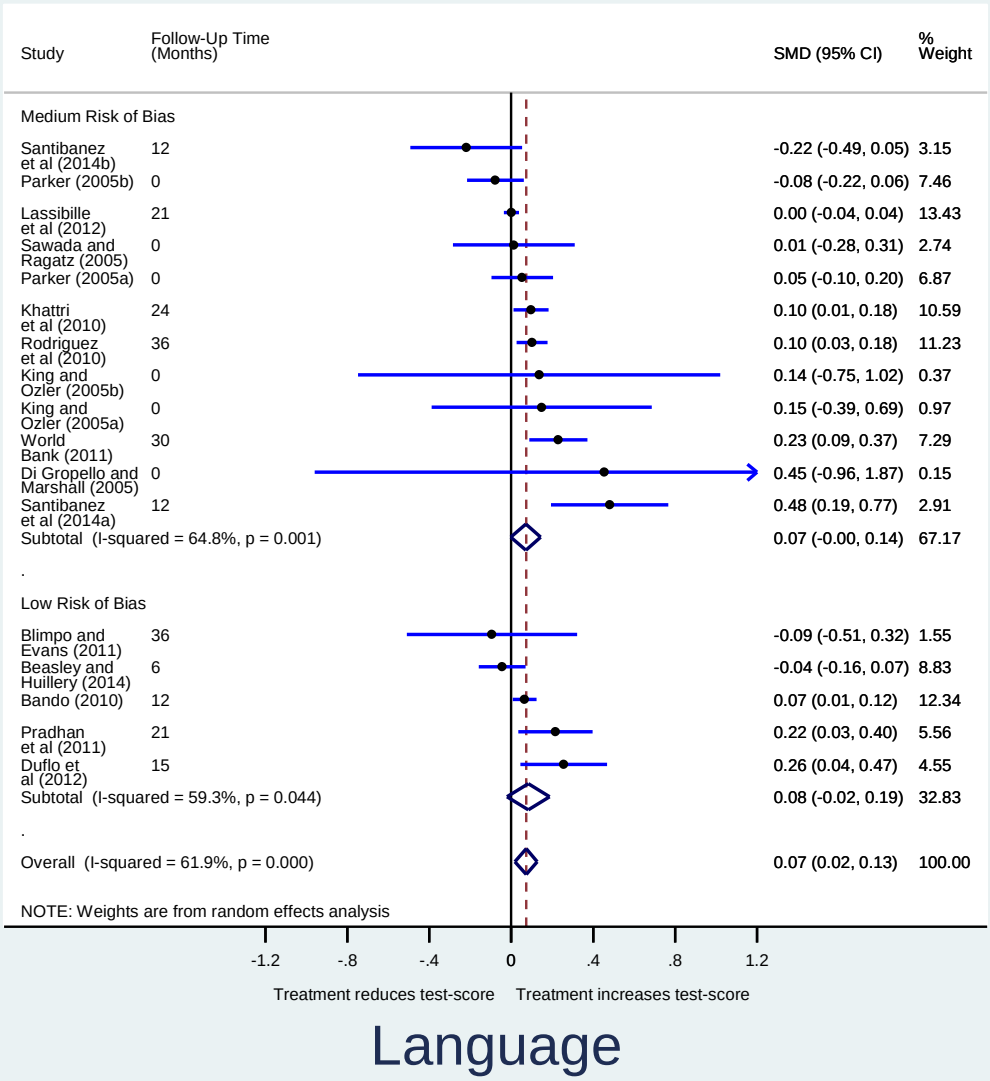
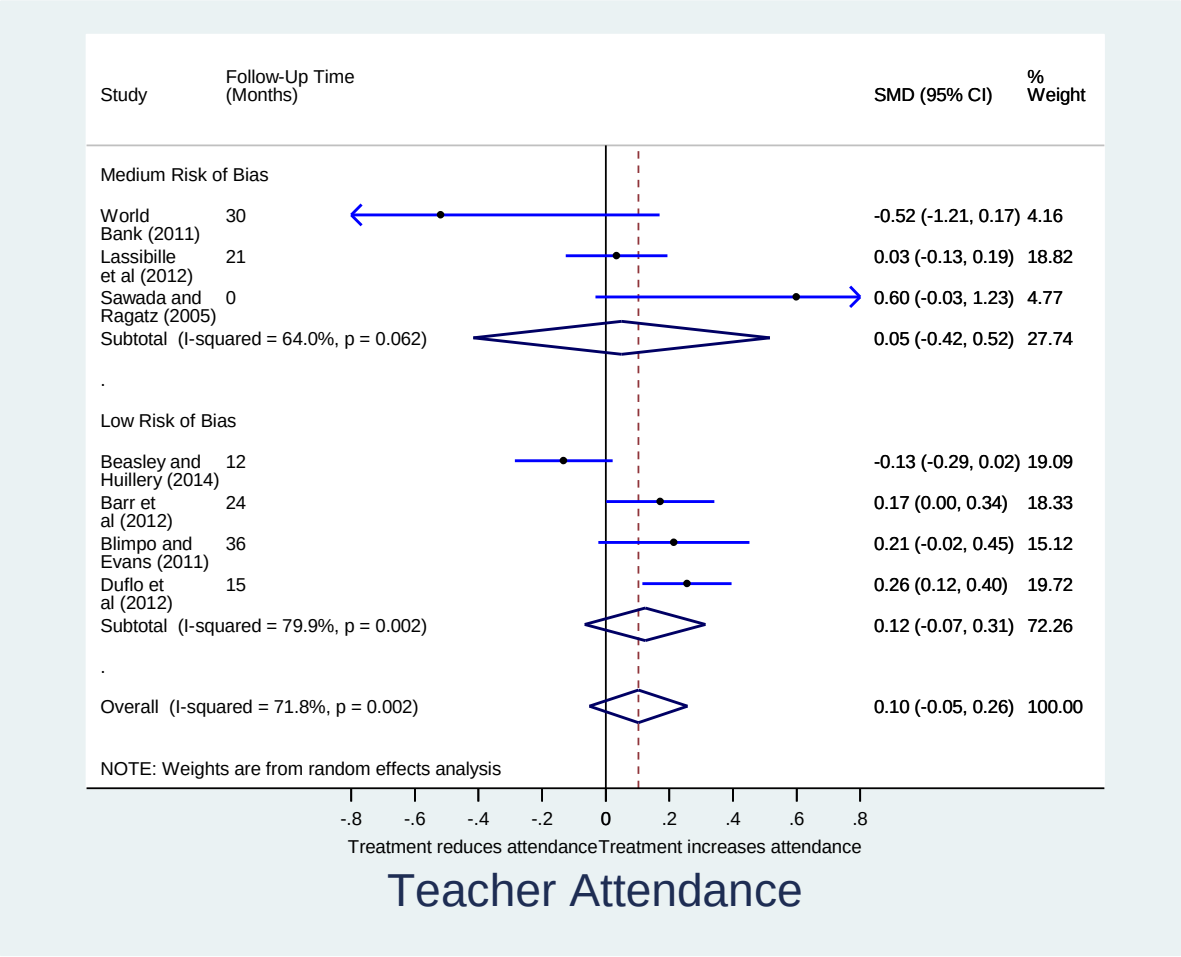


Figure 24: Effects on teacher attendance by risk of bias assessment (n=7)



4.7.2. Publication bias

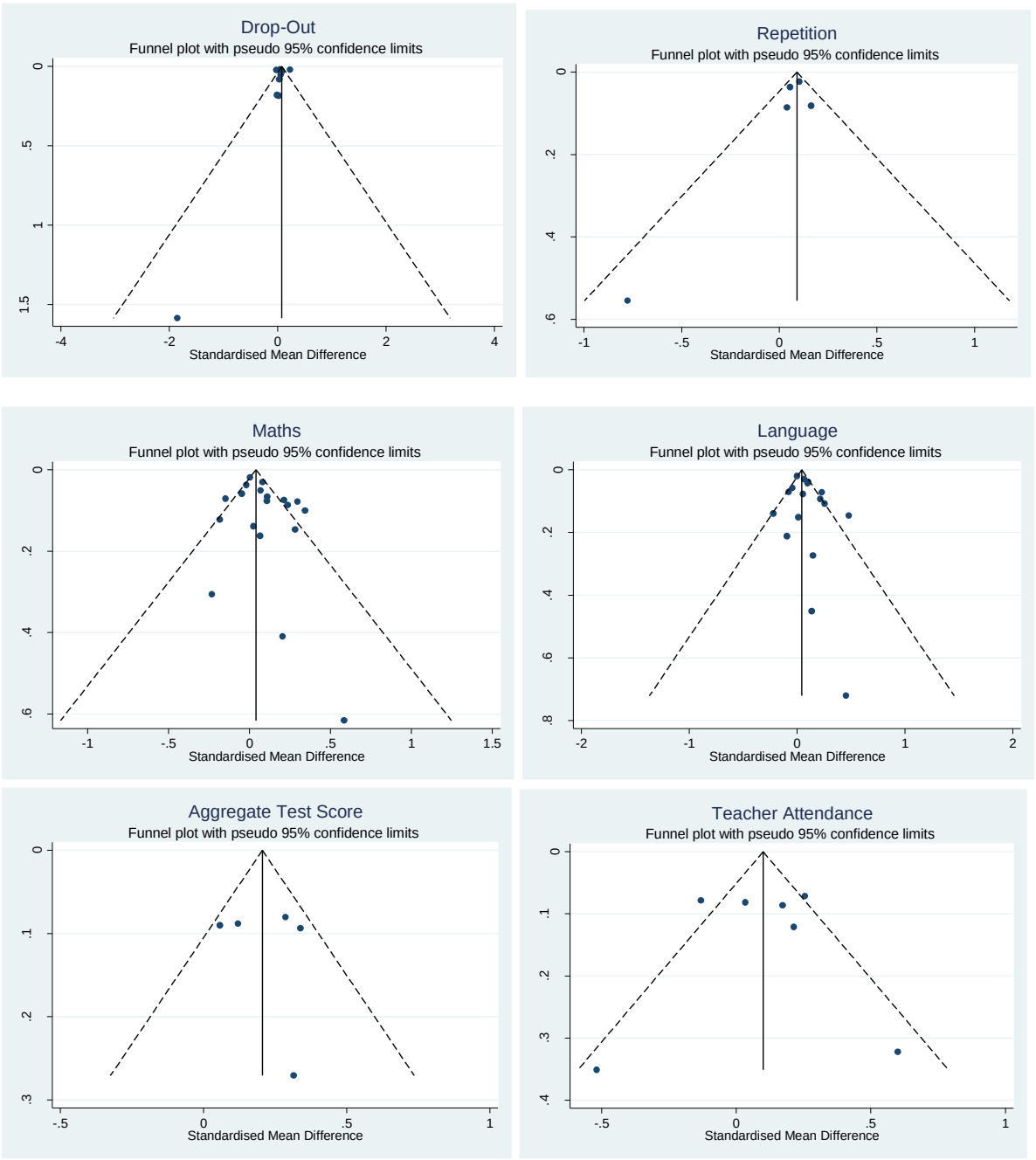
This review includes a range of impact studies in terms of publication type, including a large number of unpublished studies. 'Publication bias' denotes bias which is due to systematic differences in terms of results between studies with different kinds of publication status, particularly between published academic journal articles and unpublished reports, for example, which may arise because studies with smaller samples or with non-significant or negative findings may be less likely to be published in journals or be less likely to be located.

Sixteen – or approximately 62 percent – of the impact studies included in this review are working papers, while only 6 (23%) are peer-reviewed academic journal articles, the remainder comprising one unpublished thesis and three published book chapters. Hence we do not expect *a priori* that publication bias should be a major issue. However, we follow established procedures to test for the presence of publication bias.

First, we produce a set of funnel plots (Figure 25 below) for each of the study outcomes to examine symmetry visually. We use the absolute values of the standardised mean difference where a negative outcome is considered desirable, i.e. in the cases of drop-out and repetition, so that these appear as positive estimates on the funnel plots for ease of interpretation. Few

of the estimates included in the review had large standard errors and the plot results are relatively symmetric overall, suggesting limited evidence for publication bias, while some outcomes have too small a number of estimates to assess symmetry effectively.

Figure 25 : Publication bias funnel plots



Second, we conducted the Egger et al. (1997) test for asymmetry in the case of each outcome. The bias co-efficient estimates, their standard errors, t-statistics, p-values and confidence intervals are reported in Table 2. None of the tests finds a significant *p*-value, indicating no statistical evidence for publication bias.

Table 2: Results of Egger-tests for small-study effects (publication bias)

	Bias Co-efficient	*Std. Error*	*t*	*P>|t|*	*95% Confidence Interval*		*N Studies*	*Evidence of Bias?*
Drop-out	-0.837	1.648	-0.51	0.625	-4.637	2.963	10	No
Repetition	-0.916	0781	-1.17	0.326	-3.402	1.570	5	No
Language	0.771	0.579	1.33	0.203	-0.464	2.001	17	No
Maths	0.938	0.626	1.50	0.152	-0.383	2.258	19	No
Aggregate Test Score	0.476	2.215	0.21	0.844	-6.572	7.524	5	No
Teacher Attendance	-0.018	1.851	-0.01	0.992	-4.777	4.741	7	No

Following Duval and Tweedie (2000), we conducted a trim and fill analysis for each set of estimates by outcome. Following this routine, no trimming is performed in relation to the outcomes drop-out and repetition, so that their pooled effect sizes remain unchanged. With regard to language and mathematics, two and one estimates (for small sample studies) respectively are trimmed and filled, while the pooled effect sizes retain their original signs and significances and change very little in magnitude. For aggregate test-score and teacher attendance, no estimates are trimmed and for science the sample of estimates is too small to undertake trim and fill analysis meaningfully, while the pooled effect size is in any case not significantly different from zero. These results are consistent with the finding of a lack of evidence for publication bias, and we conclude that the substantive conclusions of the meta-analysis are not significantly affected by publication bias.

4.8. EXAMINATION OF HETEROGENEITY: STUDY SUB-GROUPS

Although some relatively weak conclusions can be drawn from the meta-analysis conducted in this review, the results are not sufficiently robust to support the conclusion that locating decision-making within schools and communities has a universally positive impact on a broad range of educational outcomes. It is perhaps not surprising that the aggregate analysis is somewhat inconclusive in this regard, given that many of the included studies report extensive heterogeneity within their individual samples. In this section, we discuss the heterogeneity factors considered within the studies themselves. As there is almost no overlap between the studies, there is little value in comparing the effects across studies, so, instead, our discussion of heterogeneity is presented in narrative format. We include the results of the studies, so that differential impacts *within* studies can be compared, but we do not standardise the results on a common scale.[13] We note here that individual studies may not be sufficiently statistically powered to assess effects on sub-groups, a problem which is compounded the smaller the sub-group sample size. Hence the findings of this analysis are interpreted cautiously: we do not discuss statistically insignificant findings.

[13] Throughout this section, we concentrate on the six outcomes included in the meta-analysis, as we do not have sufficiently robust evidence across studies regarding any additional outcomes.

4.8.1. Student-level factors

Although most included studies do not disaggregate results by student-level factors, a few do, and we report on those results in this subsection. The student-level factors investigated in at least one of the impact studies include: baseline academic ability, gender, socio-economic status, and grade level. The results are outlined in Table 3 below.

Table 3: Summary of student-level heterogeneity effects

Factor	Evidence of Differential Impact	Name of Intervention	Country	Citation	Relevant Outcomes	Results	Data source and interpretation of results
Baseline ability	Higher ability => pos impact	Unnamed RCT (SBM with various additional features)	Indonesia	Pradhan et al. (2011)	Test scores	Overall effect of linkage/election on **language scores** = 0.216** (0.093) Effect on language scores for those with lowest base scores = 0.208 (0.093); for those with highest base scores = 0.372** (0.150); Overall effect of linkage/election on **math scores** = 0.061 (0.077) Effect on math scores for those with lowest base scores = -0.067 (0.154); for those with higher (but not highest) scores = 0.184** (0.091)	Results found on page 37; method = intent-to-treat Effect of SBM with linkage/election stronger for students with higher baseline ability
Gender	Females => pos impact	Unnamed RCT (SBM with various additional features)	Indonesia	Pradhan et al. (2011)	Test scores	Overall effect of linkage/election on **language scores** = 0.216** (0.093) Effect on language scores for boys = 0.170* (0.100); for girls = 0.251** (0.098) Overall effect of linkage/election on **math scores** = 0.061 (0.077) Effect on math scores for boys = -0.003 (0.092); for girls = 0.120 (0.076)	Results found on page 37; method = intent-to-treat Effect of SBM with linkage/election on language stronger for female students but effects for boys also significantly positive (also likely to be mediated by baseline ability as girls likely to do better on baseline tests than boys)

Factor	Evidence of Differential Impact	Name of Intervention	Country	Citation	Relevant Outcomes	Results	Data source and interpretation of results
Socio-economic status (SES)	Higher SES => pos impact	PER	Colombia	Rodriguez et al. (2010)	Drop-out; Test scores	Coefficient on '*per capita* household income' = 1.019** (0.396) Coefficient on 'educational attainment (avg. parents)' = 0.490*** (0.153)	Results found on page 424; method = probit model Schools enrolling students from higher income homes and better educated families on average more likely to be successful
Grade level	Higher grades => pos impact	Autonomous Schools	Nicaragua	King & Ozler (2005)	Test scores	Effect of *de facto* autonomy on primary school **math scores** = 1.642* (0.891); on secondary math scores = -0.043 (1.525) Effect of *de facto* autonomy on primary school **language scores** = 0.822 (0.774); on secondary math scores = -0.584 (1.152)	Results found on page 37; method = fixed effects regression Impact on math scores identified at the primary level (no difference in terms of language)
	Lower grades => pos impact	SBM reform	Niger	Beasley & Huillery (2014)	Drop-out	Overall effect of intervention on drop-out = -0.00559 (0.00520) Effect on drop-out for students in Grade 1 = -0.0136* (0.00758) Effect on drop-out for students in Grade 2 = -0.00646 (0.0107) Effect on drop-out for students in Grade 6 = 0.00139 (0.00987)[14]	Results found on pages 56 and 57; method = intent-to-treat effects with interaction terms Impact on drop-out stronger for children in lower grades (although no difference in terms of other outcomes, e.g., test scores)

[14] We have not included all six grade-specific estimates here for space reasons, but the pattern is consistent, with subsequent years showing a progressively diminished effect. Full results are available in the original paper.

Factor	Evidence of Differential Impact	Name of Intervention	Country	Citation	Relevant Outcomes	Results	Data source and interpretation of results
		AGE	Mexico	Gertler et al. (2012)	Drop-out; Repetition	Overall effect on **repetition** = -0.004* (0.002) Effect on students in Grades 1, 2 or 3 = -0.007** (0.002); on students in Grades 4 or 5 = 0.002 (0.002) Overall effect on **drop-out** = 0.001 (0.002) Effect on students in Grades 1, 2 or 3 = 0.000 (0.002); on students in Grades 4 or 5 = 0.003 (0.002)	Results found on page 74; method = fixed-effects regression Significant impact on repetition for lower grades
		Autonomous Schools	Nicaragua	Parker (2005)	Test scores	Impact on **math scores** in Grade 3 sample = 3.8 (1.4)*; in Grade 6 sample = -3.7 (-2.1)** Impact on **language scores** in Grade 3 sample = 1.8 (0.7)**; in Grade 6 sample = -1.9 (-1.1)	Results found on pages 380 and 382; method = propensity score matching (nearest neighbour) Impact on test scores negative for Grade 6 sample (math) and positive for Grade 3 sample (math and language)

Factor	Evidence of Differential Impact	Name of Intervention	Country	Citation	Relevant Outcomes	Results	Data source and interpretation of results
		PER	Colombia	Rodriguez et al. (2010)	Drop-out; Test scores	Overall effect on **language scores** = 0.016*** (0.006) Effect in primary schools = 0.016*** (0.005); in secondary schools = 0.004 (0.016) Overall effect on **math scores** = -0.004 (0.008) Effect in primary schools = -0.004 (0.009); in secondary schools = -0.043 (0.025) Overall effect on **drop-out** = -0.032*** (0.003) Effect in primary schools = -0.057*** (0.007); in secondary schools = -0.044*** (0.017)	Results found on pages 420 and 421; method = DiD Impact on language test scores significantly positive for primary level; no differential impacts between primary and secondary for math scores or drop-out
		PEC-FIDE	Mexico	Santibanez et al. (2014)	Drop-out; Test scores	Impact on **math scores** in Grade 3 sample = 17.92 (9.329); in Grade 6 sample = 1.641 (8.991) Impact on **language scores** in Grade 3 sample = 28.40 (8.618); in Grade 6 sample = -12.08 (7.641) Impact on **drop-out** in Grade 3 sample = -0.0763 (0.691); in Grade 6 sample = 0.0387 (0.697)	Results found on page 105; method = PSM using DiD Finds apparently stronger effects within the 3rd grade sample for all three outcomes, although the effects are not statistically significant

Factor	Evidence of Differential Impact	Name of Intervention	Country	Citation	Relevant Outcomes	Results	Data source and interpretation of results
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.							

Only one study considers the differential impact of baseline ability (Pradhan et al., 2011), suggesting a stronger effect for students scoring higher at baseline.[15]

Gender effects are also robustly explored by only one study (Pradhan et al., 2011). They identify a positive effect for female students, but the authors acknowledge that this is result is likely to be confounded by baseline ability, as girls performed better than boys on the baseline test.

Similarly, the impact of socio-economic status is investigated by one study (Rodriguez et al., 2010); they find evidence of stronger impact on students from better-educated, wealthier families.

Six studies consider the differential impact of grade level (Beasley & Huillery, 2014; Gertler et al., 2012; King & Ozler, 2005; Parker, 2005; Rodriguez et al., 2010; Santibanez et al., 2014). Overall, the results suggest a stronger impact on students in lower grades for a range of outcomes – drop-out (Beasley & Huillery, 2014), repetition (Gertler et al., 2012), and test scores (Parker, 2005; Rodriguez et al., 2010; Santibanez et al., 2014) - but the results are not entirely consistent. King & Ozler (2005) identify a stronger effect for math in their secondary school sample, Gertler et al. (2012) do not identify a stronger effect on drop-out for lower grades, and Rodriguez et al. (2010) only identify a stronger effect on language, not on other tests. Rodriguez et al. (ibid.) also identify no difference in drop-out rates between primary and secondary students.

4.8.2. School-level factors

We next report on a number of school-level factors considered in the various studies, specifically the size of the school and the characteristics of teachers and head teachers. The results are outlined in Table 4.

[15] Bold et al. (2013) also consider baseline performance and find limited evidence that the intervention is progressive in the government treatment arm, with a larger effect identified for schools with lower baseline performance. However, as these results relate to analysis of the effect of the overall contract teacher programme, not the specific element of the programme that sought to increase autonomy at the school level, the study has not been included in the summary table.

Table 4: Summary of school-level heterogeneity effects

Factor	Differential impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
Size of school	Smaller schools => pos impact	Autonomous Schools	Nicaragua	King & Ozler (2005)	Test scores	Results of 1st stage OLS regressions for *de facto* autonomy: Large school dummy (enrolment > 4,000) = -0.189** (0.086)	Results found in Appendix (Table E7); method = OLS regression Authors identify a significantly stronger effect in small schools
		SBM reform	Niger	Beasley & Huillery (2014)	Teacher Attendance	Evidence of *positive* impact of grants on teacher attendance (coefficient on interaction term = 0.17** significant at 5% level) One-teacher schools budgeted more money for expenses related to teacher support (coefficient = 8993 FCFA**, significant at 5% level) and functioning of school committee (2100 FCFA, significant at 5% level)	Full results not available in paper, but results discussed in detail on pages 28 and 29; method = intent-to-treat effects with interaction terms Better teacher attendance in one-teacher schools. Argument that this may be because the SMC is more likely to choose to spend the grant on something of benefit to the teacher (e.g. housing), given threat of losing the teacher (i.e. 'alliance' between SMC and teacher)
	Note: Some studies finding positive impact of SBM initiatives - e.g. Sawada & Ragatz (2005) and Jimenez & Sawada (1999; 2003) re EDUCO; Di Gropello & Marshall (2005) re PROHECO - mention that the initiative tended to be implemented in smaller communities, but the sample did not allow for an explicit examination of the influence of this factor						
Teacher characteristics	No differential	AGEMAD	Madagascar	Glewwe & Maïga	Test scores	Overall effect of school-level intervention on test scores =	Results found on page 7; method = fixed effects regression

Factor	Differential impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
	impact depending on type of teacher			(2011)		0.071 (0.105) Effect on students with contract teachers = 0.089 (0.189) Effect on students with civil service teachers = -0.108 (0.095) Effect on students with student teachers = 0.317 (0.458)	Considers possibility of differential impact on kind of teacher (e.g. civil service teacher, contract teacher, student teacher) and finds no significant effects
	Contract teachers, who are less experienced => pos impact	ETP	Kenya	Duflo et al. (2012)	Teacher Attendance; Test scores	Effect of ETP (contract teacher programme) on **math scores** = 0.135* (0.075) Effect of ETP in schools with school-based management committees = 0.207*** (0.076) Effect on students of contract teachers of ETP in schools with school-based management committees = 0.237*** (0.087) Effect on students of civil service teachers of ETP in schools with school-based management committees = 0.201** (0.082) Effect of ETP (contract teacher programme) on **language**	Results found on pages 36 and 41; method = average treatment effect, with interaction terms, for test score data; linear probability model for teacher attendance, based on data from unannounced visits Main source of impact comes from contract teacher programme (ETP); SBM training strengthens the effect; effect strongest on contract teachers and students of contract teachers

Factor	Differential impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
						scores = 0.191** (0.095) Effect of ETP in schools with school-based management committees = 0.198** (0.100) Effect on students of contract teachers of ETP in schools with school-based management committees = 0.256** (0.108) Effect on students of civil service teachers of ETP in schools with school-based management committees = 0.166 (0.103) Effect of ETP on **attendance** of contract teachers = 0.011 (0.037) Effect of ETP on attendance of civil service teachers = -0.017 (0.024) Effect of ETP in schools with school-based management committees on attendance of contract teachers = 0.093*** (0.026) Effect of ETP in schools with school-based management	

Factor	Differential impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
						committees on attendance of civil service teachers = -0.024 (0.026)	
	More experienced => pos impact	RCT (two kinds of scorecard)	Uganda	Barr et al. (2012)	Teacher Attendance	Overall effect of participatory scorecard on teacher retention = 0.119** (0.06) Effect of participatory scorecard on teacher retention, when interacted with years worked at the school = 0.0334** (0.01) Effect of participatory scorecard on teacher retention, when interacted with log baseline salary = -0.0417 (0.04)	Results found on page 23; method = linear probability model (dependent variable = teacher is present during unannounced visit) Participatory version seems to work better with more experienced teaching staff. Standard treatment relatively ineffective among teachers with high salaries.
		EDUCO	El Salvador	Jimenez & Sawada (2003)	Repetition	Overall effect on repetition = -0.08 (0.45) Effect of years of teacher experience = 0.13** (1.97)	Results found on page 43; method = probit model with fixed effects Impact on repetition more pronounced in classrooms with more experienced teaching staff.
Head teacher characteristics	Strong leadership = condition	PER	Colombia	Rodriguez et al. (2010)	Drop-out; Test scores	Coefficient on 'rating school management and administration' = 1.482*** (0.462)	Results found on page 424; method = probit model Estimate a probit model of success, weighted by total number of students in a school, and find that PER's success

Factor	Differential impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
							depends on a combination of three factors: good training, high quality of educational material, and ‘first rate’ school management.
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.							

Although only two studies consider the size of school explicitly (Beasley & Huillery, 2014; King & Ozler, 2005), both find clear evidence of stronger impact on smaller schools. This may be because it is easier for school management committee members to monitor teachers when students spend the whole day with the same teacher (as is typically the case in smaller schools), or because reforms can be more directly experienced in smaller schools, given the relative simplicity of the relations between actors in comparison to larger schools with more administrative infrastructure. It is possible that this factor also helps to explain some of the positive results found in other studies (e.g. Di Gropello & Marshall, 2005; Sawada & Ragatz, 2005), as a number of the specific interventions (e.g. PROHECO, EDUCO) target communities which, by definition, are likely to have small schools.

Four studies consider the possibility of differential impact on different kinds of teachers. These results are inconclusive in the aggregate. One study (Glewwe & Maïga, 2011) finds no differential impact between different kinds of teacher.[16] The other studies do find evidence of differential impact, but the differences they identify are not consistent. Barr et al. (2012) and Jimenez & Sawada (2003) both identify stronger effects in schools with more experienced (and, in the case of Barr et al., better paid) teachers, while Duflo et al. (2012) identify stronger effects on contract teachers, who are typically *less* experienced than their civil service counterparts.

Although no studies explicitly compare schools with different head teacher characteristics, one (Rodriguez et al., 2010) identifies management and/or principal leadership as important mitigating factors, with stronger leadership being correlated with greater success of SBM initiative.

4.8.3. Community-level factors

We next report on community-level factors explored in the various studies. The results are outlined in Table 5.

[16] As with the Jimenez & Sawada (1999) study, discussed in the previous sub-section, Glewwe & Maïga (2011) has been included in the heterogeneity analysis, despite their removal from the meta-analysis for possible dependence of results, because they report on different heterogeneity effects than do Lassibille et al. (2010).

Table 5: Summary of community-level heterogeneity effects

Factor	Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
Level of development	Lower level => neg impact	AGE	Mexico	Gertler et al. (2012)	Drop-out; Repetition	Overall effect on repetition = -0.004* (0.002) Effect on students in Grades 1, 2 or 3 in low marginality communities = -0.009** (0.002) Effect on students in Grades 1, 2 or 3 in high marginality communities = -0.004 (0.003)	Results found on page 75; method = fixed-effects regression An overall impact was found for drop-out/repetition and for less marginalised communities
		PEC	Mexico	Murnane et al. (2006)	Drop-out	Overall effect = -0.274** Effect on communities at high level of development=-0.247** Effect on communities at medium level of development = -0.331** Effect on communities at low level of development = -0.15	Results found on pages 42 and 44; method = DiD estimates, obtained from fitted regression models (fixed effects) Impacts found in those communities classified as "middle" and "high" levels of development, according to Human Development Index
				Skoufias & Shapiro (2006)	Drop-out; Repetition	Overall effect on drop-out = -0.239 (0.091)** Effect on drop-out in high marginality areas = 0.428 (0.263); in low marginality areas = -0.057 (0.088)	Results on page 39; method = average effect of treatment on the treated, based on local linear regression matching estimates Statistically significant reduction in repetition in low marginality (more advantaged) communities. No

Factor	Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
						Overall effect on repetition = 0.313 (0.068) Effect on repetition in high marginality areas = 0.025 (0.396); in low marginality areas = -0.219 (0.068)***	difference between high and low marginality areas for drop-outs
Urbanicity	Urban areas => pos impact for drop-outs	PEC	Mexico	Skoufias & Shapiro (2006)	Drop-out; Repetition	Overall effect on drop-out = -0.239 (0.091) Effect on drop-out in urban areas = -0.134 (0.070)*; in rural areas = -0.038 (0.075) Overall effect on repetition = 0.313 (0.068) Effect on repetition in urban areas = -0.213 (0.045)***; in rural areas = -0.241 (0.066)**	Results on page 39; method = average effect of treatment on the treated, based on local linear regression matching estimates Significant impacts on reducing drop-outs in urban areas Significant impacts on reducing repetition in urban and rural separately
Parents' level of education	Uneducated community members on SMC => neg impact	SBM reform	Niger	Beasley & Huillery (2014)	Drop-out; Teacher Attendance; Test scores	*Negative* impact of grant on math* and French** test scores in schools with educated school committees (about one-third of standard deviation, significant at 5% level for French and 10% for math)	Full results not available in paper, but results discussed on page 28; method = intent-to-treat effects with interaction terms Conclude that limited impact on outcomes due to low levels of 'real authority'; also note that school committees with higher proportion of educated community members (defined as more than one SMC

Factor	Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
							member having completed primary education) more likely to monitor teacher attendance, although no impact on teacher attendance figures
		WSD	Gambia	Blimpo & Evans (2011)	Teacher Attendance[17]; Test scores	Overall effect of full treatment on math scores = -0.12 (0.08) Effect of full treatment on math scores in communities with higher percentage of literate adults = 1.12** (0.46) Effect of full treatment on math scores in communities in which there are *no* members of the school management committee with formal education = -0.65** (0.29) Overall effect of full treatment on language scores = -0.04 (0.09) Effect of full treatment on language scores in communities with higher percentage of literate adults =	Results found on pages 42, 44 and 45; method = average treatment effect, with interaction terms added for heterogeneity analysis Looked at 'baseline capacity' of community (i.e. literacy rate and percentage of SMC with basic education) and found that communities with higher capacity more likely to see gains as a result of the WSD reform. Argue that WSD could be counter-productive in areas where capacity is very low, although caution is needed owing to the small sample size.

[17] Teacher *absenteeism* captured in the original study, so signs were reversed prior to standardisation of effects for forest plots

Factor	Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
						0.78* (0.51) Effect of full treatment on language scores in communities in which there are *no* members of the school management committee with formal education = -0.57* (0.34)	
Level of community participation	More => pos impact	EDUCO	El Salvador	Jimenez & Sawada (1999)	Test scores	Overall effect of EDUCO on math scores, controlling for school inputs = 0.40 (0.27) Effect of EDUCO on math scores, controlling for school inputs *and* community participation = -0.77 (0.47) Effect on math scores of the number of parent association visits to classrooms in the past month = 0.14 (1.72)* Overall effect of EDUCO on language scores, controlling for school inputs = 1.57 (1.51) Effect of EDUCO on language scores, controlling for school inputs *and* community participation = 0.74 (0.65)	Results found on pages 431 and 435; method = fixed-effects regression Find that a significant proportion of effect can be explained by the level of community participation (as well as school-level inputs)

Factor	Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
						Effect on language scores of the number of parent association visits to classrooms in the past month = 0.10 (1.77)*	
		Autonomous Schools	Nicaragua	King & Ozler (2005)	Test scores	Effect of *de jure* autonomy on primary math scores = -0.232 (0.306); effect of *de facto* autonomy = 1.642* (0.891) Effect of *de jure* autonomy on primary language scores = 0.148 (0.274); effect of *de facto* autonomy = 0.822 (0.774) Effect of *de facto* administrative autonomy on primary math scores = 1.355** (0.526); of *de facto* pedagogical autonomy on primary math scores = 0.356 (0.848)	Results found on pages 37 and 38; method = fixed effects regression *De jure* autonomy not significant, but percentage of decisions taken by the community (*de facto* autonomy) is positively correlated with achievement in primary school When disaggregated, find that *de facto* administrative autonomy is more impactful than *de facto* pedagogical autonomy
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.							

Although only seven of the 26 impact studies explicitly consider community-level factors in their heterogeneity analysis, the findings in this sub-section are the most consistent in terms of contextual factors that are likely to affect the impact of school-based decision-making reforms. The community-level analysis considers three factors: the level of development of particular communities, the level of parental education within individual communities, and the level of community participation.

There is little discussion of the relative impact of school-based decision-making reforms on rural and urban areas, largely because most individual interventions are explicitly targeted at one or the other (and, therefore, individual studies do not consider differential impact in terms of urbanicity). However, one study does compare urban and rural areas (Skoufias & Shapiro, 2006), finding greater impact in urban areas. These results may be linked to the findings of four studies which investigate differential impact in terms of community disadvantage (Gertler et al., 2012; Murnane et al., 2006; Rodriguez et al., 2010; Skoufias & Shapiro, 2006). Although the four studies frame their analysis in slightly different ways, they all come to a similar conclusion: that school-based decision-making reforms are likely to have a stronger impact on more advantaged (i.e. wealthier) communities. This is a particularly important result, given that some studies showing positive impact explicitly acknowledge having avoided including more remote areas in their analysis (e.g. Glewwe & Maïga, 2011, and Lassibille et al., 2010).

These results are likely to be related to the results concerning the characteristics of community members. Given that school-based decision-making reforms often involve at least some community participation, it is just as important to investigate community member characteristics as it is to consider the characteristics of school personnel, such as teachers (as discussed in the previous sub-section). However, this factor is only investigated in two of the studies (Beasley & Huillery, 2014; Blimpo & Evans, 2011). Both studies suggest that parental education levels are an important factor, as they find that communities with a higher proportion of educated school management committee members are more likely to see positive results of school-based decision-making reforms. Beasley & Huillery (2014) argue that this is at least partially related to the level of parents' social capital, defined in terms of their relative authority within communities, suggesting that outcomes are likely to be limited in communities where parents have limited authority vis-à-vis school personnel. One would expect that these characteristics would affect the impact of school-based decision-making reforms, as both factors are likely to limit the impact of community participation in decision-making and the effect of community monitoring of school behaviour. They are also likely to be correlated with a community's overall level of development. It is therefore possible that a similar effect may be driving the results identified in the previous paragraph. Although all four studies investigating the differential impact of community disadvantage consider Latin American contexts, and the two studies considering community characteristics both focus on sub-Saharan Africa, it is reasonable to assume that areas of high disadvantage in Latin America are also characterised by similarly low levels of community human capital.

Finally, two studies investigate the possibility that some communities will opt to participate more actively in school decisions, as a result of school-based decision-making reforms, than others. The studies (Jimenez & Sawada, 1999; King & Ozler, 2005), both investigating Latin American contexts, find strong evidence that community participation levels are a critical factor. King & Ozler (2005) differentiate between communities with *de jure* autonomy (communities with a legal right to autonomy, provided by a particular reform) and those with *de facto* autonomy (communities in which participation in school decisions actually increases significantly as a result of the reform). They find positive effects only in communities with *de facto* autonomy, suggesting that giving communities authority to make decisions is only impactful if communities then elect to capitalise on their new autonomy. King & Ozler also disaggregate this effect and find that it is in the domain of administrative decisions that impact can really be identified; communities electing to engage with pedagogical decisions see less impact than those engaging with administrative decisions, such as raising additional funds and providing incentives to teachers.

4.8.4. National-level factors

As we explicitly excluded studies based on country-level comparisons, we found very little robust analysis of national-level factors. However, one such factor – the possibility of interaction effects between school-based decision-making reforms and other reforms in a given context – was considered by one included study, so the results are reported here.

School-based decision-making reforms are almost always implemented alongside other education reforms, many of which are led by central authorities. Although many studies acknowledge the possibility of interaction between reforms, most did not explicitly investigate the possibility that other reforms might affect the impact of the specific intervention in question. However, Gertler et al. (2012) did examine this question and found that the proportion of teachers under *Carrera Magisterial* (a centralised pay-for-performance scheme that rewards teachers for strong results on student assessments) significantly reduced repetition [-0.004* (0.002); significant at 90% level]. They also found that the proportion of students receiving *Oportunidades* vouchers in a school had a significant impact on drop-out [0.014** (0.002); significant at 95% level]. These reforms, therefore, are potential confounders affecting the overall results of the study. As no other study explicitly considers the potentially confounding effect of other reforms, some of the studies may have overestimated the impact of the school-based decision-making interventions under investigation.

4.8.5. Implementation factors

In addition to the student-level and contextual factors described in the previous sub-sections, the specific manner in which reforms are implemented might also be expected to differentially affect outcomes. For instance, one would expect to see different effects if devolution of decision-making is accompanied by additional financing for schools or if those assuming authority are offered training on their new responsibilities. Some school-based

management interventions, such as TEEP in the Philippines, have been implemented as part of a broader programme of education reform; schools participating in TEEP received money for infrastructure/materials and pedagogical training, in addition to support for increased school-community partnership. One would assume that multi-faceted reforms like TEEP might have a stronger impact than narrower reforms focused exclusively on changing the level of decision-making authority.

Despite the likelihood that such implementation decisions would impact results, most of the included studies do not explicitly investigate any implementation factors, as they focus instead on the overall impact of a particular intervention. However, a small number of included studies using experimental designs (Blimpo & Evans, 2011; Bold et al., 2013; Duflo et al., 2012; Pradhan et al., 2011; World Bank, 2011) do consider implementation factors by creating a number of discrete treatment arms, each constituting a different combination of elements. In this sub-section, we discuss six implementation factors considered by this small sample of experiments: the incorporation of a grant, the incorporation of training, the incorporation of a report card or other accountability mechanism, the mechanism by which school management committee members are selected, the relationship between schools and the surrounding community (outside of school management committees), and the implementing body. Where relevant and appropriate, we also reference supporting evidence from the other impact studies.

We start by highlighting the results of the experiment conducted by Pradhan et al. (2011) in Indonesia, as this study is the only one in the review to explicitly consider the differential impact of a range of implementation factors. The randomised control trial outlined in this study comprised a number of treatment arms, each of which included either training, elections, facilitation of collaboration between school management committees and village councils (a factor they call “linkage”), or some combination of the three. Overall, they find no effect within the control group (receiving only a grant), nor do they find any effect on schools receiving only the grant and training. However, they do find impact in schools where elections and/or linkage were facilitated. The full results are outlined in Table 6.

Table 6: Summary of comparative results from Pradhan et al. (2011)[18]

	Grant	*Training*	*Elections*	*Linkage*	*Linkage & Election*	*Linkage & Training*	*Training & Election*
Drop-out (n=517)	-0.005 (0.005)	0.007 (0.006)	-0.003 (0.006)	-0.002 (0.006)	-0.005 (0.011)	0.003 (0.006)	0.004 (0.006)
Repetition (n=517)	-0.004 (0.008)	-0.006 (0.005)	-0.001 (0.005)	0.007 (0.005)	0.007 (0.008)	0.001 (0.009)	-0.006 (0.008)
Average test score (n=11,463)	0.129 (0.094)	-0.049 (0.069)	0.049 (0.069)	0.165** (0.067)	0.216** (0.093)	0.116 (0.086)	0.002 (0.101)

The authors' conclusion from these results is that elements that support existing school management committees are unlikely to have an effect, whereas elements that introduce new participants (e.g. elections and linkage) are likely to substantially impact outcomes. Although these findings are the result of only one study, they raise interesting questions that would benefit from further attention in future studies.

Grants

We next consider the potential impact of providing **grants** to schools as part of a school-based decision-making intervention. Many school-based decision-making interventions follow a grant-giving model, whereby selected schools are given grants to fund school improvement plans developed by school management committees. In other models, schools are given grants for explicit purposes, e.g. the hiring of contract teachers (as discussed in Bold et al., 2013; and Duflo et al., 2012). Although these models differ, they all comprise increased decision-making at the level of the school and an increase in school funding through the provision of a grant.

In fact, no study in the sample offers insight into the marginal impact of allocating grants, because all of the experiments including a grant component allocate grants to all of the treatment arms. Receipt of the grant is typically the 'control' condition, which is then compared to other treatments in which the base grant is supplemented by an additional intervention, e.g. training of the school management committee (see, for example, Blimpo & Evans, 2011; Bold et al., 2013; Duflo et al., 2012). We therefore cannot draw any robust

[18] Results found on page 37; method = intent-to-treat; effect sizes not standardised, reproduced here on the original scale

conclusions around the differential impact of providing a grant. However, we can draw some tentative conclusions by comparing the overall results of studies in the sample which do and do not include a grant component. A summary of studies investigating interventions including a grant is presented in Table 7.

Table 7: Summary of evidence relating to grants

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference (*p*-value)[19]	Notes
Pos impact identified overall; reform includes grant	PEC	Mexico	Bando (2010)	Drop-out; Test scores	Overall effect on drop-out = -0.045 (0.025)** Overall effect on math scores = 0.081 (0.008)*** Overall effect on language scores = 0.065 (0.027)**	Grant provided to fund School Improvement Plan; includes matching funds for monies raised locally
			Murnane et al. (2006)	Drop-out	Overall effect on drop-out = -0.068 (0.050)**	
			Skoufias & Shapiro (2006)	Drop-out; Repetition	Overall effect on drop-out = -0.069 (0.009)*** Overall effect on repetition = -0.104 (>0.001)***	
	PEC-FIDE	Mexico	Santibanez et al. (2014)[20]	Drop-out; Test scores	Overall effect on drop-out (Grade 3) = -0.020 (0.920) Overall effect on math scores (Grade 3) = 0.282 (0.054)* Overall effect on language scores (Grade 3) = 0.481 (0.001)***	Grant amount depends on size of school; can be spent on training, interventions for children 'at risk', materials, equipment, or infrastructure

[19] As we are comparing across studies in these tables, we have elected to use the standardised effect sizes, rather than the data in their original form. However, caution is advised, as these figures show the overall effect of school-based decision-making (for interventions with and without grants). They do not show the effects of the grants *per se*.
[20] Positive results for Grade 3 sample only.

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference (*p*-value)[19]	Notes
	Autonomous Schools	Nicaragua	King & Ozler (2005)[21]	Test scores	Overall effect on math scores (secondary) = 0.205 (0.630) Overall effect on language scores (primary) = 0.148 (0.601) Overall effect on language scores (secondary) = 0.136 (0.770)	All communities participating in the programme receive a grant; the grant appears to be insufficient on its own, given the apparent low impact in communities with low *de facto* autonomy
	TEEP	Philippines	Khattri et al. (2010)	Test scores	Overall effect on math scores = 0.110 (0.097)* Overall effect on language scores = 0.097 (0.026)**	Grants linked to School Improvement Plans; many chose to use funds to support construction
			Yamauchi & Liu (2012)	Test scores	Overall effect on math scores = 0.297 (<0.001)*** Overall effect on aggregate test scores = 0.287 (<0.001)***	
	BESRA	Philippines	World Bank (2013)	Test scores	Overall effect on math scores = 0.343 (<0.001)*** Overall effect on aggregate test scores = 0.339 (<0.001)***	Grants linked to School Improvement Plans
			Yamauchi (2014)	Test scores	Overall effect on aggregate test scores = 0.315 (0.247)	
	PSI	Sri Lanka	World Bank	Teacher Attendance;	Overall effect on teacher attendance	Grants linked to School Improvement

[21] Positive results on math score for secondary sample only

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference (*p*-value)[19]	Notes
			(2011)	Test scores	= 0.519 (0.140) Overall effect on math scores = 0.213 (0.004)*** Overall effect on language scores = 0.230 (0.002)***	Plans
Mixed effect identified overall; intervention includes grant	AGE	Mexico	Gertler et al. (2012)	Drop-out; Repetition	Overall effect on drop-out = 0.022 (0.322) Overall effect on repetition = -0.055 (0.134)	Grants can only be used for infrastructure and materials (not wages); usually phased in at $500-700 a year; parents required to participate more in school in exchange
	RCT (two kinds of scorecard)	Uganda	Barr et al. (2012)	Teacher Attendance	Overall effect on teacher attendance = 0.172 (0.047)**	Both treatment arms received grants; effect only identified in arm using participatory scorecard
	ETP	Kenya	Bold et al. (2013)	Test scores	Overall effect on aggregate test scores = 0.057 (0.537)	All treatment arms received grant for hiring contract teachers; effect only identified when implemented by NGO; stronger effect also identified when combined with local hiring and training of SMC members
			Duflo et al. (2012)	Teacher Attendance; Test scores	Overall effect on teacher attendance = 0.256 (<0.001)*** Overall effect on math scores = 0.237 (0.006)*** Overall effect on language scores =	All treatment arms received grant for hiring contract teachers, but grant appears to have little impact on its own; effects differ depending on type of teacher (e.g. no effect identified in classrooms with civil service teachers, even within schools receiving the

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference (*p*-value)[19]	Notes
					0.256 (0.018)**	grant); impact also affected by whether or not school committee members were trained
	SBM reform	Niger	Beasley & Huillery (2014)	Drop-out; Teacher Attendance; Test scores	Overall effect on drop-out = -0.056 (0.286) Overall effect on math scores = -0.048 (0.422) Overall effect on language scores = -0.044 (0.460) Overall effect on teacher attendance = -0.132 (0.092)*	Relatively small grant, particularly in one-teacher schools (where no investments were made in construction, possibly due to small size of grant); No restrictions regarding its use; Evidence that many communities invested in small businesses to gain capital (e.g. agricultural projects), rather than making investments likely to have a direct impact on student learning
Neg impact identified overall; intervention includes grant	WSD	Gambia	Blimpo & Evans (2011)	Teacher Attendance; Test scores	Overall effect on teacher attendance = -0.215 (0.076)* Overall effect on math scores = -0.184 (0.134) Overall effect on language scores = -0.094 (0.670)	Grants linked to School Improvement Plans; both treatment arms included a grant component; had to use grants for teaching/learning activities; slow disbursement of grants likely to have affected outcomes; baseline 'capacity' does not seem to affect the likelihood that grant will have impact
Pos impact identified; does not include	AGEMAD	Madagascar	Lassibille et al. (2010)/Glewwe & Maïga (2011)[22]	Drop-out; Repetition; Test scores	Overall effect on drop-out = -0.027 (0.753) Overall effect on repetition = -0.163	Both studies of AGEMAD find possible effects of school-level intervention; does not include any grant component

[22] As Glewwe & Maiga (2011) did not appear in the forest plots, we can only report a standardised mean difference for Lassibille et al. (2010) in this table. However, both studies found positive effects.

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference (*p*-value)[19]	Notes
grant					(0.045)** Overall effect on math scores = 0.005 (0.801) Overall effect on language scores = 0.001 (0.966)	
	SBM	Philippines	San Antonio (2008)	Test scores	Overall effect on aggregate test scores = 0.121 (0.173)	Overall positive effect; no grant component
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.						

This comparison shows a mixed picture, in terms of the potential impact of including grants as a component of school-based decision-making reforms. Although a number of studies show positive impact of reforms including grants, others show mixed – or even negative – impacts. The studies investigating the AGEMAD programme in Madagascar and the early version of the SBM reform in the Philippines (neither of which included a grant), meanwhile, suggest that school-based decision-making reforms can be effective without providing grants to schools.

It is perhaps unsurprising that we cannot draw any firm conclusions around the importance of incorporating grants into school-based management reforms, as the particularities of the grant elements are themselves likely to have a differential impact. For instance, the size of the grant is likely to matter, as does any restrictions around their use. As discussed in Beasley & Huillery (2014), small grants may have little impact in some contexts, as may grants that can be spent on anything within the school (as opposed to being restricted to expenditures likely to have a direct impact on learning). The manner in which grants are disbursed to schools is also likely to affect the impact of the programme.

Training

We turn next to the potential impact of **training** school personnel and/or school committee members as an explicit component of school-based decision-making reforms.

In addition to the Pradhan et al. (2011) study discussed above, three other experiments included in the review explicitly investigate the marginal impact of incorporating a training element into a school-based decision-making intervention (Blimpo & Evans, 2011; Bold et al., 2013; Duflo et al., 2012). The results of these experiments are presented in Table 8. As these results offer comparisons *within* studies, the original results are shown, rather than the standardised effects.

Table 8: Summary of experimental evidence on training

Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
ETP	Kenya	Duflo et al. (2012)	Teacher Absenteeism; Test scores	Overall effect of ETP on math scores = 0.135* (0.075) Effect of ETP, plus training of school-management committees, on math scores = 0.207*** (0.076) Overall effect of ETP on language scores = 0.191** (0.095) Effect of ETP, plus training of school-management committees, on language scores = 0.198** (0.100)	Results found on page 41; method = average treatment effect, with interaction terms Some arms in the experiment include training of SMC members; others do not. Training found to have an effect. Authors surmise training is particularly important for mitigating elite capture in hiring of contract teachers (i.e. hiring of relatives for positions). Training also provides an opportunity for SMC members to identify which community members are responsible for monitoring teacher attendance
		Bold et al. (2013)	Test scores	Overall effect of ETP with local hiring on aggregate test scores = 0.057 (0.090) Effect of ETP, with SMC training, on aggregate test scores = 0.122 (0.094)	Results found on page 40; method = intent-to-treat Stronger effect identified (in NGO arm) when grants for contract teachers combined with local hiring and training of SMC members; training in government arm less effective
WSD	Gambia	Blimpo &	Teacher	Overall effect of grant on teacher	Training to community members included as part of initiative, although using a

Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
		Evans (2011)	Attendance; Test scores	absenteeism[23] = -0.22 (1.76) Effect of grant plus WSD training on teacher absenteeism = -3.11* (1.75) Overall effect of grant on math scores = -0.09 (0.07) Effect of grant plus WSD training on math scores = -0.12 (0.08) Overall effect of grant on language scores = -0.13 (0.08) Effect of grant plus WSD training on language scores = -0.04 (0.09)	training-of-trainers cascade model; training adapted to local languages
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.					

[23] As teacher *absenteeism* considered in study, sign reversed prior to standardisation for forest plots

Both studies of ETP in Kenya suggest that training increases the impact of the programme. However, this result is not replicated in Blimpo and Evans (2011), who find that, although training seems to increase the impact on teacher attendance, it does not appear to have a similarly positive effect on student learning (as measured through test scores).

In addition to this experimental evidence, it was possible to compare studies of reforms with and without a training element, as we did when examining the potential impact of grants. Table 9 presents a summary of the studies investigating interventions including training. As in Table 7, we show the standardised effects here, as we are looking across studies.

Table 9: Summary of evidence relating to training

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference[24]	Notes
Pos impact identified overall; intervention includes training	EDUCO	El Salvador	Jimenez & Sawada (2003)	Repetition	Overall effect on repetition = -0.039	Intervention includes an explicit training component
	PEC	Mexico	Bando (2010)	Drop-out; Test scores	Overall effect on drop-out = -0.045 (0.025)** Overall effect on math scores = 0.081 (0.008)*** Overall effect on language scores = 0.065 (0.027)**	Head teacher trained as part of initiative
			Murnane et al. (2006)	Drop-out	Overall effect on drop-out = -0.068 (0.050)**	
			Skoufias & Shapiro (2006)	Drop-out; Repetition	Overall effect on drop-out = -0.069 (0.009)*** Overall effect on repetition = -0.104 (>0.001)***	
	PEC-FIDE	Mexico	Santibanez et al. (2014)[25]	Drop-out; Test scores	Overall effect on drop-out (Grade 3) = -0.020 (0.920) Overall effect on math scores (Grade 3) = 0.282 (0.054)* Overall effect on language scores	Training for head teachers and SMC heads provided as part of initiative

[24] The same caution as that specified for Table 12 applies here; these results show the overall effect of school-based decision-making for interventions with and without training. They do not show the effect of training specifically.

[25] Positive impact only identified in Grade 3 sample

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference[24]	Notes
					(Grade 3) = 0.481 (0.001)***	
	TEEP	Philippines	Khattri et al. (2010)	Test scores	Overall effect on math scores = 0.110 (0.097)* Overall effect on language scores = 0.097 (0.026)**	Head teachers trained as part of initiative
			Yamauchi & Liu (2012)	Test scores	Overall effect on math scores = 0.297 (<0.001)*** Overall effect on aggregate test scores = 0.287 (<0.001)***	
	BESRA	Philippines	Yamauchi (2014)	Test scores	Overall effect on aggregate test scores = 0.315 (0.247)	School staff trained as part of initiative
			World Bank (2013)	Test scores	Overall effect on math scores = 0.343 (<0.001)*** Overall effect on aggregate test scores = 0.339 (<0.001)***	
	AGEMAD	Madagascar	Lassibille et al. (2010)/Glewwe & Maïga (2011)[26]	Drop-out; Repetition; Test scores	Overall effect on drop-out = -0.027 (0.753) Overall effect on repetition = -0.163 (0.045)** Overall effect on math scores = 0.005 (0.801)	School staff trained as part of initiative. Training provided directly to school staff (not via district or sub district officials)

[26] Only results from Lassibille et al. (2010) are reported here, as we did not standardise the results of Glewwe & Maiga (2011)

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference[24]	Notes
					Overall effect on language scores = 0.001 (0.966)	
	SBM (initial)	Philippines	San Antonio (2008)	Test scores	Overall effect on aggregate test scores = 0.121 (0.173)	Training provided to all participants, although content of training differed depending on treatment arm (those receiving intervention trained on collaboration; those not receiving intervention trained on rights as education stakeholders)
Mixed impact; intervention includes training	EDUCO	El Salvador	Sawada & Ragatz (2005)/Jimenez & Sawada (1999)[27]	Teacher attendance; Test scores	Overall effect on teacher attendance = -0.560 Overall effect on math scores = 0.065 Overall effect on language scores = 0.012	Intervention includes an explicit training component
	AGE	Mexico	Gertler et al. (2012)	Drop-out; Repetition	Overall effect on drop-out = 0.022 (0.322) Overall effect on repetition = -0.055 (0.134)	Community members trained as part of initiative
	PER	Colombia	Rodriguez et al. (2010)	Drop-out; Test scores	Overall effect on drop-out = -0.232 Overall effect on math scores = -0.019 Overall effect on language scores =	Training of teachers included as part of initiative. Authors estimate a probit model of success, weighted by total number of students in a school, and find that PER's success depends on a combination of three factors: good

[27] Only results from Sawada & Ragatz (2005) are reported here, as we did not standardise the results of Jimenez & Sawada (1999)

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised Mean Difference[24]	Notes
					0.102	training, high quality of educational material, and 'first rate' school management.
	SBM reform	Niger	Beasley & Huillery (2014)	Drop-out; Teacher Attendance; Test scores	Overall effect on drop-out = -0.056 (0.286) Overall effect on math scores = -0.048 (0.422) Overall effect on language scores = -0.044 (0.460) Overall effect on teacher attendance = -0.132 (0.092)*	Training to community members included as part of initiative
Neg impact identified overall; intervention includes training	WSD	Gambia	Blimpo & Evans (2011)	Teacher Attendance; Test scores	Overall effect on teacher attendance = -0.215 (0.076)* Overall effect on math scores = -0.184 (0.134) Overall effect on language scores = -	Training to community members included as part of initiative, although using a training-of-trainers cascade model; training adapted to local languages
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.						

As with the evidence relating to grants, the comparison presents a mixed picture, in terms of the importance of providing training as part of school-based decision-making reforms. Intuitively, it would seem important to train school personnel and community members on any new decision-making responsibilities within the context of a devolution reform; this may be the reason why nearly all of the interventions incorporate some training component. Rather than a discussion of *whether* training should be included, therefore, it seems more important to discuss the *manner in which* training is provided. Although there is no systematic evidence from this group of studies to support any conclusions around *who* should be trained (i.e. school personnel or community members), there is evidence to suggest that the *trainers* may matter. In particular, the two studies investigating AGEMAD (Glewwe & Maïga, 2011; Lassibille et al., 2010) suggest that training must be provided directly to schools in order for school-based decision-making reforms to have a positive effect, as a 'train the trainers' cascade model led by the district or sub-district employees was not found to be effective.

Accountability mechanisms (e.g. report cards)

The next factor addressed by a few of the included studies is the incorporation of an **accountability mechanism** as an explicit component of school-based management reform. There is already a substantial body of literature on the impact of accountability mechanisms on educational outcomes. As this review focuses on changes in decision-making authority, rather than on mechanisms that might improve the functioning of existing school-level decision-making structures, we have not reviewed much of this literature.[28] However, one of the experiments in the review does explicitly consider the marginal impact of adding a report card to a school-based decision-making intervention (World Bank, 2011). Surprisingly, the study finds that the addition of the report card actually reduced the impact of the intervention, rather than increasing it. Table 10 outlines the results of the study (in the original scale).

[28] A recent review commissioned by the World Bank (Bruns et al, 2011) provides an excellent overview of this literature.

Table 10: Results of World Bank (2011)[29]

Outcome	*Results of PSI programme*	*Results of PSI programme with additional report card element*
Teacher absenteeism	9.592 (6.490)	6.505 (5.866)
Math test scores	0.220*** (0.0767)	0.0321 (0.0789)
Language test scores	0.226*** (0.0712)	-0.0806 (0.0715)

Notes: *** , **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.

In addition, five other included studies discuss interventions which include school report cards. Table 11 presents a summary of these five studies. As with the other tables showing standardised effects, the results do not explicitly demonstrate the impact of including report cards; they show the overall impact (standardised across studies) for interventions with and without a report card element.

[29] Results found on pages 18 and 19; method = fixed effects regression

Table 11: Summary of evidence relating to report cards

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised mean difference	Notes
Positive impact identified overall; intervention includes report card	TEEP	Philippines	Khattri et al. (2010)	Test scores	Overall effect on math scores = 0.110 (0.097)* Overall effect on language scores = 0.097 (0.026)**	Report card included as part of initiative
			Yamauchi & Liu (2012)	Test scores	Overall effect on math scores = 0.297 Overall effect on aggregate test scores = 0.287	
	AGEMAD	Madagascar	Lassibille et al. (2010)/ Glewwe & Maïga (2011)[30]	Drop-out; Repetition; Test scores	Overall effect on drop-out = -0.027 (0.753) Overall effect on repetition = -0.163 (0.045)** Overall effect on math scores = 0.005 (0.801) Overall effect on language scores = 0.001 (0.966)	Report card included as part of initiative
Mixed effect; intervention includes report card	Scorecard	Uganda	Barr et al. (2012)	Teacher Attendance ; Test scores	Overall effect on teacher attendance = 0.172 (0.047)**	Both arms of experiment include report card; difference identified between participatory and standard arms indicates a positive impact of giving participants a voice in the development of the report card
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.						

[30] Only results from Lassibille et al. (2010) are reported here, as we did not standardise the results of Glewwe & Maiga (2011)

It is difficult to synthesise the evidence relating to the incorporation of accountability mechanisms as a part of school-based decision-making reforms, as the one study showing a negative result (World Bank, 2011) does not offer any explanation as to why schools receiving the added element of a report card might have performed worse in the evaluation than did those who did not. The other studies considering interventions with a report card element (i.e. those looking at the TEEP programme in the Philippines and the AGEMAD programme in Madagascar) show positive effects, although it is unclear if any of the observed impact can be attributed to the report card itself. The only study to explicitly consider the manner in which report cards are developed and used (Barr et al., 2012) suggests that report cards developed through a participatory process are likely to have a positive impact, while those developed by central authorities are not. Barr et al. also argue that accountability mechanisms, such as report cards, are likely to be particularly effective in contexts where accountability is generally low.

Elections

The final implementation factor relevant to a number of interventions in the sample is the mechanism through which school management committee members are selected, i.e. whether **elections** are organised to fill posts on committees. No experiments explicitly consider the marginal impact of elections, except for Pradhan et al. (2011). Furthermore, very few studies even discuss the mechanism through which committee members are selected. However, the overall standardised effects from those that do are compared in Table 12 below.

Table 12: Summary of evidence relating to elections

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Standardised mean difference	Notes
Positive impact identified overall; SMC members elected	EDUCO	El Salvador	Jimenez & Sawada (2003)	Repetition	Overall effect on repetition = -0.039	Elections held to fill school management committee positions
			Sawada & Ragatz (2005)/Jimenez & Sawada (1999)[31]	Teacher Attendance; Test scores	Overall effect on teacher attendance = -0.560 Overall effect on math scores = 0.065 Overall effect on language scores = 0.012	
	PSI	Sri Lanka	World Bank (2011)	Teacher Absenteeism; Test scores	Overall effect on teacher attendance = 0.519 (0.140) Overall effect on math scores = 0.213 (0.004)*** Overall effect on language scores = 0.230 (0.002)***	SMC members elected as part of the reform
Mixed impact identified overall; SMC members elected	SBM reform	Niger	Beasley & Huillery (2014)	Drop-out; Teacher Absenteeism; Test scores	Overall effect on drop-out = -0.056 (0.286) Overall effect on math scores = -0.048 (0.422) Overall effect on language scores = -0.044 (0.460) Overall effect on teacher attendance = -0.132 (0.092)*	SMC members elected as part of the reform
Notes: ***, **, * indicates findings are statistically significant at 99%, 95% and 90% confidence levels.						

[31] Only results from Sawada & Ragatz (2005) are reported here, as we did not standardise the results of Jimenez & Sawada (1999)

The results pertaining to elections are inconclusive, as the sample includes studies showing both positive and mixed effects of reforms including election components.

Implementing body

The final factor to consider in this sub-section is the **body responsible for implementing the reform**. This factor is not considered by most of the studies, as most examine the impact of individual interventions. However, one study (Bold et al., 2013) considers this factor in detail and concludes that the implementing body is the single most important implementation factor affecting outcomes. Bold et al. exploit the unusual circumstance arising in Kenya in 2009, in which a contract teacher reform, initially implemented by an NGO in the Western part of the country, was adopted by the central government and scaled up to the national level within the time frame of the NGO programme evaluation. As a result of these unique circumstances, the authors were able to examine the differential impact of the programme depending on the implementing body. Their results suggest that, although the programme was quite effective when implemented by the NGO, it had no impact when implemented by the government [effect of government implementation = -0.163 (0.095)*; effect of NGO implementation = 0.184 (0.088)**)]. [32] As with the results of the Pradhan et al. (2011) experiment (outlined above), these results must be treated with caution, as they only pertain to one of the included studies – and, in fact, many of the studies showing positive impact pertain to reforms implemented by central government authorities (albeit often with the support of the World Bank). However, this is not universally the case. The studies of the AGEMAD programme in Madagascar (Glewwe & Maïga, 2011; Lassibille et al., 2010) indirectly support Bold et al.'s conclusion, as they acknowledge that the school-level trainings (found to have the greatest impact) were provided by an NGO. Although not discussed by the authors, this could be a crucial factor in the results, given that no effect was identified in the treatment arms relying on district and sub-district level authorities to implement the reform. Although not mentioned in reference to this particular point, Beasley & Huillery (2014) suggest in their study that school-based management reforms were ineffective in Niger because of a preference amongst community members for central government control over public services. Although we cannot draw any firm conclusions around this point, it appears that government-led reforms may be more (or less) effective depending on the context and, in particular, depending on the relationship between central and local authorities and the existence of strong or weak accountability within the overall education system.

[32] Results found on page 39; method = intent-to-treat

4.8.6. Other factors

Finally, two additional factors are likely to affect the results of the impact studies considered in this review: the level of compliance with the proposed intervention, and the time elapsed between the implementation of a given reform and the study investigating its impact.

Unfortunately, we have very little information relating to the level of compliance, as most studies do not report on this factor. There are, however, a few exceptions. Pradhan et al. (2011) note that, due to resistance to the reform in some communities in Indonesia, only some of the treatment communities intended to implement elections did so in practice. Blimpo & Evans (2011) acknowledge that the slow disbursement of grant monies to both groups of treatment schools resulted in differential exposure, as some communities received their grants much earlier than others. Within the government arm of their study, Bold et al. (2013) also acknowledge imperfect compliance with some of the specifications of the contract teacher evaluation, namely that certain schools did not retain contract teachers within a specific year, thereby leading to likely spill-over effects on students in other years. Finally, the 2013 study of BESRA in the Philippines, conducted by the World Bank, includes a brief comment on the high level of compliance with the policy. As Yamauchi (2014) examines the same policy, one can assume that his results also reflect a high level of compliance with the intended intervention.

It was, however, possible to examine the possibility of differential impact, depending on the length of exposure to the reforms under investigation. As discussed in the introduction to this report, studies in the U.S. have indicated that school-based management reforms are unlikely to have an impact on test scores until they have been established for at least eight years. This could be because schools initially see a decline in performance as school personnel adapt to the new structures, or because school-based management reforms are likely to have a more immediate impact on proximal outcomes (e.g. teacher attendance), which then have a more gradual impact on student learning over time. In the forest plots in Sections 4.4 and 4.5, we include the follow-up time for longitudinal studies with an endline and a baseline. However, follow-up time is not necessarily the same as the length of exposure to a particular intervention; some studies take data from a year or two prior to the implementation of a reform as their baseline, which results in unequal follow-up time and length of exposure, whereas cross-sectional studies always have different follow-ups and exposure lengths, given that their lack of baseline results in a notation of ‘zero’ for follow-up time on the forest plots. Generally, this factor was not explicitly acknowledged in the studies. However, seven of the studies do explicitly include time-lag in their heterogeneity analysis. The results of these studies are presented in Table 13 below (in their original scale).

Table 13: Summary of time-lag effects

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
Short exposure => pos impact	EDUCO	El Salvador	Jimenez & Sawada (1999)	Test scores	Overall effect on **math scores** = 0.40 (0.27) Effect on math scores for schools built prior to 1995 = -1.93 (0.97) Effect on math scores for schools built in 1995 = 3.21 (1.75)* Effect on math scores for schools built in 1996 = -0.49 (0.28) Overall effect on **language scores** = 1.57 (1.51) Effect on language scores for schools built prior to 1995 = 0.82 (0.59) Effect on language scores for schools built in 1995 = 3.26 (2.55) Effect on language scores for schools built in 1996 = 0.43 (0.35)	Results found on page 432; method = fixed effects regression Find highest impact for those communities joining the programme in 1995 (one year prior to the study), although coefficients not significant; suggests this may be evidence of a possible Hawthorne effect

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
	ETP	Kenya	Duflo et al. (2012)	Test scores	Effect on **math scores** for students of ETP contract teachers in schools with school-based management committees = 0.237*** (0.087) Effect after an additional year of exposure = 0.080 (0.074) Effect on **language scores for** students of ETP contract teachers in schools with school-based management committees = 0.256** (0.108) Effect after an additional year of exposure = 0.075 (0.100)	Results found on page 43; method = OLS regression Finds evidence of fadeout of effect on test score over time
	AGE	Mexico	Gertler et al. (2012)	Drop-out; Repetition	Overall effect on **repetition** = -0.004* (0.002) Effect on students in Grades 1, 2 or 3 *after one year* = -0.007** (0.002); *after more than one year* = -0.007* (0.003) Effect on students in Grades 14 and 5 *after one year* = -0.001 (0.002); *after more than one year* = 0.003 (0.003) Overall effect on **drop-out** =	Results found on page 74; method = fixed effects regression Finds that impact is mostly identified in first year of the program; impact does not appear to change over time

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
					0.001 (0.002) Effect on students in Grades 1, 2 or 3 *after one year* = 0.000 (0.002); *after more than one year* = -0.000 (0.002) Effect on students in Grades 4 or 5 *after one year* = 0.003 (0.002); *after more than one year* = 0.002 (0.002)	
Long exposure => pos impact	PEC	Mexico	Bando (2010)	Drop-out; Test scores	Effect on **drop-out** of schools with no prior experience in programme = -0.015 (0.015) Effect on drop-out of schools after 5 years in programme = -0.155 (0.069)** Effect on **math scores** of schools with no prior experience in programme = -0.023 (0.012)** Effect on math scores of schools after 5 years in programme = 0.081 (0.030)*** Effect on **language scores** of schools with no prior experience in programme =	Results found on page 80 and 81; method = fixed effects regression Finds stronger effect for those schools that have been in the programme the longest (i.e. the maximum of five years)[33]

[33] Results for interim years not included due to space constraints; full results available in original paper.

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
					-0.024 (0.012)** Effect on language scores of schools after 5 years in programme = 0.065 (0.030)**	
			Murnane et al. (2006)	Drop-out	Effect on **drop-out** after 1 year in programme = -0.056 Effect on drop-out after 2 years in programme = -0.165*** Effect on drop-out after 3 years in programme = -0.274***	Results found on page 42; method = difference-in-difference with fixed effects Size of effect is correlated with the number of years in the programme; the increase in impact over time holds true for communities at all levels of development (although the starting points differ)
	Autonomous Schools	Nicaragua	King & Ozler (2005)	Test scores	Results of 1st stage OLS regressions for *de facto* autonomy: # of years since autonomous school established = 0.004** (0.002)	Result found in Appendix E on Table E7; method = OLS regression The time lag between the establishment of a school and the date of data collection is positively correlated with the impact of the programme
No significant difference	AGEMAD	Madagascar	Glewwe & Maïga (2011)	Test scores	Overall effect of school-level intervention on test scores after five months in programme = 0.009 (0.151)	Results found on pages 4 and 5; method = fixed effects regression Collects data at two points –

Differential Impact	Name of Intervention	Country	Citation	Outcomes	Results	Data source and interpretation of results
					Overall effect of school-level intervention on test scores after two years in programme = 0.068 (0.134)	once after five months and once after two years – and finds no significant effect at either point

The evidence on this point is inconsistent. Some studies (e.g. Duflo et al., 2012; Gertler et al., 2012; Jimenez & Sawada, 1999; and Santibanez et al., 2014) identify a possible 'Hawthorne effect', whereby schools show positive results in the first year (possibly due to the energy and momentum created by the new reform), which do not continue to increase with prolonged exposure. A similar effect is identified in Khattri et al. (2010) and Yamauchi (2014), although neither study explicitly presents data on this point. However, other studies (e.g. Bando, 2010; King & Ozler, 2005; Murnane et al., 2006) identify stronger results in communities with longer exposure to the intervention. As studies in both groups examine similar outcomes, it is difficult to draw any conclusions around the differential impact of length of exposure.

4.9. ENABLERS AND BARRIERS

In this section, we attempt to provide some answers to the second review question – "What are the barriers to (and enablers of) effective models of school-based decision-making?" – by combining the results of the heterogeneity analysis with relevant qualitative evidence from the included studies. As a few of the impact studies used mixed methods, some of the qualitative evidence cited here comes from the impact studies discussed in the previous sub-sections, but here we also draw on evidence from the nine non-causal studies included in the review.

4.9.1. Barriers to effective school-based decision-making

We start with the potential barriers to impact identified by the included studies.

First, it appears that **poverty** can act as a barrier to effective school-based decision-making reforms. As discussed in the previous section, a number of impact studies suggest that devolving decisions to the school level does not have a positive effect on the poorest, most disadvantaged communities. This finding is also supported by evidence from some of the non-causal studies in the sample. In Nicaragua, for instance, Fuller & Rivarola (1998) found that schools in severely impoverished areas were, unsurprisingly, unlikely to raise additional revenue from the surrounding communities. In the same context, Gershberg & Meade (2005) found parental contributions to be a significant component of autonomous school budgets, suggesting that disadvantaged communities without access to such additional monies would be unlikely to experience similar benefits under the autonomous schools model.

This finding is likely to be linked to the evidence suggesting that **low levels of 'capacity' within communities** also act as a barrier to impact. Communities with high levels of illiteracy and/or with few educated parents do not seem to benefit from devolution of decisions to the community level. In their study of Whole School Development programme in the Gambia, Blimpo & Evans (2011) go so far as to argue that devolution may be detrimental in such contexts:

> "In countries where [the gap in capacity between local and central levels] is small … a decentralized policy would be superior because of the added value of localized information. However, if the gap is sufficiently high in favor of the central government, then the localized information plays less of a role because the communities are not well equipped to act on them." (p. 29)

In their cross-country study, Hanushek et al. (2011) reach a similar conclusion, arguing that autonomy reforms improve student achievement in more developed countries but actually undermine it in less developed areas. Reimers & Cardenas (2007) expand this argument by suggesting that *schools* must also have a certain baseline capacity in order to benefit from school-based decision-making reforms. In their analysis of Mexico's PEC programme, they find that leadership and 'coherence of vision among school staff' can act as significant enablers – or barriers – to impact (p. 38). Considering this question from the perspective of teachers, Bjork (2003) found that teachers in Indonesia felt they did not have the capacity to implement the curricular component of that country's school-based management reform points, nor did they feel adequately supported to use the autonomy given to them. As schools in wealthier areas are more likely to begin school-based management reforms at a higher baseline institutional capacity, this reinforces the argument that school-based decision-making is more likely to benefit more advantaged communities.

There are a variety of reasons why the capacity of institutions and communities can act as a barrier to effective school-based decision-making reforms. First, in order for such reforms to be effective, school personnel and community members must understand the nature of the reform and crucially must also be able to propose changes that are likely to affect student learning within the school. There is evidence from a number of studies that neither of these conditions is met in many lower-income contexts. Although both studies identify overall positive impact of school-based management reforms, Santibanez et al. (2014) and Parker (2005) note that communities in Mexico and Nicaragua did not always fully grasp the nature and the objective of school-based decision-making reforms in those two countries. Bandur (2008) raises similar concerns in his analysis of the national school-based management reform in Indonesia. In the Nicaraguan context, this lack of understanding was actually found to translate into active resistance in certain communities (Fuller & Rivarola, 1998). Pradhan et al. (2011) also identify resistance to the election of school committee members within some communities in Indonesia, although it is not clear if this resistance was the result of a lack of understanding or an active attempt to block potential changes to the status quo. Beasley & Huillery (2014) note that, although school-based management reforms assume that community members know what should be done to improve educational outcomes, the evidence suggests that this is not always the case. In their study, they find that school management committees in rural communities frequently opted to spend their grants on agricultural projects, instead of school materials, teacher incentives or other initiatives likely to affect educational outcomes. In a credit-constrained environment such as Niger, it is unsurprising that communities might choose to invest grants in projects that can be used to generate income in the long term; however, although potentially a wise economic decision, such investment is unlikely to improve student learning in the region. In a very different

context, Di Gropello & Marshall (2005) note a similar barrier, as they argue that parents with little or no formal education residing in rural areas may find it difficult to even know how much learning is actually taking place in schools, never mind know what might need to be done to address any deficiencies. Secondly, community members – particularly parents - must have a certain amount of status in order to play an active role on school management committees. As discussed in Beasley & Huillery (2014) and in Gertler et al. (2012), this does not tend to be the situation in rural, poor communities, where school personnel are often perceived as authority figures due to their relatively high levels of education. This political dynamic is likely to limit active participation in school decisions and result in the formation of committees that simply 'rubber stamp' decisions made by school personnel. All of these reasons may explain why early interventions devolving decisions to the school level, such as EDUCO in El Salvador, restricted participation in school management decisions to literate members of the community, a requirement which does not appear to feature in similar models of school-based management implemented more recently in other low income contexts.

Another potential barrier highlighted by the included studies is the **potentially limited effectiveness of government-led reforms in some contexts.** As discussed in the previous section, the study examining this barrier in detail is Bold et al. (2013), which finds that a contract teacher programme demonstrating strong evidence of impact when implemented by an NGO had no effect when implemented by the government at the national level. Bold et al. suggest that this is at least partially due to the limited capacity of under-resourced governments to monitor the implementation of complex reforms. Although they do not frame their analysis in a similar fashion, Lassibille et al. (2010) and Glewwe & Maïga (2011) indicate a similar result in their analysis of the AGEMAD programme in Madagascar, as they only find evidence of impact within schools benefiting from direct training by NGO representatives. No impact could be identified within schools that had been trained by district or sub-district employees (who had themselves been trained by the NGO). As Madagascar also struggles with weak monitoring within the government system, this may be indicative of the limited capacity of district and sub-district officials to implement the reform without assistance. This is an important finding, given that governments often opt to scale up reforms based on pilot studies in which NGOs have played an active role in implementation. Such programmes are unlikely to have a similar impact at the national level without sufficient monitoring capacity and accountability mechanisms, both of which are often limited in low income contexts. Indeed, there may be reason to suspect that government officials may actively hinder the effectiveness of school-based management reforms, as was identified by both Bandur (2008) and Vernez et al. (2012) in Indonesia, where provincial and district officials were found to actively interfere in school decision-making processes.

Another interpretation of this finding is that communities are only likely to benefit from autonomy over school decisions if there is already an **active desire for autonomy within the community**. In their study of eight Latin American countries (Argentina, Bolivia,

Brazil, Chile, Colombia, Dominican Republic, Honduras and Peru), Gunnarsson et al. (2008) investigate the relationship between school autonomy and student test scores in math and language. They determine that school autonomy (as defined by formal decision-making authority) and parental/community participation are not highly correlated, suggesting that local authority over educational decisions is as much a matter of local choice as central policy. Although school autonomy alone does not seem to have a significant impact on student test scores, parental participation does, once controls for endogeneity are put in place. They conclude that decentralisation to schools is a beneficial policy when communities demonstrate an interest in participating in educational decisions but that, if such interest is not evident, central decision-making may be more effective. King & Ozler's (2005) analysis of *de jure* versus *de facto* autonomy within communities supports the same conclusion, as does Jimenez & Sawada's (1999) investigation of the impact of community participation levels within EDUCO schools.[34]

Finally, the studies highlight the fact that school-based decision-making reforms can only affect the immediate circumstances of a given school or community. Even in the event that a reform is effective within a community, school-based management reforms cannot address many **external factors** that can act as significant barriers to impact. Although there are myriad external factors affecting educational outcomes, the included studies reference five that appear to have a strong effect, at least in some contexts:

1) The strength of the national teacher's union

Bold et al. (2013) argue that the strength of Kenya's teachers union was one of the reasons for the relative failure of the national scale-up of the contract teacher programme. Once the programme was implemented at the national level, there was strong political backlash from the union, and their mobilisation of civil service teachers against the reform appears to have been a major factor in its limited success. Although not explicitly examined in their study, King & Ozler (2005) note that one reason for the success of the Autonomous Schools initiative in Nicaragua in the late 1990s was the low likelihood of strike activity following the 1990 election. When school-based decision-making reforms change teacher conditions and hiring/firing practices, teachers unions are likely to get involved and, potentially, limit any possible impact. This factor is only likely to affect high decentralisation contexts, in which personnel decisions are devolved to the school level.

2) The strength of the teacher job market

Another factor likely to limit the impact of reforms devolving personnel decisions is the strength of the teacher job market in the region. Barr et al. (2012) note that a shortage of teachers tends to reduce the willingness of school management committees to exercise their authority to fire ineffective teachers, given the potential lack of a suitable replacement. Parker (2005) discusses the same factor in her study.

[34] EDUCO schools are often upheld as a model of community participation, as there is clear evidence of higher levels of parental participation in EDUCO, versus traditional public, schools (Sawada & Ragatz, 2005; de Umanzor et al, 1997).

3) Teacher ability

Learning outcomes are unlikely to improve as a result of school-based management reforms if the teachers are simply not equipped to teach certain subjects. Lassibille et al. (2010) highlight this factor as a potential reason why students in their sample improved in math and Malagasy but not in French, a subject they argue that many teachers in Madagascar are ill-equipped to teach. Blimpo & Evans (2011) also discuss this as a barrier to impact in the Gambian context.

4) Constraints imposed by the central system

Teachers within schools are often affected by central-level decisions, even within decentralised contexts. Teacher attendance, for instance, is often the result of inefficient mechanisms for distributing salaries in rural areas. Although teachers in some contexts may be absent because of low motivation or limited interest in the profession, many miss school for legitimate reasons, including travelling to banks in regional or provincial capitals in order to collect their salaries. In such contexts, school-based decision-making reforms can only have a limited impact on teacher attendance, as teachers will still need to miss school on pay-day (as discussed in Blimpo & Evans, 2011; and Lassibille et al., 2010). Blimpo & Evans (2011) also mention the negative impact of the shift system in over-crowded areas, an efficiency reform often implemented by central authorities in resource-constrained contexts.

5) Security

The security of a region can also act as a barrier to impact. Although no studies in this review analyse the impact of school-based decision-making reforms on conflict-affected areas, many reference security in passing, generally in reference to areas *not* included in the study catchment area. It is important to remember that conflict (or the threat of conflict) is likely to have a negative impact on school-level decision-making, particularly given that studies often explicitly avoid conducting data collection in hard to reach and/or insecure areas. Pradhan et al. (2011), for instance, note that their study was conducted in a "peaceful, well-resourced area", while Beasley & Huillery (2014) opted to exclude certain communities from the data collection in their evaluation following the outbreak of conflict in some regions of Niger. The exclusion of insecure areas from any evaluation of a school-based management reform is likely to upwardly bias the results, so this is an important factor to consider when interpreting the results of the individual studies.

4.9.2. Enablers of effective school-based decision-making

In addition to highlighting a number of potential barriers, the included studies point to a number of enablers of effective school-based decision-making reforms.

First, it appears that **smaller schools** are particularly likely to benefit from local decision-making authority, likely because it is easier for school management committees to monitor teachers and stay informed about conditions at the school. Beasley & Huillery (2014) note

that the only schools in their sample that benefited from school-based management were the one-teacher schools, with teacher attendance tending to improve following the implementation of the reform. School management committees in these contexts were more likely to use their grants to support benefits for the teachers, and the authors conjecture that this may be because parents in one-teacher-school communities may recognise that they are highly dependent on the teachers' continued motivation and are therefore more likely to establish an alliance with the teacher, instead of an adversarial relationship. This may, in turn, have a positive impact on teacher behaviour in these communities.

Second, it seems that **devolving personnel decisions**, in addition to financial and other management decisions, enables the possibility that school-based decision-making will affect teacher behaviour, including teacher attendance. Although other forms of decentralisation may be useful in other ways, it appears to be necessary to give schools and communities some control over hiring and firing of teachers in order to have any significant impact on teacher absenteeism. Sawada & Ragatz (2005) credit this aspect of the EDUCO programme with much of its success, as do King & Ozler (2005) in reference to Nicaragua's Autonomous Schools programme. The effectiveness of such models, however, appear to depend at least partially on the teacher job market. The possibility of long-term employment may also play a role in enabling impact, as teachers hired by school-management committees on short-term contracts may be more motivated if they believe they will ultimately be able to secure longer-term contracts (as discussed in Duflo et al., 2012; and Jimenez & Sawada 2003).

Third, it appears that school-based decision-making reforms are more effective when they incorporate **certain elements**, such as training for committee members. Although the incorporation of such components can act as enablers, it is important to highlight that they must be implemented effectively in order to perform such a function. It does not appear that simply providing a grant or a training programme, incorporating elections or requiring an accountability mechanism such as a report card has a consistently positive impact on outcomes. Rather, additional elements appear to be particularly useful if they incentivise behaviour that is likely to increase motivation and community participation (e.g. by requiring that grants be spent in ways that support teaching or involving the community in the development of the school report card).

Finally, one potentially important enabler is **giving parents the majority voting power on school management committees**. Duflo et al. (2012) suggest that parental majority on Kenyan school management committees is one of the reasons why local hiring addresses issues of elite capture in that context. It was not possible to investigate this potential enabler in any detail in this review, as studies typically indicate that decision-making authority is 'shared' between parents and community members without specifying which groups hold the voting majority. Furthermore, concerns around community capacity remain, in that parental majority may only be an effective enabler in contexts where parents have sufficient status and authority within the community to affect change.

4.10. INTEGRATION OF FINDINGS

As most studies did not include data relating to the full list of barriers and enablers outlined in the preceding section, it was not possible to formally test the impact of these factors on the outcomes of interest in this review. Furthermore, as some of the enablers and barriers pertain to some outcomes and not others (e.g. parental majority as being a potential enabler in terms of teacher attendance but not necessarily student learning), it was not possible to summarise the findings of the review in one coherent table. Instead, we opted to integrate the findings from the two phases of the review by using the data sets to inform a revision of our original conceptual framework (presented in Section 1.3 as Figure 1). This section reports on this revision process.

The first revision to the original framework was to replace the 'mechanisms' with the broad intervention types outlined in Section 4.2 (i.e. 'high', 'medium' and 'low' decentralisation). We then elected to disaggregate the original diagram, by creating individual frameworks depicting the causal pathways relating to two of the intervention types.[35]

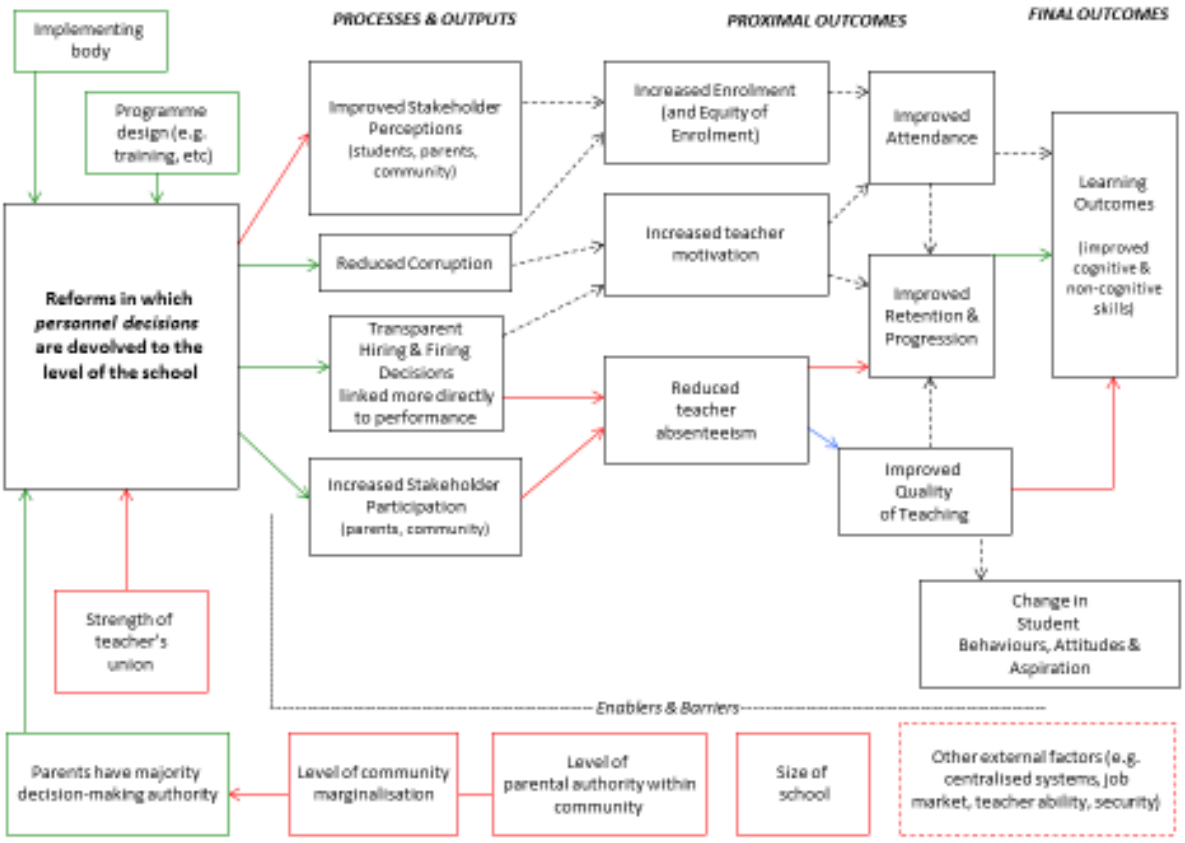
As we did not find evidence of any causal pathways not included in the original diagram, the adapted frameworks do not show dramatically different pathways to impact. They do, however, depict a modified list of enablers and barriers, drawn from the analysis in the preceding sections of this chapter. Furthermore, the revised versions graphically depict the strength of – and gaps in – the evidence base represented by the included studies in this review. Colours are used to denote the **strength of a given causal link**: red arrows are used when a causal link seems sound, based on the evidence; green is used to indicate links which appear to depend on implementation and context; and blue indicates areas where the evidence suggests that the assumed causal link does not necessarily hold. Shading is then used to denote **where we do or do not have evidence** within this review: solid lines are used for links investigated by the included studies, while dashed lines indicate areas where we are missing evidence.

4.10.1. Pathways to impact: devolving personnel decisions to school level

In models of school-based decision-making classified as 'high' decentralisation, schools and communities have decision-making authority over nearly all aspects of school management. Most importantly, the school (or, typically, the school management committee) has authority over both financial and personnel decisions, including the authority to hire/fire teachers and to pay salaries. The pathways to impact relating to this model of school-based decision-making are depicted in Figure 28.

[35] Although we identified three intervention types in the included studies, we created only two adapted frameworks, as the third type ('low' decentralisation) only featured in one of the impact studies.

Figure 28: Adapted Framework A: personnel decisions



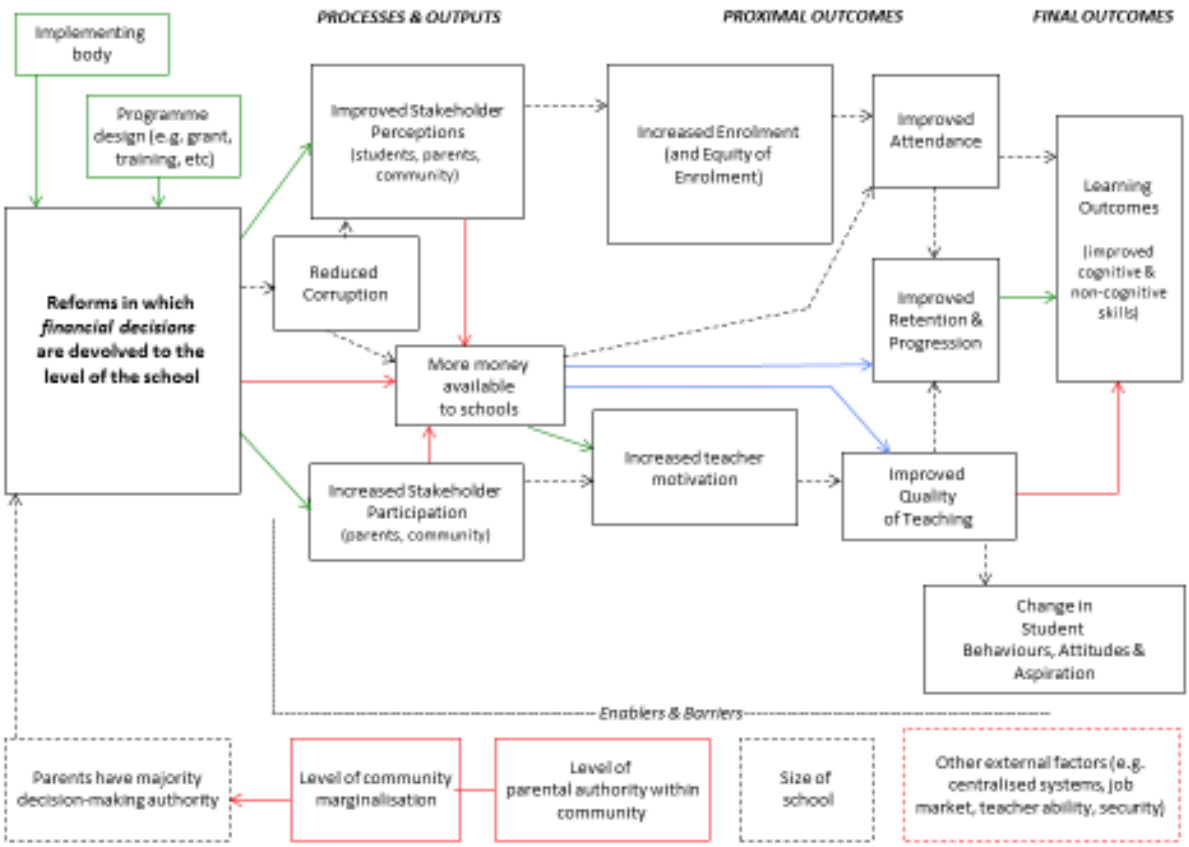
As is evident from the studies examining the impact of differential levels of participation on outcomes, devolving decision-making to school level does not always result in increased stakeholder participation in school activities. However, when participation does increase – and when school management committees have the authority to hire and fire teachers – the evidence suggests that teacher attendance does improve. We know less about how this may translate into student learning. In fact, improved teacher attendance does not appear to result in increased teacher effort or improved quality of teaching in many contexts. The link between teacher attendance and student learning is likely to depend on a number of other external factors, including teacher ability, community characteristics and the specific design of the school-based decision-making reform.

4.10.2. Pathways to impact: devolving financial decisions to school level

In 'medium' decentralisation models, schools do not have the authority to hire and fire teachers. However, they do have authority over non-personnel financial decisions. This authority usually comprises oversight of grants related to School Improvement Plans and/or the school budget, as well as legal authority to raise independent monies on behalf of the school.

Figure 29: Adapted Framework B: financial decisions

The pathways to impact for 'medium' decentralisation reforms are even less clear than those for 'high' decentralisation reforms. There is evidence to suggest that devolving financial



decisions to the school level often results in an increased amount of money available to the school, either due to the receipt of a grant or to the fundraising activities of school management committees. However, increased money does not appear to translate into educational outcomes, particularly in poorer communities.

5. Conclusions

5.1. SUMMARY OF MAIN RESULTS

Overall, we find that devolving decision-making to the level of the school appears to have a somewhat negative effect on drop-out in certain contexts and on repetition when looking across studies.[36] Effects on test-scores are more robust, being positive and significant in the aggregate (between 0.10 and 0.20 SMD), particularly in middle income countries. While pooled effects on teacher attendance are not significant overall, there is some evidence that these effects are stronger in contexts of high decentralisation and of low income.

Most of the included studies do not conduct any sub-group analysis relating to individual characteristics, such as gender and student background; those that do differ in their findings. However, there is some evidence to suggest that school-based decision-making reforms have a stronger impact on wealthier students with more educated parents. It also appears that school-management reforms may be particularly impactful on children in younger grade levels.

School-based decision-making reforms appear to be less effective in disadvantaged communities, particularly if parents and community members have low levels of education and low status relative to school personnel. Devolution also appears to be ineffective when communities do not choose to actively participate in decision-making processes. Small schools, however, may find school-based decision-making interventions to be effective, particularly if community members opt to establish a collaborative, rather than an adversarial, relationship with teachers.

School-based decision-making reforms can be implemented in a variety of ways. Training appears to be an important element of any school-based management reform, although this may be more effective when delivered directly to schools by NGOs, rather than via government authorities, at least in contexts with weak monitoring and accountability mechanisms. Grants do not always have an impact on educational outcomes, although sufficiently large grants targeted explicitly at investments likely to increase learning may have a positive effect.

[36] It is worth reminding the reader that a negative impact is the desired outcome for drop-out and repetition.

Overall, we can conclude that devolving decision-making authority to the school level can have a positive impact on educational outcomes, but that such positive effects are only likely to occur in more advantaged contexts in which community members are largely literate and have sufficient status to participate as equals in the decision-making process.

5.2. QUALITY OF THE EVIDENCE

Although only 27 studies met the criteria for robust studies of impact, the studies themselves were of relatively high quality, with seven classified at low risk of bias and 20 classified at medium risk. We could not identify any significant differences in the effects indicated by low and medium risk studies.

There are, however, two important caveats relating to the quality of the evidence synthesised in this review:

1) Many of the included studies report on small evaluations implemented within particular regions and/or by NGOs or other external actors (e.g. Barr et al., 2012; Pradhan et al., 2011). Considering the results of Bold et al.'s (2013) analysis of NGO-led versus government-led interventions, it is important to acknowledge that the sample of studies included in this review may overestimate the potential impact of school-based decision-making reforms when implemented at a national level.
2) Second, we must acknowledge that there is intense debate within the international development community (and, more explicitly, within the field of economics) around the relative quality of the various methods used in the studies included in this review. The relative rigour and utility of using different techniques for estimating attribution is hotly contested within the field, as is evidenced by the fact that some of the included studies explicitly cross-reference (and question) other studies in the sample. Yamauchi & Liu (2012), for instance, query the control group constructed by Khattri et al. (2010), while Parker (2005) argues that King & Ozler's (2005) study is limited by both selection and attribution bias. Murnane et al. (2006) build explicitly on Skoufias & Shapiro (2006) by adding pre-selection trends as an additional control for selection bias, and Sawada & Ragatz (2005) build on Sawada's previous work (with Jimenez in 1999) by incorporating propensity-score matching into the analysis. We elected to include all studies meeting our risk of bias criteria, regardless of any negative assessments from competing studies in the sample, but we acknowledge that there are ongoing debates around the relative robustness of the various methods utilised by the different authors.

5.3. IMPLICATIONS FOR POLICY AND PRACTICE

Our findings carry a number of implications for policy and practice.

First, the evidence suggests that school-based decision-making reforms in highly disadvantaged communities are unlikely to be successful. The level of parental participation appears to be key and this, in turn, is linked to the real authority/status and cultural capital

of community members. One potentially relevant benchmark is proposed by Blimpo & Evans (2011), who explicitly recommend that communities need a minimum of 45 percent overall literacy in order to benefit from school-based management. This suggests that policy makers are likely to see greater impact of school-management reforms is more advantaged areas, although this raises obvious equity concerns.

Second, the involvement of school management committees in personnel decisions (particularly hiring and firing) appears to play an important role in improving proximal outcomes, particularly teacher attendance. However, the impact of devolving personnel decisions is also likely to be linked to the overall teacher job market and the possibility of long-term employment. Policy proposals should therefore take into account the current and prospective job market conditions for teachers when anticipating the potential impact of school-based decision-making reforms.

Third, the specifics of programme design appears to be crucial. Given the limited evidence on implementation factors in this review, we cannot conclude with certainty that incorporating certain elements (e.g. training or grants) into school-based management reforms are universally advisable. However, it does appear that the details of such supplementary elements (e.g. restrictions on the use of grants; the implementing body responsible for training; etc.) may play an important enabling role. The evidence also suggests that, at least in some contexts, impact on student learning may take longer than is often allowed within evaluation timelines. This suggests that evaluations with longer timelines may be necessary in order to identify any sustained impact. Where donors are involved, this also means that decentralisation reforms may require sustained donor commitment over the long term.

Finally, our review suggests that policy makers should proceed with caution when using the results from small-scale pilot programmes to inform national programming. Although further research is needed on this point, the limited evidence identified within this review suggests that this is an important consideration for policy makers.

5.4. IMPLICATIONS FOR RESEARCH

As evidenced by the large number of titles identified during our initial search, there is a vast literature on school-based management in lower-income contexts. However, much of the existing literature is descriptive in nature, and many of the empirical studies of school-based decision-making reforms that do exist are only able to investigate changes in perception and/or participation within communities. Although we were able to identify a relatively large number of impact studies for this review, the included studies represent limited geographic diversity and focus only on a small number of discrete interventions (some of which are small-scale pilots). There is, therefore, a general need for further robust analysis of the impact(s) of the large-scale (i.e. national) school-based decision-making reforms that have recently been implemented in a range of national contexts. Within this, there is a clear need to examine the potentially negative impacts of these reforms, particularly given the widespread adoption of such policies around the world. The limited data on time effects

identified within this review also suggests that there is scope for further longitudinal investigation of how school-based management reforms play out over time.

Additional research is also needed into the relative impact of different kinds of school-based decision-making interventions. Most of the studies included in this review investigated the impact of school-based management versus no school-based management, as opposed to evaluating the differential impact of different models of reforms. The few exceptions (e.g. Pradhan et al., 2011) offer important insights into the specific effects of different models; there is a need for further investigation in this vein in other countries and regions. Further research into the relationship between the enabling factors – and barriers – highlighted in this review and particular outcomes would also be beneficial, as would additional study of the ways in which formal and informal relationships between parents and teachers differentially affect the outcomes of school-based management interventions in different contexts.

Finally, it is important to acknowledge that, although this review has highlighted a number of potential enablers and barriers, the limited evidence base within the included studies has prevented us from drawing any robust conclusions around the conditions necessary for positive impact. There is a significant body of qualitative evidence that considers these factors, but it was not possible to comprehensively synthesise this body of literature within the constraints of this review. A future review of the same topic, utilising a different review methodology, could usefully complement the findings of this study. There also remains a need for further evidence in order to answer important process and context questions linked to when, why and where decentralisation efforts are likely to be effective.

5.5. LIMITATIONS

We completed this review following the standards for systematic reviewing established by the Campbell Collaboration. As the Campbell criteria require that any study of impact must include a counterfactual (actual or constructed), we were necessarily limited to considering evidence of impact on outcomes included in identified impact studies. We had initially intended to mitigate any possible bias in our results by including evidence from a broader range of literature, but our identification of a relatively large number of impact studies prevented us from accessing the full range of qualitative evidence relating to school-based management. As a result, the review is somewhat limited in its scope. We are particularly aware that we were unable to draw on any studies investigating any negative or unintended consequences of school-based decision-making reforms, given that such outcomes do not feature explicitly in any of the included impact studies. We know that devolving decisions to the level of the school can have negative consequences, such as elite capture and disharmony between ethnic groups, and we note that a few of the impact studies in our sample did identify some unintended consequences of the school-based decision-making reforms under investigation (e.g. Duflo et al. (2012) note that school management committees in Kenya seem to be more likely to hire male teachers; Murnane et al. (2006) identified a significant increase in the administrative burden on schools as a result of the PEC programme in

Mexico). However, we could not discuss these issues in any detail in the review, given the focus of the impact studies identified. Our focus on quantitative studies may also have precluded our ability to discuss outcomes usually considered harder-to-measure.

The review team was also limited by time and resource constraints, which necessitated a number of decisions which may have restricted the breadth of our review findings. First, our inability to complete forward citation chasing during the search phase of the review may have limited our ability to synthesise current evidence not yet available in the public domain. Second, our decision to focus only on qualitative evidence relating to interventions discussed in the impact literature necessarily limited our ability to discuss a broader range of contextual and implementation factors.

A recent paper by Evans and Popova (2015) argues that divergent conclusions from systematic reviews tend to be driven by a reliance on different samples of research studies, which, in turn, are driven by differing criteria for inclusion. We are aware that our inclusion criteria has influenced our results and may have served the limit the utility of our findings. The way in which we conceptualised a 'change in decision-making to the level of the school' is also likely to have limited the depth of our analysis. It may specifically have been useful to include studies which evaluated interventions designed to improve the functioning of existing school-based decision-making mechanisms, as these may have contributed valuable evidence to the section on implementation factors. Such studies could usefully be examined in a subsequent review. Similarly, our specific concern with the impact of changes in decision-making at the level of the school means that we have excluded interventions organised by outside agencies (e.g. donor agencies, NGOs) external to the school, where there has been no active agency by local stakeholders. As there are indications that interventions designed by outside agencies are likely to be more successful, if less sustainable (Bold et al., 2013), the exclusion of studies considering such interventions may have impacted the results of our review.

Furthermore, the included studies represent only some of the contexts in which school-based management reforms have been implemented. Some countries which have implemented school-based decision-making reforms do not feature in the sample (e.g. Brazil, Guatemala), while other countries (e.g. Mexico and the Philippines) are over-represented. Given that context clearly plays a crucial role in the success of school-based decision-making reforms, the limited geographic diversity of the included studies limits the quality of our analysis.

In addition to limitations related to the review methodology, the evidence base itself carries limitations. In particular, the lack of studies comparing different ways in which it might be possible to shift decision-making from higher levels to the level of the school restricted our ability to compare the relative effectiveness of different approaches. Similarly, the lack of information in the studies about the cost of particular intervention types precluded us from discussing cost-effectiveness in this review.

5.6. AGREEMENTS AND DISAGREEMENTS WITH OTHER REVIEWS

Although there are no other systematic reviews on school-based management following the Campbell Collaboration criteria, there are two comprehensive literature reviews available on the topic (Santibanez, 2007; World Bank, 2007). Our findings are broadly similar to the conclusions reached by both reviews, in that both identified moderate impact on drop-out and repetition and mixed impact on student learning. The most significant difference that can be identified is the size and geographic breadth of the body of evidence reviewed. In 2007, the World Bank Education team was only able to identify 13 impact studies (all of which focused on Latin American initiatives). Santibanez identified slightly more studies (19 from low income contexts), but most of these (16) also focused on Latin America. Our review, in contrast, includes 26 impact studies, representing 13 countries in Latin America (5 countries), sub-Saharan Africa (5 countries) and South/Southeast Asia (3 countries).

6. Other Topics

6.1. DEVIATIONS FROM THE PUBLISHED PROTOCOL

The methods employed in this review deviated from the method outlined in the published protocol in a few respects:

1) During the search process, we refined our list of search terms. Although largely similar to the list in the published protocol, the final search strategy differed in a few minor respects. The full search strategy is available in Appendix 8.2.
2) Due to time constraints, we consulted a slightly abbreviated list of electronic databases and websites from the list published in the protocol. We are confident that our final list represents a broad range of disciplinary perspectives and is likely to have captured unpublished and 'grey' literature as well as formally published studies. The limited number of additional studies identified during citation chasing confirms that our initial search was comprehensive. Time pressures also prevented us from using the Web of Science, Google Scholar or Scopus to do any forward citation chasing; instead, we relied on reference following and expert checking to verify our final list of studies.
3) Once we began the full-text screening phase, we realised that we needed to add an additional exclusion criterion. As 'external' interventions (implemented by external bodies without any evident stakeholder involvement in the process), and interventions attempting to improve the functioning of existing devolved decision-making structures, cannot really be understood to constitute a change in decision-making authority, any studies investigating such interventions were excluded from synthesis.
4) Given the large number of impact studies that we found through our search, we elected to modify our inclusion criteria for Review Question 2, by limiting our analysis of non-causal studies to those pertaining to one of the interventions investigated through the impact studies included in the review.
5) During data extraction, we elected to modify the code lists in order to simplify their use. Although there is no difference in the substantive content, the order and formatting of the code lists in Appendix 8.4 differs slightly from those included in the published protocol.

6) As we could identify no consistent intervention-outcome pairs, it was not possible to complete separate narrative assessment for each pair (as specified on page 24 of our protocol). Instead, we elected to conduct in-depth narrative analysis of heterogeneity.
7) We were unable to complete any aggregate sub-group analysis, as the included studies rarely report separate estimates for a common set of sub-groups.
8) It was also not possible to formally test the impact of any identified enabling and constraining factors, given the heterogeneity of the final sample of studies and the limited number of studies with data pertaining to such factors. The diversity of findings also prevented us from assembling one aggregated 'Summary of Findings' table. Instead, we opted to create individual tables for each of the identified areas of heterogeneity within the study sample and to integrate the data sets through a revision of the initial conceptual framework.

6.2. PLANS FOR UPDATING THE REVIEW

The members of the review team will update the review if and when new rigorous evidence (and suitable funding) becomes available.

6.3. AUTHOR CONTRIBUTIONS

As Team Leader of the review, Roy Carr-Hill contributed to all aspects of the process. Specific contributions included appraisal and assessment of risk of bias of all included impact studies; advising on the methods used during meta-analysis; assistance with synthesis of both the impact and the non-causal studies; and drafting of sections of the review report.

Caine Rolleston was responsible for the meta-analysis, spearheading the calculation of standardised effect sizes and the creation of forest and funnel plots. He also wrote all sections of the report pertaining to the meta-analysis (both methodology and results) and contributed to the assessment of risk of bias of all included impact studies.

Tejendra Pherali contributed to the quality appraisal and synthesis of non-causal studies.

Rebecca Schendel directed the overall review process, while also contributing to each phase. Specific contributions included assisting with screening; spearheading quality appraisal of non-causal studies; conducting the heterogeneity analysis; integrating the data sets; and writing the final review report.

Edwina Peart and Emma Jones conducted the searches and completed the majority of the screening of studies. They also assisted with quality appraisal of the non-causal studies.

6.4. DECLARATIONS OF INTERESTS

None of the team members have any financial interests in the review, nor have any team members been involved in any other systematic review focused on this topic or in the development of any of the interventions investigated.

6.5. ACKNOWLEDGEMENTS

Claire Stansfield and Carol Vigurs at the EPPI Centre developed the search strategy and provided invaluable assistance during the search process. Charley Nussey and Amy Smail also assisted with searching, screening and full text retrieval during the early phases of the review. We have also benefited from the advice of Kelly Dickson in the EPPI Centre, as well as substantial support from Hugh Waddington at 3ie.

6.6. SOURCES OF SUPPORT

This review is supported by DFID, as part of their Systematic Reviews Programme.

7. References

7.1. REFERENCES TO INCLUDED STUDIES

Impact Studies (n=26)

Bando, R. (2010). *The Effect of School Based Management on Parent Behavior and the Quality of Education in Mexico*. Unpublished PhD thesis.

Barr, A., Bategeka, L., Guloba, M., Kasirye, I., Mugisha, F., Serneels, P. & Zeitlin, A. (2012). *Management and motivation in Ugandan primary schools: an impact evaluation report*. PEP Working Paper. Nairobi: Partnership for Economic Policy.

Beasley, E. & Huillery, E. (2014). *Willing but Unable: Short-Term Experimental Evidence on Parent Empowerment and School Quality*. Unpublished manuscript. Available at: http://www.povertyactionlab.org/publication/willing-unable-short-term-experimental-evidence-parent-empowerment-and-school-quality.

Blimpo, M. & Evans, D.K. (2011). *School-Based Management and Educational Outcomes: Lessons from a Randomized Field Experiment*. Unpublished manuscript. Available at: http://siteresources.worldbank.org/EDUCATION/Resources/Blimpo-Evans_WSD-2012-01- 12.pdf.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A. & Sandefur, J. (2013). *Scaling-up What Works:Experimental Evidence on External Validity in Kenyan Education*. CSAE Working Paper WPS/2013-04. Oxford: Centre for the Study of African Economies.

Di Gropello, E. & Marshall J.H. (2005). 'Teacher effort and schooling outcomes in rural Honduras.' In: E. Vegas (ed), *Incentives to improve teaching*. Washington DC: World Bank, pages 307-358.

Duflo, E., Dupas. P. & Kremer, M. (2012). *School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools*. NBER Working Paper No. 17939. Cambridge, MA: National Bureau of Economic Research.

Gertler, P., Patrinos, H.A. & Rubio-Codina, M. (2012). 'Empowering parents to improve education: Evidence from rural Mexico.' *Journal of Development Economics*, 99(1): 68-79.

Glewwe, P. & Maïga, E. (2011). *The Impacts of School Management Reforms in Madagascar: Do the Impacts Vary by Teacher Type?* Unpublished manuscript.

Jimenez, E. & Sawada, Y. (1999). 'Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program.' *The World Bank Economic Review*, 13(3): 415-441.

Jimenez, E. & Sawada, Y. (2003). *Does Community Management Help Keep Kids in*

Schools? Evidence Using Panel Data from El Salvador's EDUCO Program. CIRJE Discussion Paper F-236. Available at: http://www.cirje.e.u-tokyo.ac.jp/research/dp/2003/2003cf236.pdf.

Khattri, N., Ling. C. & Jha, S. (2010). *The Effects of School-based Management in the Philippines: An Initial Assessment Using Administrative Data*. World Bank Policy Research Working Paper 5248. Washington, DC: World Bank.

King, E.M. & Ozler, B. (2005). *What's Decentralization got to do with learning?* 21COE Discussion Paper No. 54. Kyoto: Kyoto University.

Lassibille, G., Tan, J.P., Jesse, C. & Van Nguyen, T. (2010). 'Managing for Results in Primary Education in Madagascar: Evaluating the Impact of Selected Workflow Interventions.' *World Bank Economic Review*, 24(2): 303-329.

Murnane, R.J., Willett, J.B. & Cardenas, S. (2006). *Did the Participation of Schools in Programa Escuelas de Calidad (PEC) Influence Student Outcomes?* Unpublished manuscript.

Parker, C.E. (2005). 'Teacher incentives and student achievement in Nicaraguan autonomous schools.' In: E. Vegas (ed), *Incentives to improve teaching*. Washington DC: World Bank, pages 359-388.

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Alishjabana, A., Gaduh, A. & Prama Artha, R. (2011). *Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia*. World Bank Policy Research Working Paper 5795. Washington, DC: World Bank.

Rodriguez, C., Sanchez, F. & Armenta, A. (2009). 'Do Interventions at School Level Improve Educational Outcomes? Evidence from a Rural Program in Colombia.' *World Development*, 38(3): 415-428.

San Antonio, D.M. (2008). 'Creating Better Schools through Democratic School Leadership.' *International Journal of Leadership in Education*, 11(1): 43-62.

Santibañez, L., Abreu-Lastra, R. & O'Donoghue, J. (2014). 'School based management effects: Resources or governance change? Evidence from Mexico.' *Economics of Education Review*, 39: 97-109.

Sawada, Y. & Ragatz, A.B. (2005). 'Decentralization of education, teacher effort, and educational outcomes.' In: E. Vegas (ed), *Incentives to improve teaching*. Washington DC: World Bank, pages 255-306.

Skoufias, E. & Shapiro, J. (2006). *Evaluating the impact of Mexico's quality schools program: the pitfalls of using non-experimental data*. Impact Evaluation Series No. 8. Washington, DC: World Bank. Available at: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2006/10/12/000016406_20061012150223/Rendered/PDF/wps4036.pdf.

World Bank. (2011). *An Impact evaluation of Sri Lanka's policies to improve the performance of schools and primary school students through its school improvement and school report card programs*. South Asia Human Development Unit Report No. 35. Washington, DC: World Bank.

World Bank. (2013). *Republic of the Philippines Basic Education Public Expenditure Review*

Phase 2 School Based Management in The Philippines: An Empirical Investigation. East Asia and Pacific Unit Report No: ACS2212. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/16076/ACS22120PER0P10ase0II0Final0Report.pdf?sequence=1

Yamauchi, F. (2014). *An Alternative Estimate of School-based Management Impacts on Students' Achievements: evidence from the Philippines*. World Bank Policy Research Working Paper 6747. Washington, DC: World Bank. Available at: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2014/01/16/000158349_20140116114632/Rendered/PDF/WPS6747.pdf.

Yamauchi, F. & Liu, Y. (2012). *Impacts of an early stage education intervention on students' learning achievement : evidence from the Philippines*. World Bank Policy Research Working Paper 6246; Impact Evaluation Series No. 72. Washington, DC: World Bank. Available at: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2012/10/18/000158349_20121018144416/Rendered/PDF/wps6246.pdf.

Other Studies (n=9)

Bandur, A. (2008). A study of the implementation of school-based management in Flores primary schools in Indonesia. Unpublished PhD thesis.

Bjork, C. (2003). 'Local Responses to Decentralization Policy in Indonesia.' *Comparative Education Review*, 47 (2): 184-216.

de Umanzor S., Soriano, I., Vega, M.R., Jimenez, E., Rawlings, L., & Steele, D. (1997). *El Salvador's EDUCO Program: A First Report on Parents' Participation in School-Based Management*. Working Paper Series on Impact of Education Reforms, Paper No. 4. Washington, DC: World Bank. Available at: http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079934475/547667-1135281552767/ElSalvador_EDUCO.pdf.

Fuller B. & Rivarola, M. (1998). *Nicaragua's Experiment to decentralize schools: views of parents, teachers and directors*. Working Paper Series on Impact of Education Reforms, Paper No. 5. Washington, DC: World Bank. Available at: http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079934475/547667-1135281552767/Nicaragua_Decentralize_Schools.pdf.

Gershberg, A.I. & Meade, B. (2005). 'Parental Contributions, School-Level Finances and Decentralization: An Analysis of Nicaraguan Autonomous School Budgets.' *Comparative Education*, 41 (3): 291-308.

Gunnarsson V., Orazem P.F., Sanchez M.A., & Verdisco, A. (2008). *Does Local School Control Raise Student Outcomes?: Theory and Evidence on the Roles of School Autonomy and Community Participation*. Working Paper No. 09012. Ames, IA: Iowa State University. Available at:

http://www.econ.iastate.edu/sites/default/files/publications/papers/p5504-1009-06-19.pdf.

Hanushek, E.A., Link, S., & Woessmann, L. (2011). *Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA*. NBER Working Paper No. 17591. Washington, DC: National Bureau of Economic Research. Available at: http://papers.nber.org/papers/w17591.

Reimers, F. & Cardenas, S. (2007). 'Who Benefits from School-Based Management in Mexico?' *Prospects: Quarterly Review of Comparative Education*, 37 (1): 37-56.

Vernez, G., Karam, R., & Marshall. J.H. (2012). *Implementation of School-Based Management in Indonesia*. Monograph. Santa Monica, CA: RAND Corporation. Available at: http://www.rand.org/pubs/monographs/MG1229.html.

7.2. REFERENCES TO STUDIES EXCLUDED IN FINAL STAGES

Study excluded for missing data (n=1)

Carnoy, M., Gove, A.K., Loeb, S., Marshall, J.H., and Socias, M. (2008) 'How Schools and Students Respond to School Improvement Programs: The Case of Brazil's PDE'. *Economics of Education Review* 27(1): 22-38

Studies excluded during quality appraisal (n=19)

(2013). *Interim Support to Education Programme (INSTEP) Project Completion Review*. London: DFID.

Abdinoor, A. (2008). 'Community Assumes the Role of State in Education in Stateless Somalia'. *International Education* 37(2): 43-61.

Akyeampong, K. (2011). *(Re)Assessing the Impact of School Capitation Grants on Educational Access in Ghana*. CREATE Pathways to Access Research Monograph No. 71. Brighton: University of Sussex.

Amirrachman, A., Syafi'i, S. and Welch, A. (2008). 'Decentralising Indonesian education: the promise and the price'. *World Studies in Education* 9(1): 31-53.

Chowdhury, M.D., Al-Mahmood, A., Bashar, M.A., and Ahmed, J.U. (2011). *Localization of Digital Content for Use in Secondary Schools of Bangladesh*.

Condy, A. (1998). *Improving the Quality of Teaching and Learning Through Community Participation: Achievements, Limitations and Risks: Early lessons from the Schooling Improvement Fund in Ghana*. Social Development Working Paper No. 2. London: DFID.

Cossou, M. (2000). *Recherche opérationnelle sur la coopération en éducation de base dans les pays francophones d'Afrique de l'Ouest : cas du Bénin* (Operational research on cooperation in basic education in the francophone countries of West Africa: the case of Benin). Montreal: Fondation Paul Gérin-Lajoie; Ottawa: International Development Research Centre.

Dowd, A. and Namathaka, L. (2007). 'Malawi, 1994–2003: Training on a National Scale'. In D. Glassman, J. Naidoo, and F. Woods (eds), *Community schools in africa: Reaching the unreached*. New York: Springer.

Ekosiswoyo, R., Evans, D.P., Thair, M., and Wello, M.B. (2007). *Final evaluation: Managing Basic Education (MBE) Project*. Washington, DC: The Mitchell Group.

Holger, D. (2007). *School decentralization in the context of globalizing governance: international comparison of grassroots responses*. Dordrecht: Springer.

Jones, A. (2005). 'Conflict, development and community participation in education: Pakistan and Yemen'. *Internationales Asienforum* 36(3-4): 289-310.

Pailwar, V.K., and Mahajan, V. (2005). 'Janshala in Jharkhand: An Experiment with Community Involvement in Education'. *International Education Journal* 6(3): 373-385.

Tate, S, and Amedie, W.Y. (2011). *Mid-term evaluation of the USAID community-school partnership program for education and health*.

Updadhaya, H., Dubey, N., and Shrestha, O. (2007). *Understanding School Autonomy: A Study on Enabling Conditions for School Effectiveness*. Kathmandu: Research Centre for Educational Innovation and Development.

Vasquez, W.F. (2012). 'Supply-Side Interventions and Student Learning in Guatemala'. *International Review of Education* 58(1): 9-33.

Wadesango, N. (2012). 'The influence of teacher participation in decision-making on their occupational morale.' Journal of social sciences 31(3): 361-369.

Wanzare, Z. (2012). 'Instructional Supervision in Public Secondary Schools in Kenya'. Educational Management Administration & Leadership 40(2): 188-216.

Yousuf, M.I., Alam, M.T., Sajjad, M.L, and Imran, M. (2010). 'Amelioration of Educational Conditions through School Management Committees.' Journal of College Teaching & Learning 7(9): 47-52.

Yuki, T., Mizuno, K., Ogawa, K., and Mihoko, S. (2013). 'Promoting gender parity in basic education: lessons from a technical cooperation project in Yemen'. International Review of Education 59(1): 47-66.

Studies excluded as not about an included intervention (n=45)

Abebe, W. (2012). *School Management and Decision-making in Ethiopian Government Schools: Evidence from the Young Lives Qualitative School Survey*. Young Lives Working Paper 86.

American Institutes for Research. (2006) *Haiti : education 2004 project -- impact evaluation study*.

Amjad, R. and MacLeod, G. (2014). 'Academic effectiveness of private, public and private–public partnership schools in Pakistan'. *International Journal of Educational Development* 37: 22-31.

Arcia, G. and Belli, H. (1999). *Rebuilding the Social Contract: School Autonomy in* Nicaragua. LCSHD Paper Series No. 40. Washington: World Bank.

Arvind, G.R. (2009). 'Local Democracy, Rural Community, and Participatory School Governance'. *Journal of Research in Rural Education* 24(2): 1-13.

Barnhardt, S., Karlan, D., and Khemani, S. (2005). *Participation In A School Incentive Program In Karnataka.*

Barrs, J. (2005). 'Factors Contributed by Community Organizations to the Motivation of Teachers in Rural Punjab, Pakistan, and Implications for the Quality of Teaching'. *International Journal of Educational Development* 25(3): 333-348.

Blöchliger, H., Égert, B. and Fredriksen, K. (2013). *Fiscal Federalism and its Impact on Economic Activity, Public Investment and the Performance of Educational Systems*. OECD Economics Department Working Papers No. 1051.

Brunette, T., Chimombo, J., Chiuye, G. and Tilson, T. (2011). *USAID/Malawi: Education Decentralization Support Activity (EDSA) Mid-Term Evaluation*.

Cueto, S., Torero. M., León. J and Deustua, J. (2008). *Asistencia docente y rendimiento escolar: el caso del programa META*. Documento de Trabajo 53. Lima: GRADE.

Dang, H. and King, E. (2013). *Incentives and Teacher Effort: Further Evidence from a* Developing Country.

Diko, N.N. (2006) 'Culture production of the educated person: a case study of a rural co-educational high school in the Eastern Cape'. *Agenda* 68: 88-94.

El-Baradei, L. and Amin, K. (2010). 'Community participation in education: a case study of the boards of trustees' experience in the Fayoum governorate in Egypt'. *Africa education review* 7(1): 107-138.

Eskeland, G.S. and Filmer, D. (2002). *Autonomy, Participation, and Learning in Argentine Schools: Findings and Their Implications for Decentralization*. Washington, DC: World Bank.

Evans, D., Purwadi, A., Setiadi, A., Losert, L., Wello, M., Bimo, N., Noni, N., Tate, S. and Amd, S.T. (2012). *Indonesia: Decentralized Basic Education Project Final Evaluation Volume I: Main Report*. Washington, DC: USAID.

Gamage, D.T. and Sooksomchitra, P. (2004). 'Decentralisation And School-Based Management In Thailand'. *International Review of Education* 50(3-4): 289-305.

Gao, X., Barkhuizen, G.P. and Chow, A.W.K. (2011). 'Research engagement and educational decentralisation: problematising primary school English teachers' research experiences in China'. *Educational Studies* 37(2): 207-219.

Garnier, M., Diallo, M., Diallo, M., Diallo, T., Koivogui, A., Leno, P., and Sako, M. (2005). *Community Participation, Quality and Equity in Guinea's Schools: Evaluation Report of the PACEEQ project: 2001-2005*.

Garnier, M. and Tigri, F. (2013). *Girls' education & community participation project (GECP): Final evaluation*.

Gershberg, A.I., Meade, B. and Andersson, S. (2009). 'Providing Better Education Services to the Poor: Accountability and Context in the Case of Guatemalan Decentralization'. *International Journal of Educational Development* 29(3): 187-200.

Gomez, J., Valdivia, L., and Lambert, V. (2001). *Evaluation of the Falconbridge Foundation School Sponsorship Program*.

Goyal, S. and Pandey, P. (2013). 'Contract teachers in India'. *Education Economics* 21(5): 464-484.

Hoadley, U., Christie, P., and Ward, C.L. (2009). 'Managing to Learn: Instructional

Leadership in South African Secondary Schools'. *School Leadership & Management* 29(4): 373-389.

Hyde, K., Kadzamira, E., Sichinga, J., Chibwana, M., and Ridker, R. (1997). *Village based schools in Mangochi, Malawi : an evaluation*.

Jones, N., Lyytikainen, M., Mukherjee, M., Gopinath, R.M. (2007). *Local institutions and social policy for children. Opportunities and constraints of participatory service delivery*. UNICEF/Young Lives Social Policy Paper 001.

King, E. and Cordeiro-Guerra, S. (2005). 'Education reforms in East Asia: Policy, process and impact'. In *East Asia decentralizes: Making local government work*. Washington, DC: World Bank.

Kingdon, G. and Muzammil, M. (2010). 'The school governance environment in Uttar Pradesh, India: implications for teacher accountability and effort'. *Journal of development studies* 49(2): 251-269.

Laugharn, P. (2007). *Negotiating "Education for Many": Enrolment, Drop-out and Persistence in the Community Schools of Kolondieba, Mali*. CREATE Pathways to Access Research Monograph No. 14.

Leclercq, F. (2002). *The Impact of Education Policy Reforms on the School System:A Field Study of EGS and Other Primary Schools in Madhya Pradesh*. CSH Occasional Paper. Paris: Universite de Paris; New Delhi: Centre de Sciences Humaines.

Marchelli, H.C. (2001). *Decentralization and Privatization of Education in El Salvador: Assessing the Experience*. National Center for the Study of Privatization in Education Occasional Paper. New York, NY: Columbia University.

Marshall, J. (2004). *EQIP School Grants Program Evaluation: Final Report*.

Mncube, V. and Harber, C. (2010). 'Chronicling Educator Practices and Experiences in the Context of Democratic Schooling and Quality Education in South Africa'. *International Journal of Educational Development* 30(6): 614-624.

Motala, S. (2009). 'Privatising Public Schooling in Post-Apartheid South Africa: Equity Considerations'. *Compare: A Journal of Comparative and International Education* 39(2): 185-202.

Paes de Barros, E. and Mendonca, R. (1998). 'The impact of three institutional innovations in Brazilian Education.' In Savedoff, W.D. (ed), *Organization matters: agency problems in health and education in Latin America*. Washington, DC: Inter-American Development Bank.

Poppema, M. (2009). 'Guatemala, the Peace Accords and Education: A Post-Conflict Struggle for Equal Opportunities, Cultural Recognition and Participation in Education'. *Globalisation, Societies and Education* 7(4): 383-408.

Pryor, J. (2005). 'Can Community Participation Mobilise Social Capital for Improvement of Rural Schooling? A Case Study from Ghana'. *Compare: A Journal of Comparative Education* 35(2): 193-203.

Rajbhandari, M.M.S. (2007). *Community Readiness for Self-Managed School*.

Rajeev, S. (2005). 'Identifying a Framework for Initiating, Sustaining and Managing Innovations in Schools'. *Psychology and Developing Societies* 17: 51-80.

Sahasewiyon, K. (2004). 'Working locally as a true professional: case studies in the development of local curriculum through action research in the context of Thai schools'. *Educational Action Research* 12(4): 493-514.

Sayed, Y. and Soudien, C. (2005). 'Decentralisation and the Construction of Inclusion Education Policy in South Africa'. *Compare: A Journal of Comparative Education* 35(2): 115-125.

Shenker, S.D. (2012). 'Towards a world in which many worlds fit?: Zapatista autonomous education as an alternative means of development'. *International journal of educational development* 32(3): 432-443.

Soudien, C. and Sayed, Y. (2004). 'A New Racial State? Exclusion and Inclusion in Education Policy and Practice in South Africa'. *Perspectives in Education* 22(4): 101-116.

Toi, A. (2010). 'An Empirical Study of the Effects of Decentralization in Indonesian Junior Secondary Education'. *Educational Research for Policy and Practice* 9(2): 107-125.

Transparency International. (2010). *Africa Education Watch: Good Governance Lessons for Primary Education*.

Ye, Wangbei. (2012). 'Beyond State Planned School-Based Curriculum Development: One Chinese School's Story'. *International Journal of Educational Reform* 21(4): 253-275.

7.3. EXISTING REVIEWS CONSULTED DURING INITIAL SEARCH

Bruns, B., Filmer, D., & Patrinos, H. A. (2012). *Making Schools Work: New Evidence on Accountability Reforms*. Washington, DC: World Bank. DOI: 10.1596/978-0-8213-8679-8.

Guerrero, G., Leon, J., Zapata, M., Sugimaru, C., & Cueto, S. (2012). *What works to improve teacher attendance in developing countries? A systematic review*. London: EPPICentre, Social Science Research Unit, Institute of Education, University of London. Available at: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3377.

Petrosino, A., Morgan, C., Fronius, T.A., Tanner-Smith, E.E., & Boruch, R.F. (2012). *Interventions in developing nations for improving primary and secondary school enrollment of children: A systematic review*. Campbell Systematic Reviews 2012: 19. Available at: http://www.campbellcollaboration.org/lib/project/123/.

Santibanez, L. (2007). *School-based management effects on educational outcomes: A literature review and assessment of the evidence base*. Toluca: Centro de Investigación y Docencia Económicas. Available at: http://www.libreriacide.com/librospdf/DTAP-188.pdf.

Westhorp, G., Walker, B. and Rogers, P. (2014). *Under what circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor? A realist synthesis*. London: EPPI-Centre, Social Science Research Unit, Institute of Education. Available at: http://r4d.dfid.gov.uk/pdf/outputs/SystematicReviews/Community-accountability-2014-Westhorp-report.pdf.

World Bank. (2007). *What Do We Know About School-Based Management?* Washington, DC: World Bank.

7.4. SUPPORTING LITERATURE

Andrabi, T., Das, J., & Khwaja, A. (2009). *Report cards: The impact of providing school and child test scores on educational markets*. Unpublished manuscript. Washington, DC: World Bank.

Anderson, L.M., Petticrew, M., Rehfuess, E., Armstrong, R., Ueffing, E., Baker, P., Francis, D., & Tugwell, P. (2011). Using logic models to capture complexity in systematic reviews. *Research Synthesis Methods* 2 (1), 33-42. DOI: 10.1002/jrsm.32.

Atherton, P. & Kingdon, G. (2010). *The relative effectiveness and costs of contract and regular teachers in India*. CSAE Working Paper Series 2010-15. Oxford: Centre for the Study of African Economies

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2008). *Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India*. *Policy Research Working Paper 4584*. Washington, DC: World Bank. DOI: 10.1596/1813-9450-4584.

Bardhan, P., & Mookherjee, D. (2000). Capture and governance at local and national levels. *American Economic Review* 90 (2), 135-139. DOI: 10.1257/aer.90.2.135.

Bardhan, P., & Mookherjee, D. (2005). Decentralizing antipoverty program delivery in developing countries. *Journal of Public Economics* 89 (4), 675-704. DOI: 10.1016/j.jpubeco.2003.01.001.

Barrera-Osorio, F., & Linden, L. (2009). *The use and misuse of computers in education: Evidence from a randomized experiment in Colombia*. *Policy ResearchWorking Paper (Impact Evaluation Series) 4836 (29)*. Washington, DC: World Bank.

Barrera-Osorio, F., Fasih, T., Patrinos, H.A., and Santibanez, L. (2009). *Decentralized Decision-Making in Schools: The Theory and Evidence on School-based Management*. Washington, DC: World Bank.

Becker, W.B.J., Hedges, L.V., & Pigott, T.D. (undated). *Statistical Analysis Policy Brief*. Oslo: The Campbell Collaboration. Available at: http://www.campbellcollaboration.org/artman2/uploads/1/C2_Statistical_Analysis_Policy_Brief-2.pdf.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons, Ltd.

Carr-Hill, R.A. (2012). Finding and then counting out-of-school children. *Compare: A journal of comparative and international education* 42 (2), 187-212. DOI: 10.1080/03057925.2012.652806.

Carr-Hill, R., Rolleston, C., Pherali, T., & Schendel, R. (2014). *The Effects of School-Based Decision Making on Educational Outcomes in Low and Middle Income Contexts: A Systematic Review (protocol)*. Available at: http://www.campbellcollaboration.org/lib/project/325/.

Carr-Hill, R., Hopkins, M., Lintott, J., & Riddell, A. (1999). *Monitoring the performance of*

educational programmes in developing countries. DFID Education Research Paper No. 37. London: DFID. Available at: http://r4d.dfid.gov.uk/PDF/Outputs/Misc_Education/paper37.pdf. Last accessed April 14, 2014.

Cochrane Effective Practice and Organisation of Care Group. (2014). *Suggested risk of bias criteria for EPOC reviews*. Ottawa: EPOC. Available at: http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Suggested%20risk%20of%20bias%20criteria%20for%20EPOC%20reviews.pdf. Last accessed April 10, 2014.

Condy, A. (1998). *Improving the quality of teaching and learning through community participation: Achievements, limitations and risks*. SDD/DFID Working Paper. London: DFID. Available at: http://www.eldis.org/vfile/upload/1/document/0708/DOC6157.pdf. Last accessed April 14, 2014.

De Grauwe, A., Lugaz, C., Balde, D., Diakhate, C., Dougnon, D., Moustapha, M., & Odushina, D. (2005). *Does decentralization lead to school improvement? Findings and lessons from research in West Africa*. Paris: IIEP. Available at: http://www.equip123.net/JEID/articles/1/1-1.pdf. Last accessed April 14, 2014.

DFID. (2014). *How To Note: Assessing the Strength of Evidence*. London: DFID.

Duflo, E., Dupas, P. & Kremer, M. (2011). *School governance, pupil-teacher-ratios, and teacher incentives experimental evidence from kenyan primary schools*. Unpublished Working Paper.

Gertler, P., Patrinos, H.A., & Rubio-Codina, M. (2008). *Impact evaluation for school-based management reform. Policy Research Working Paper. Doing Impact Evaluation Series 10*. Washington, DC: World Bank. Available at: http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_10.pdf. Last accessed April 14, 2014.

Glassman, D., Naidoo, J., & Wood, F. (2007). *Community schools in Africa: Reaching the unreached*. New York, NY: Springer.

Glewwe, P., Ilias, N., & Kremer, M. (2003). *Teacher Incentives. NBER Working Paper Series (Working Paper 9671)*. Cambridge, MA: National Bureau of Economic Research. Available at: http://www.nber.org/papers/w9671.pdf. Last accessed April 14, 2014.

Gough, D., Oliver, S., & Thomas, J. (2014). *An introduction to systematic reviews*. London: Sage Publications.

Hammerstrom, K., Wade, A., Hanz, K., & Jorgensen, A-M.K. (2009). *Searching for Studies: Information retrieval methods group policy brief*. Oslo: The Campbell Collaboration. Available at: http://www.campbellcollaboration.org/artman2/uploads/1/C2_Information_retrieval_policy_brief_new_draft.pdf.

Higgins, J.P.T, & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Available at: http://handbook.cochrane.org/

Hombrados, J.G., & Waddington, H. (2012). *Internal validity in social experiments and*

quasi-experiments: an assessment tool for reviewers. Mimeo. London: International Initiative for Impact Evaluation (3ie).

Keef, S.R., & Roberts, L.A. (2004). The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology* 57 (Part 1): 97-129. DOI: 10.1348/000711004849303.

Krishnaratne, S., White, H., & Carpenter, E. (2013). *Quality education for all children? What works in education in developing countries. Working Paper 20*. New Delhi: International Initiative for Impact Evaluation (3ie).

Lugaz, C., De Grauwe, A., Balde, D., Diakhate, C., Dougnon, D., Moustapha, M., & Odushina, D. (2010). *Schooling and decentralization: patterns and policy implications in Francophone West Africa*. Paris: IIEP.

Noyes, J, & Lewin, S. (2011). Supplemental guidance on selecting a method of qualitative evidence synthesis, and integrating qualitative evidence with Cochrane intervention reviews. In J. Noyes et al. (Eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions*.

Oliver, S., Dickson, K., & Newman, M. (2012). Getting started with a review. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (p. 66-82). London: Sage Publications.

Pherali, T., Smith, A., & Vaux, T. (2011). *A political economy analysis of education in Nepal*. Kathmandu: EU.

Rocha Menocal, A., & Sharma, B. (2008). *Joint evaluation of citizens' voice and accountability: Synthesis report*. London: DFID. Available at: http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3425.pdf. Last accessed April 14, 2014.

Rose, P. (2003). Community participation in school policy and practice in Malawi: Balancing local knowledge, national policies and international agency priorities. *Compare: A journal of comparative and international education* 33 (1), 47-64. DOI: 10.1080/03057920302597.

Shadish, W. & Myers, D. (2004). Research Design Policy Brief. Oslo: The Campbell Collaboration. Available at: http://www.campbellcollaboration.org/artman2/uploads/1/C2_Research_Design_Policy_Brief-2.pdf.

Snilstveit, B. (2012). Systematic reviews: from 'bare bones' reviews to policy relevance. *Journal of Development Effectiveness* 4 (3), 388-408. DOI: 10.1080/19439342.2012.709875.

Thomas, J., Harden, A., & Newman, M. (2012). Synthesis: Combining results systematically and appropriately. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (p. 179-226). London: Sage Publications.

UNESCO. (2014). *Teaching and learning: Achieving quality for all. EFA Global Monitoring Report 2013/4*. Paris: UNESCO. Available at: http://unesdoc.unesco.org/images/0022/002256/225660e.pdf. Last accessed April 14, 2014.

Unterhalter, E. (2012). Silences, stereotypes and local selection: Negotiating policy and practice to implement the MDGs and EFA. In A. Verger, H.K. Altinyelken, & M. Novelli (Eds.), *Global Education Policy and International Development: New Agendas, Issues and Policies* (p. 79-100). London: Contiuum.

Waddington, H., White, H., Snilstveit, B., Hombrados, J.G., Vojtkova, M., Davies, P., Bhavsar, A., Eyers, J., Koehlmoos, T.P., Petticrew, M., Valentine, J.C., & Tugwell, P. (2012). How to do a good systematic review of effects in international development: a tool kit. *Journal of Development Effectiveness* 4 (3), 359-387. DOI: 10.1080/19439342.2012.711765.

World Bank. (2004). *World Development Report: Making services work for poor people*. Washington, DC: World Bank.

World Bank. (2014). *Country and lending groups*. Washington, DC: World Bank.

8. Appendices

8.1. LIST OF SEARCH LOCATIONS

Education databases (electronic)

- AEI (Australian Education Index)
- BEI (British Education Index)
- ERIC (Education Resources Information Centre)

Multidisciplinary databases (electronic)

- ASSIA (Applied Social Science Index and Abstracts)
- IBSS (International Bibliography of the Social Sciences)

Other bibliographic databases and catalogues

- AJOL (African Journals Online)
- Asia Journals Online
- BLDS (British Library of Development Studies)
- CREATE (Consortium for Research on Educational Access, Transitions and Equity)
- IDEAS RePEc (Research Papers in Economics)
- IDRIS (International Development Research Centre Development Research Information System)
- IEA (International Association for the Evaluation of Educational Achievement)
- LAMJOL (Latin American Journals Online)
- National Bureau for Economic Research (NBER)
- SIGLE (Open Grey)
- UNBISNET (United Nations Bibliographic Information System)

Organisational databases or websites with potentially relevant publications lists

- 3ie RIDIE (Registry for International Development Impact Evaluations)
- Abdul Latif Jameel Poverty Action Lab (J-PAL)
- African Development Bank Evaluation Reports
- Asian Development Bank Evaluation Reports
- CEGA (Centre for Effective Global Action)
- DFID (Research for Development)
- DIME (Development Impact Evaluation Initiative)

- Inter-American Development Bank Evaluation Reports
- IE2 Impact Evaluation Repository (World Bank)
- IIEP (International Institute of Educational Planning)
- IPA (Yale University Innovations for Poverty Action Center)
- JOLIS (World Bank and IMF Library Catalogue)
- OECD (Organisation for Economic Co-Operation and Development ilibrary)
- SIDA (Swedish International Development Agency: Unit for Research Cooperation)
- UNESCdoc (United Nations Educational, Scientific and Cultural Organisation)
- USAID (Development Experience Clearinghouse)

8.2. DETAILED SEARCH STRATEGY

EBSCO Host Databases Search Strategy Outline:

Concepts based on change in decision making OR mechanisms of change AND developing countries AND date limit

- DE= Descriptors
- TX= All text
- TI=title
- AB=Abstract
- N2 within 2 words in any order

ERIC (search conducted 18th July 2014)

S1	TI (decentral* OR devolv* OR governance) n2 school OR AB (decentral* OR devolv* OR governance) n2 school	1504
S2	TI (decentral* OR devolv* OR governance) n2 education OR AB (decentral* OR devolv* OR governance) n2 education	1167
S3	TI ("school based management" OR SBM) OR AB ("school based management" OR SBM)	691
S4	TI ("shared decision making" OR SDM) OR AB ("shared decision making" OR SDM)	633
S5	Ti "school management committee*" OR AB "school management committee*"	13
S6	TI accountability n2 school OR AB accountability n2 school	1671
S7	TI accountability n2 education OR AB accountability n2 education	1096
S8	TI "report cards" OR AB "report cards"	768
S9	TI "principal leadership" OR AB "principal leadership"	428
S10	TI "School level planning" OR AB "School level planning"	15
S11	TI "school autonomy" OR AB "school autonomy"	179
S12	Ti "parent-teacher association" OR AB "parent-teacher association"	160
S13	Ti "community participation" n2 school OR AB "community participation" n2 school	101
S14	Ti "community participation" n2 education OR AB "community participation" n2 education	70
S15	TI "community based management" OR AB "community based management"	23

S16	TI (decentral* OR devolv* OR governance) n2 budget* OR AB (decentral* OR devolv* OR governance) n2 budget*	122
S17	Ti "resource allocation" n2 school OR AB"resource allocation" n2 school	107
S18	TI "resource allocation" n2 education OR AB "resource allocation" n2 education	66
S19	TI "capitation grant*" OR AB "capitation grant*"	13
S20	TI "block grant*" n2 school OR AB "block grant*" n2 school	14
S21	TI "block grant*" n2 education OR AB "block grant*" n2 education	68
S22	Ti (decentral* OR devolv* OR governance) n2 curriculum OR AB (decentral* OR devolv* OR governance) n2 curriculum	211
S23	TI (decentral* OR devolv* OR governance) n2 pedagog* OR AB (decentral* OR devolv* OR governance) n2 pedagog*	15
S24	TI "contract teachers" OR AB "contract teachers"	15
S25	TI "supply teachers" OR AB "supply teachers"	29
S26	Ti curriculum n2 local OR AB curriculum n2 local	564
S27	Ti pedagog* n2 local OR AB pedagog* n2 local	18
S28	TI "teacher allocation" OR AB "teacher allocation"	16
S29	TI "teacher distribution" OR AB "teacher distribution"	25
S30	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29	17398
S31	DE "School Administration"	5287
S32	DE "School Based Management"	1972
S33	DE "Teacher Leadership"	907
S34	DE "Instructional Leadership"	5982
S35	DE "School Restructuring"	4420
S36	DE "School Organization"	4163
S37	DE "School Statistics"	2395
S38	DE "Private School Aid"	656
S39	DE "School Support"	1663
S40	DE "School Funds"	1558
S41	DE "School District Autonomy"	1279
S42	DE "decentralization"	2332
S43	DE "report cards"	511
S44	DE "teacher distribution"	302
S45	S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44	29180

S46	S30 OR S45	42613
S47	TX Afghan* OR Libya# OR Albania# OR Macedonia# OR Algeria# OR Madagasca# OR Samoa# OR Malawi* OR Angola# OR Malaysia# OR Argent* OR Maldiv* OR Armenia# OR Mali OR Malian OR Azerbaij* OR "Marshall Islands" OR Bangladesh# OR Mauritania# OR Belarus* OR Mauriti* OR Belize OR Mexic* OR Benin OR Micronesia# OR Bhutan OR Moldov* OR Bolivia# OR Mongolia# OR Bosnia# OR Montenegr* OR Botswan* OR Morocc* OR Brazil* OR Mozambique OR Bulgaria# OR Myanmar OR Burkin* OR Namibia# OR Burundi* OR Nepal* OR "Cabo Verde" OR Nicaragua* OR Cambodia# OR Niger* OR Cameroon OR "African Republic" OR Pakistan* OR Chad OR Palau# OR China OR Chinese OR Panama# OR Colombia# OR "Papua New Guinea" OR Comoros OR Paraguay* OR Congo* OR Palestin# OR Peru* OR Philippin* OR "Costa Rica#" OR Romania# OR "Cote d'Ivoire" OR Rwanda# OR "Ivory coast" OR Cuba# OR Djibouti* OR "Sao Tome" OR Dominica# OR Senegal* OR Serbia# OR Ecuador* OR Seychelles OR Egypt OR Egyptian OR "Sierra Leone" OR "El Salvador" OR "Solomon Islands" OR Eritrea# OR Somalia# OR Ethiopia# OR "South Africa" OR Fiji# OR "South Sudan" OR Gabon* OR "Sri Lanka" OR Gambia# OR "St. Lucia" OR Georgia# OR "St. Vincent" OR Grenadines OR Ghana OR Ghanaian OR Sudan* OR Grenada* OR Surinam* OR Guatemala* OR Swaziland OR Guinea* OR Syrian OR Syria OR Palestin* OR "Guinea Bissau" OR Tajikistan OR Guyana# OR Tanzania# OR Haiti* OR Thailand OR Thai OR Hondura# OR "Timor Leste" OR Hungar* OR Togo OR Togolese OR India# OR Tonga# OR Indonesia# OR Tunisia# OR Iran OR Iranian OR Turkey OR Turkish OR Iraq# OR Turkmenistan OR Jamaica# OR Tuvalu OR Jordan* OR Uganda# OR Kazakhstan# OR Ukrain* OR Kenya# OR Uzbekistan OR Kiribati OR Vanuatu OR Korea# OR Venezuela# OR Kosov* OR Vietnam* OR Kyrgyz Republic OR "West Bank" OR Gaza OR Lao OR Laos OR Yemen# OR Lebanon OR Lebanese OR Zambia# OR Lesotho OR Zimbabw* OR Liberia#	109739
S48	TX (Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America") or ((developing OR "low income" OR "less developed" OR "lesser developed" OR "middle income" OR "under developed" OR "underdeveloped" OR "low and middle income" OR "lower income") N1 (countr* OR nation OR nations OR world)) or ((African OR Asian OR "South American" OR "Central American" OR "West Indian") N1 (nations OR countries OR economy OR economies)) or ((underserved OR "under served" OR deprived OR poor) N1 (countr* OR nation OR nations OR world)) OR ((LMIC OR LMICS OR "third world") N3 (countr* OR nation OR nations))	31783
S49	S47 OR S48	122690
S50	S46 AND S49	3552
S51	publication date from 2000	1644

ProQuest Database Search Strategy Outline:

Concepts based on change in decision making OR mechanisms of change AND developing countries AND date limit

- TI=title
- AB=Abstract
- SU = Subject (Index Terms)
- TX= All text
- Near/2 within 2 words in any order

ASSIA (search conducted 28th July 2014)

S1.	ti,ab((decentral* OR devolv* OR governance) near/2 school)	16
S2.	ti,ab((decentral* OR devolv* OR governance) near/2 education)	59
S3.	ti,ab(("school based management" OR SBM))	16
S4.	ti,ab(("shared decision making" OR SDM))	445
S5.	ti,ab("school management committee*")	2
S6.	ti,ab(accountability near/2 school)	44
S7.	ti,ab(accountability near/2 education)	33
S8.	ti,ab("report cards")	53
S9.	ti,ab("principal leadership")	5
S10.	ti,ab ("School level planning")	0
S11.	ti,ab("school autonomy")	5
S12.	ti,ab("parent-teacher association")	2
S13.	Ti, ab("community participation" near/2 school)	4
S14.	ti,ab("community participation" near/2 education)	7
S15.	ti,ab("community based management")	14
S16.	ti,ab((decentral* OR devolv* OR governance) near/2 budget*)	18
S17.	ti,ab("resource allocation" near/2 school)	1
S18.	ti,ab("resource allocation" near/2 education)	3
S19.	Ti,ab("capitation grant*")	0
S20.	ti,ab("block grant*" near/2 school)	1
S21.	ti,ab("block grant*" near/2 education)	1
S22.	ti,ab((decentral* OR devolv* OR governance) near/2 curriculum)	1
S23.	ti,ab((decentral* OR devolv* OR governance) near/2 pedagog*)	1
S24.	ti,ab("contract teachers")	3
S25.	ti,ab("supply teachers")	2
S26.	ti,ab(curriculum NEAR/2 local)	17
S27.	ti,ab(pedagog* near/2 local)	2
S28.	ti,ab("teacher allocation")	0
S29.	Ti,ab("teacher distribution")	3
S30.	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29	752
S31.	SU “Parent-Teacher Collaboration”	16
S32.	SU “School Governors”	6

S33.	S30 OR S31 OR S32	774
S34.	TX Afghan* OR Libya# OR Albania# OR Macedonia# OR Algeria# OR Madagasca# OR Samoa# OR Malawi* OR Angola# OR Malaysia# OR Argent* OR Maldiv* OR Armenia# OR Mali OR Malian OR Azerbaij* OR "Marshall Islands" OR Bangladesh# OR Mauritania# OR Belarus* OR Mauriti* OR Belize OR Mexic* OR Benin OR Micronesia# OR Bhutan OR Moldov* OR Bolivia# OR Mongolia# OR Bosnia# OR Montenegr* OR Botswan* OR Morocc* OR Brazil* OR Mozambique OR Bulgaria# OR Myanmar OR Burkin* OR Namibia# OR Burundi* OR Nepal* OR "Cabo Verde" OR Nicaragua* OR Cambodia# OR Niger* OR Cameroon OR "African Republic" OR Pakistan* OR Chad OR Palau# OR China OR Chinese OR Panama# OR Colombia# OR "Papua New Guinea" OR Comoros OR Paraguay* OR Congo* OR Palestin# OR Peru* OR Philippin* OR "Costa Rica#" OR Romania# OR "Cote d'Ivoire" OR Rwanda# OR "Ivory coast" OR Cuba# OR Djibouti* OR "Sao Tome" OR Dominica# OR Senegal* OR Serbia# OR Ecuador* OR Seychelles OR Egypt OR Egyptian OR "Sierra Leone" OR "El Salvador" OR "Solomon Islands" OR Eritrea# OR Somalia# OR Ethiopia# OR "South Africa" OR Fiji# OR "South Sudan" OR Gabon* OR "Sri Lanka" OR Gambia# OR "St. Lucia" OR Georgia# OR "St. Vincent" OR Grenadines OR Ghana OR Ghanaian OR Sudan* OR Grenada* OR Surinam* OR Guatemala* OR Swaziland OR Guinea* OR Syrian OR Syria OR Palestin* OR "Guinea Bissau" OR Tajikistan OR Guyana# OR Tanzania# OR Haiti* OR Thailand OR Thai OR Hondura# OR "Timor Leste" OR Hungar* OR Togo OR Togolese OR India# OR Tonga# OR Indonesia# OR Tunisia# OR Iran OR Iranian OR Turkey OR Turkish OR Iraq# OR Turkmenistan OR Jamaica# OR Tuvalu OR Jordan* OR Uganda# OR Kazakhstan# OR Ukrain* OR Kenya# OR Uzbekistan OR Kiribati OR Vanuatu OR Korea# OR Venezuela# OR Kosov* OR Vietnam* OR Kyrgyz Republic OR "West Bank" OR Gaza OR Lao OR Laos OR Yemen# OR Lebanon OR Lebanese OR Zambia# OR Lesotho OR Zimbabw* OR Liberia OR (Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America") or ((developing OR "low income" OR "less developed" OR "lesser developed" OR "middle income" OR "under developed" OR "underdeveloped" OR "low and middle income" OR "lower income") N1 (countr* OR nation OR nations OR world)) or ((African OR Asian OR "South American" OR "Central American" OR "West Indian") N1 (nations OR countries OR economy OR economies)) or ((underserved OR "under served" OR deprived OR poor) N1 (countr* OR nation OR nations OR world)) OR ((LMIC OR LMICS OR "third world") N3 (countr* OR nation OR nations))	47336
S35.	S33 (Limited by Publication Date Post 1st January 2000)	634
S36.	S34 AND S35	**55**

BEI (search conducted 29th July 2014)

S1.	ti,ab((decentral* OR devolv* OR governance) near/2 school)	16
S2.	ti,ab((decentral* OR devolv* OR governance) near/2 education)	137
S3.	ti,ab(("school based management" OR SBM))	34
S4.	ti,ab(("shared decision making" OR SDM))	5
S5.	ti,ab("school management committee*")	2

S6.	ti,ab(accountability near/2 school)	54
S7.	ti,ab(accountability near/2 education)	63
S8.	ti,ab("report cards")	3
S9.	ti,ab("principal leadership")	13
S10.	ti,ab ("School level planning")	0
S11.	ti,ab("school autonomy")	33
S12.	ti,ab("parent-teacher association")	1
S13.	Ti, ab("community participation" near/2 school)	4
S14.	ti,ab("community participation" near/2 education)	5
S15.	ti,ab("community based management")	1
S16.	ti,ab((decentral* OR devolv* OR governance) near/2 budget*)	1
S17.	ti,ab("resource allocation" near/2 school)	4
S18.	ti,ab("resource allocation" near/2 education)	2
S19.	Ti,ab("capitation grant*")	0
S20.	ti,ab("block grant*" near/2 school)	0
S21.	ti,ab("block grant*" near/2 education)	2
S22.	ti,ab((decentral* OR devolv* OR governance) near/2 curriculum)	8
S23.	ti,ab((decentral* OR devolv* OR governance) near/2 pedagog*)	2
S24.	ti,ab("contract teachers")	3
S25.	ti,ab("supply teachers")	17
S26.	ti,ab(curriculum NEAR/2 local)	29
S27.	ti,ab(pedagog* near/2 local)	1

S28.	ti,ab("teacher allocation")	0
S29.	Ti,ab("teacher distribution")	0
S30.	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29	470
S31.	SU “Institutional Autonomy”	220
S32.	SU “Professional Autonomy”	320
S33.	SU “School Governors”	423
S34.	SU “School Governing Bodies”	286
S35.	SU “Local Management of Schools”	649
S36.	SU “School based”	617
S37.	SU “Community control”	30
S38.	SU “School councils”	223
S39.	SU “Participative Decision Making”	264
S40.	S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39	3502
S41.	TX Afghan* OR Libya# OR Albania# OR Macedonia# OR Algeria# OR Madagasca# OR Samoa# OR Malawi* OR Angola# OR Malaysia# OR Argent* OR Maldiv* OR Armenia# OR Mali OR Malian OR Azerbaij* OR "Marshall Islands" OR Bangladesh# OR Mauritania# OR Belarus* OR Mauriti* OR Belize OR Mexic* OR Benin OR Micronesia# OR Bhutan OR Moldov* OR Bolivia# OR Mongolia# OR Bosnia# OR Montenegr* OR Botswan* OR Morocc* OR Brazil* OR Mozambique OR Bulgaria# OR Myanmar OR Burkin* OR Namibia# OR Burundi* OR Nepal* OR "Cabo Verde" OR Nicaragua* OR Cambodia# OR Niger* OR Cameroon OR "African Republic" OR Pakistan* OR Chad OR Palau# OR China OR Chinese OR Panama# OR Colombia# OR "Papua New Guinea" OR Comoros OR Paraguay* OR Congo* OR Palestin# OR Peru* OR Philippin* OR "Costa Rica#" OR Romania# OR "Cote d'Ivoire" OR Rwanda# OR "Ivory coast" OR Cuba# OR Djibouti* OR "Sao Tome" OR Dominica# OR Senegal* OR Serbia# OR Ecuador* OR Seychelles OR Egypt OR Egyptian OR "Sierra Leone" OR "El Salvador" OR "Solomon Islands" OR Eritrea# OR Somalia# OR Ethiopia# OR "South Africa" OR Fiji# OR "South Sudan" OR Gabon* OR "Sri Lanka" OR Gambia# OR "St. Lucia" OR Georgia# OR "St. Vincent" OR Grenadines OR Ghana OR Ghanaian OR Sudan* OR Grenada* OR Surinam* OR Guatemala* OR Swaziland OR Guinea* OR Syrian OR Syria OR Palestin* OR "Guinea Bissau" OR Tajikistan OR Guyana# OR	9535

	Tanzania# OR Haiti* OR Thailand OR Thai OR Hondura# OR "Timor Leste" OR Hungar* OR Togo OR Togolese OR India# OR Tonga# OR Indonesia# OR Tunisia# OR Iran OR Iranian OR Turkey OR Turkish OR Iraq# OR Turkmenistan OR Jamaica# OR Tuvalu OR Jordan* OR Uganda# OR Kazakhstan# OR Ukrain* OR Kenya# OR Uzbekistan OR Kiribati OR Vanuatu OR Korea# OR Venezuela# OR Kosov* OR Vietnam* OR Kyrgyz Republic OR "West Bank" OR Gaza OR Lao OR Laos OR Yemen# OR Lebanon OR Lebanese OR Zambia# OR Lesotho OR Zimbabw* OR Liberia OR (Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America") or ((developing OR "low income" OR "less developed" OR "lesser developed" OR "middle income" OR "under developed" OR "underdeveloped" OR "low and middle income" OR "lower income") N1 (countr* OR nation OR nations OR world)) or ((African OR Asian OR "South American" OR "Central American" OR "West Indian") N1 (nations OR countries OR economy OR economies)) or ((underserved OR "under served" OR deprived OR poor) N1 (countr* OR nation OR nations OR world)) OR ((LMIC OR LMICS OR "third world") N3 (countr* OR nation OR nations))	
S42.	S40 Limited by Publication Date Post 1st January 2000	344
S43.	S42 AND S43	**137**

AEI (search conducted 29th July 2014)

S1.	ti,ab((decentral* OR devolv* OR governance) near/2 school)	246
S2.	ti,ab((decentral* OR devolv* OR governance) near/2 education)	178
S3.	ti,ab(("school based management" OR SBM))	161
S4.	ti,ab(("shared decision making" OR SDM))	37
S5.	ti,ab("school management committee*")	3
S6.	ti,ab(accountability near/2 school)	181
S7.	ti,ab(accountability near/2 education)	107
S8.	ti,ab("report cards")	18
S9.	ti,ab("principal leadership")	65
S10.	ti,ab ("School level planning")	0
S11.	ti,ab("school autonomy")	29
S12.	ti,ab("parent-teacher association")	1
S13.	Ti, ab("community participation" near/2 school)	35
S14.	ti,ab("community participation" near/2 education)	19
S15.	ti,ab("community based management")	1
S16.	ti,ab((decentral* OR devolv* OR governance) near/2 budget*)	6
S17.	ti,ab("resource allocation" near/2 school)	33
S18.	ti,ab("resource allocation" near/2 education)	18

S19.	Ti,ab("capitation grant*")	0
S20.	ti,ab("block grant*" near/2 school)	1
S21.	ti,ab("block grant*" near/2 education)	0
S22.	ti,ab((decentral* OR devolv* OR governance) near/2 curriculum)	45
S23.	ti,ab((decentral* OR devolv* OR governance) near/2 pedagog*)	4
S24.	ti,ab("contract teachers")	8
S25.	ti,ab("supply teachers")	4
S26.	ti,ab(curriculum NEAR/2 local)	75
S27.	ti,ab(pedagog* near/2 local)	12
S28.	ti,ab("teacher allocation")	0
S29.	Ti,ab("teacher distribution")	0
S30.	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29	1174
S31.	SU “Institutional Autonomy”	256
S32.	SU “School-Government Relationship	1036
S33.	SU “Professional Autonomy”	140
S34.	SU “School Restructuring”	528
S35.	S30 OR S31 OR S32 OR S33 OR S34	1053
S36.	TX Afghan* OR Libya# OR Albania# OR Macedonia# OR Algeria# OR Madagasca# OR Samoa# OR Malawi* OR Angola# OR Malaysia# OR Argent* OR Maldiv* OR Armenia# OR Mali OR Malian OR Azerbaij* OR "Marshall Islands" OR Bangladesh# OR Mauritania# OR Belarus* OR Mauriti* OR Belize OR Mexic* OR Benin OR Micronesia# OR Bhutan OR Moldov* OR Bolivia# OR Mongolia# OR Bosnia# OR Montenegr* OR Botswan* OR Morocc* OR Brazil* OR Mozambique OR Bulgaria# OR Myanmar OR Burkin* OR Namibia# OR Burundi* OR Nepal* OR "Cabo Verde" OR Nicaragua* OR Cambodia# OR Niger* OR Cameroon OR "African Republic" OR Pakistan* OR Chad OR Palau# OR China OR Chinese OR Panama# OR Colombia# OR "Papua New Guinea" OR Comoros OR Paraguay* OR Congo* OR Palestin# OR Peru* OR Philippin* OR "Costa Rica#" OR Romania# OR "Cote d'Ivoire" OR Rwanda# OR "Ivory coast" OR Cuba# OR Djibouti* OR "Sao Tome" OR Dominica# OR Senegal* OR Serbia# OR Ecuador* OR Seychelles OR Egypt OR Egyptian OR "Sierra Leone" OR "El Salvador" OR "Solomon Islands" OR Eritrea# OR Somalia# OR Ethiopia# OR "South Africa" OR Fiji# OR "South Sudan" OR Gabon* OR "Sri Lanka" OR Gambia# OR "St. Lucia" OR Georgia# OR "St. Vincent" OR Grenadines OR Ghana OR Ghanaian OR Sudan* OR Grenada* OR Surinam* OR Guatemala* OR Swaziland OR Guinea* OR Syrian OR Syria OR Palestin* OR "Guinea Bissau" OR Tajikistan OR Guyana# OR Tanzania# OR Haiti* OR Thailand OR Thai OR Hondura# OR "Timor Leste" OR Hungar* OR Togo OR Togolese OR India# OR Tonga# OR Indonesia# OR Tunisia# OR Iran OR Iranian OR Turkey OR Turkish OR Iraq# OR Turkmenistan OR Jamaica# OR Tuvalu OR	8688

	Jordan* OR Uganda# OR Kazakhstan# OR Ukrain* OR Kenya# OR Uzbekistan OR Kiribati OR Vanuatu OR Korea# OR Venezuela# OR Kosov* OR Vietnam* OR Kyrgyz Republic OR "West Bank" OR Gaza OR Lao OR Laos OR Yemen# OR Lebanon OR Lebanese OR Zambia# OR Lesotho OR Zimbabw* OR Liberia OR (Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America") or ((developing OR "low income" OR "less developed" OR "lesser developed" OR "middle income" OR "under developed" OR "underdeveloped" OR "low and middle income" OR "lower income") N1 (countr* OR nation OR nations OR world)) or ((African OR Asian OR "South American" OR "Central American" OR "West Indian") N1 (nations OR countries OR economy OR economies)) or ((underserved OR "under served" OR deprived OR poor) N1 (countr* OR nation OR nations OR world)) OR ((LMIC OR LMICS OR "third world") N3 (countr* OR nation OR nations))	
S37.	S35 Limited by Date 1st January 2000	677
S38.	S36 AND S37	**131**

IBSS (search conducted 29th July 2014)

S1.	ti,ab((decentral* OR devolv* OR governance) near/2 school)	109
S2.	ti,ab((decentral* OR devolv* OR governance) near/2 education)	254
S3.	ti,ab(("school based management" OR SBM))	51
S4.	ti,ab(("shared decision making" OR SDM))	89
S5.	ti,ab("school management committee*")	5
S6.	ti,ab(accountability near/2 school)	71
S7.	ti,ab(accountability near/2 education)	55
S8.	ti,ab("report cards")	47
S9.	ti,ab("principal leadership")	6
S10.	ti,ab ("School level planning")	104
S11.	ti,ab("school autonomy")	29
S12.	ti,ab("parent-teacher association")	3
S13.	Ti, ab("community participation" near/2 school)	6
S14.	ti,ab("community participation" near/2 education)	9
S15.	ti,ab("community based management")	48
S16.	ti,ab((decentral* OR devolv* OR governance) near/2 budget*)	62
S17.	ti,ab("resource allocation" near/2 school)	3

S18.	ti,ab("resource allocation" near/2 education)	6
S19.	Ti,ab("capitation grant*")	2
S20.	ti,ab("block grant*" near/2 school)	1
S21.	ti,ab("block grant*" near/2 education)	3
S22.	ti,ab((decentral* OR devolv* OR governance) near/2 curriculum)	4
S23.	ti,ab((decentral* OR devolv* OR governance) near/2 pedagog*)	4
S24.	ti,ab("contract teachers")	6
S25.	ti,ab("supply teachers")	5
S26.	ti,ab(curriculum NEAR/2 local)	18
S27.	ti,ab(pedagog* near/2 local)	7
S28.	ti,ab("teacher allocation")	39
S29.	Ti,ab("teacher distribution")	0
S30.	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29	1071
S31.	SU “Educational Reform”	664
S32.	S30 OR S31	1714
S33.	TX Afghan* OR Libya# OR Albania# OR Macedonia# OR Algeria# OR Madagasca# OR Samoa# OR Malawi* OR Angola# OR Malaysia# OR Argent* OR Maldiv* OR Armenia# OR Mali OR Malian OR Azerbaij* OR "Marshall Islands" OR Bangladesh# OR Mauritania# OR Belarus* OR Mauriti* OR Belize OR Mexic* OR Benin OR Micronesia# OR Bhutan OR Moldov* OR Bolivia# OR Mongolia# OR Bosnia# OR Montenegr* OR Botswan* OR Morocc* OR Brazil* OR Mozambique OR Bulgaria# OR Myanmar OR Burkin* OR Namibia# OR Burundi* OR Nepal* OR "Cabo Verde" OR Nicaragua* OR Cambodia# OR Niger* OR Cameroon OR "African Republic" OR Pakistan* OR Chad OR Palau# OR China OR Chinese OR Panama# OR Colombia# OR "Papua New Guinea" OR Comoros OR Paraguay* OR Congo* OR Palestin# OR Peru* OR Philippin* OR "Costa Rica#" OR Romania# OR "Cote d'Ivoire" OR Rwanda# OR "Ivory coast" OR Cuba# OR Djibouti* OR "Sao Tome" OR Dominica# OR Senegal* OR Serbia# OR Ecuador* OR Seychelles OR Egypt OR Egyptian OR "Sierra Leone" OR "El	413735

	Salvador" OR "Solomon Islands" OR Eritrea# OR Somalia# OR Ethiopia# OR "South Africa" OR Fiji# OR "South Sudan" OR Gabon* OR "Sri Lanka" OR Gambia# OR "St. Lucia" OR Georgia# OR "St. Vincent" OR Grenadines OR Ghana OR Ghanaian OR Sudan* OR Grenada* OR Surinam* OR Guatemala* OR Swaziland OR Guinea* OR Syrian OR Syria OR Palestin* OR "Guinea Bissau" OR Tajikistan OR Guyana# OR Tanzania# OR Haiti* OR Thailand OR Thai OR Hondura# OR "Timor Leste" OR Hungar* OR Togo OR Togolese OR India# OR Tonga# OR Indonesia# OR Tunisia# OR Iran OR Iranian OR Turkey OR Turkish OR Iraq# OR Turkmenistan OR Jamaica# OR Tuvalu OR Jordan* OR Uganda# OR Kazakhstan# OR Ukrain* OR Kenya# OR Uzbekistan OR Kiribati OR Vanuatu OR Korea# OR Venezuela# OR Kosov* OR Vietnam* OR Kyrgyz Republic OR "West Bank" OR Gaza OR Lao OR Laos OR Yemen# OR Lebanon OR Lebanese OR Zambia# OR Lesotho OR Zimbabw* OR Liberia OR (Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America") or ((developing OR "low income" OR "less developed" OR "lesser developed" OR "middle income" OR "under developed" OR "underdeveloped" OR "low and middle income" OR "lower income") N1 (countr* OR nation OR nations OR world)) or ((African OR Asian OR "South American" OR "Central American" OR "West Indian") N1 (nations OR countries OR economy OR economies)) or ((underserved OR "under served" OR deprived OR poor) N1 (countr* OR nation OR nations OR world)) OR ((LMIC OR LMICS OR "third world") N3 (countr* OR nation OR nations))41	
S34.	S32 AND S33	361
S35.	S34 Limited by Date after 1st January 2000	**322**

Search Terms for Website Searches

English	**French**	**Spanish**	**Portuguese**
parent-teacher association	Association des parents d'élèves;	Asamblea de Padres, Consejo de Padres de Familia, Consejo de Participación social	associação entre pais e professores
School-based management	gestion par l'école; gestion autonome des écoles; décideurs au niveau des établissements scolaires; conseils de gestion des établissements scolaires	gestión escolar autonóma / autonómica, organización escolar autónoma	administração baseada na escola
community-based management (note - search with education/school)	gestion communitaire AND école/éducation	Organización escolar comunitaria AND escuela/educación	administração baseada na comunidade AND escola/educação

English	French	Spanish	Portuguese
community participation (note - search with education/school)	participation communitaire AND école/éducation	Participación comunitaria, participación de la comunidad AND escuela/educación	participação da comunidade AND escola/educação
school boards	commission scolaire	Comité escolar; consejo escolar	conselho escolar
school management committee	Comité de gestion scolaire, conseil de l'école	Consejos de administración escolar, consejos de gestión escolar	administração baseada em comissões escolares
school autonomy	autonomie scolaire; autonomie de l'école; Autonomie des établissements scolaires	Autonomía escolar, autonomía de la escuela	autonomia escolar
school governance	gouvernance scolaire	Gobierno del centro escolar, gobernanza del centro escolar, administración escolar, dirección escolar	governância escolar, governança escolar
decentralisation	décentralisation, déconcentrée	Descentralización educativa (some countries use Federalización too), municipalización	descentralização, municipalização
decentralised	décentralisée	Descentralizado/a, (federalizado/a)	descentralizado
decentralization	N/A	N/A	N/A
decentralized	N/A	N/A	N/A
devolution	dévolution	Not used	Not used
devolved management	la dévolution des pouvoirs de décision aux écoles	Not used	Not used
decentralised decision-making	la décentralisation des pouvoirs de décision aux écoles	(the translation would be toma descentralizada de decisiones, but I've never seen it used as a term in itself, the decision is part of the general management: gestión like Gestión escolar descentralizada, Administración escolar descentralizada)	processo de decisão descentralizado
decentralized decision-making	N/A	N/A	N/A

English	French	Spanish	Portuguese
school report cards	N/A (no term seperate from what given to students to reflect their marks)	N/A (no term seperate from what given to students to reflect their marks)	N/A (no term seperate from what given to students to reflect their marks)
capitation grants	subventions proportionnelles	N/A (only terms available are for students, not schools)	N/A (only terms available are for students, not schools)
supply teachers	enseignants intérimaires; enseignants suppléants	profesores / docentes substitutos/sustitutos	professores temporários
contract teachers	enseignants contractuels	profesores / docentes asalariados	professores contratados
curriculum reform	réforme du curriculum	reforma(s) curricular(es)	reforma curricular
curriculum relevance	pertinence du curriculum	relevancia curricular, pertinencia curricular	relevância/aplicabilidade curricular
accountability (note - only search with education/school)	responsabilisation AND école/éducation	responsabilidad AND escuela/educación	responsabilidade AND escola/educação

8.3. CONTACTED AUTHORS

Author's Surname	Institution
Abrereu-Lastra	Fundacion Idea (Mexico)
Barr	Georgetown University
Beasley	Institut d'études politiques de Paris
Beatty	Center for Global Development
Evans	World Bank
Di Gropello	World Bank
Duflo	Massachusetts Institute of Technology
Gertler	World Bank
Glewwe	University of Minnesota
Jesse	World Bank
Jha	New York University
Jimenez	World Bank
Kremer	Harvard University
Lassibile	Universite de Bourgogne
Ling	World Bank
Murnane	Harvard Graduate School of Education
Ng'ang'a	Unknown
O'Donohue	Unknown
Parker	Education Development Center
Patrinos	World Bank
Pradhan	University of Amsterdam

Author's Surname	Institution
Ragatz	World Bank
Sanchez	Universidad de los Andes (Colombia)
Santibanez	RAND
Sawada	World Bank; University of Tokyo
Shapiro	London School of Economics
Skoufias	World Bank
Suryadarma	Australian National University
Tan	World Bank
Van Nguyen	World Bank
Yamauchi	World Bank

8.4. CODE LISTS

8.4.1. Exclusion criteria for title and abstract screening

1. Exclude Duplicate
    a. Any title which matches another title in your allocation exactly (e.g. same date, author and title)
2. Exclude Language
    a. Studies available only in a language other than English, French, Spanish or Portuguese
3. Exclude Publication Status
    a. Sources that report second-hand on empirical findings, such as committee minutes, newspaper articles and the like
        i. Sources that are likely to include first-hand reporting of empirical findings (either published literature – such as journal articles, books, conference papers and institutional grey literature, including reports and process evaluations - or unpublished - such as dissertations and theses, empirical studies showing null and/or negative results and the like) should be included
4. Exclude Geographic context
    a. Studies without any data from any LMIC (as classified at the time of the intervention), excluding those in Europe & former USSR
        i. *Please refer to World Bank Historical Classification Table*
5. Exclude Level of Education
    a. Studies that do not include any data on primary or secondary education
6. Exclude No SBDM
    a. Studies in which *no change in the level of decision-making is apparent*, OR
    b. Studies that investigate a change in decision-making to a level higher than the school/community (e.g. from central to district government), OR
    c. Studies that investigate a change in decision-making to the individual or family level (e.g. individual voucher programmes)
7. Exclude Date Data Collection
    a. Studies in which *all data* were collected prior to 1990

8.4.2. Exclusion criteria for full text screening

1. Exclude Duplicate
    a. Any title which matches another title in your allocation exactly (e.g. same date, author and title)
2. Exclude Language
    a. Studies available only in a language other than English, French, Spanish or Portuguese
3. Exclude Publication Status
    a. Sources that report second-hand on empirical findings, such as committee minutes, newspaper articles and the like
        i. Sources that are likely to include first-hand reporting of empirical findings (either published literature – such as journal articles, books, conference papers and institutional grey literature, including reports and process evaluations - or unpublished - such as dissertations and theses, empirical studies showing null and/or negative results and the like) should be included
4. Exclude Geographic context
    a. Studies without any data from any LMIC (as classified at the time of the intervention), excluding those in Europe & former USSR
5. Exclude Level of Education
    a. Studies that do not include any data on primary or secondary education
6. Exclude No SBDM
    a. Studies in which *no change in the level of decision-making is apparent*, OR
    b. Studies that investigate a change in decision-making to a level higher than the school/community (e.g. from central to district government), OR
    c. Studies that investigate a change in decision-making to the individual or family level (e.g. individual voucher programmes)
7. Exclude Date Data Collection
    a. Studies in which *all data* were collected prior to 1990
8. Exclude Theoretical
    a. Studies which include no empirical data
        i. Note: Data can be collected in any manner – e.g. through quantitative research, qualitative research, document analysis, etc. – but the study must report at least some empirical findings and present an empirical methodology in order to be included
9. Exclude No Outcomes
    a. Studies which do not include any data on educational outcomes (neither proximal nor final)

8.4.3. Initial coding list

1. Single or multiple study
    a. If title is a summary of other studies and must be disaggregated for coding, CODE AS Summary Title
        i. *Note: If a study is coded as a summary title, no further coding is necessary at this stage*
    b. If not, continue to next coding set
2. Country context

a. Exclude context: Any study without any data from any LMIC, excluding those in Europe & former USSR
    i. *Note: If a country has been classified as a low or middle income country at some stage since 1995, the study should be retained for further coding*
    ii. *Note: Studies analysing data from more than one country can be included at this stage, even if they also reference high income contexts) – but exclude multi-country studies which reference only one LMIC*
    iii. *Note: If a study should be excluded on context, no further coding is necessary*
b. If data have been collected from *more than one* LMIC, CODE AS Multiple Country
c. If data have been collected from one LMIC, CODE AS name of individual country

3. Study design
    a. Exclude Not Empirical: Any study in which there is no identifiable method
        i. *Note: If a study should be excluded as not empirical, no further coding is necessary*
    b. Otherwise, CODE AS specific method
        i. RCT: Experimental designs using randomised or quasi-randomised assignment to the reform/intervention
        ii. Regression discontinuity design: Studies in which assignment to treatment/intervention group is based on known allocation rules including a cut-off rule on a continuous or ordinal policy variable
        iii. Natural experiment: Studies in which assignment to treatment/intervention group is due to a natural experiment (e.g. exogenous geographical/political variation)
        iv. Other quasi-experimental: Studies with a quasi-experimental design in which assignment to treatment/intervention group is based on other selection mechanisms (e.g. self-selection by participating schools)
        v. Longitudinal before-and-after: Before-and-after studies which collect longitudinal data at baseline and endline
        vi. Cross-sectional before-and-after with comparison group: Before-and-after studies which collect cross-sectional endline data from a treatment and a comparison group
        vii. Propensity score matching: Studies which collect cross-sectional endline data from a treatment group and an equivalent group created through propensity score matching
        viii. Covariate matching: Studies which collect cross-sectional endline data from a treatment group and an equivalent group created through covariate matching
        ix. Difference-in-difference: Studies which control for confounding using a difference-in-difference technique
        x. Fixed effects regression: Studies which control for confounding using a fixed effects regression technique

xi. Instrumental variables: Studies which control for confounding using an instrumental variables technique
xii. Interrupted time-series regression: Studies which control for confounding using an interrupted time-series regression analysis with at least 3 data collection points both before and after the intervention
xiii. Other regression-based study design: Studies using regression which do not fit any of the study designs listed above
xiv. Other quantitative design: Purely quantitative study using a different technique from the above
xv. Purely qualitative study

c. Any study combining quantitative and qualitative techniques should be CODED AS Mixed Methods
    i. *Note: Mixed methods studies should receive two code – one for the specific quantitative method employed and the Mixed Methods code*

4. SBDM reform
    a. Exclude Decentralisation to Higher/Lower Level: Studies that are solely related to educational decentralisation to a level *higher* than the school (e.g. decentralisation to districts) or *lower* than the school (e.g. decentralisation to families, in the form of vouchers and the like)
        i. *Note: If a study should be excluded on these grounds, no further coding is necessary*
    b. Exclude No SBDM: Studies which are about schools but in which *no change in the level of decision-making is apparent*
        i. *Note: We can include studies about any kind of decision-making reform – e.g. school management reforms, funding reforms, or curricular/pedagogical reforms – but the study must clearly report on a change in decision-making authority. Interventions which merely take place within a school but over which the school has no decision-making authority should be excluded.*
        ii. *Note: If a study should be excluded on these grounds, no further coding is necessary*
    c. Otherwise, CODE ALL that are relevant:
        i. Financial: Studies investigating contexts in which schools have been given authority over financial decision-making
        ii. Personnel: Studies investigating contexts in which schools have been given authority over decisions about personnel (e.g. hiring, firing, training, qualifications)
        iii. Other management: Studies investigating contexts in which schools have been given authority over other management decisions (e.g. not financial or personnel-related)
        iv. Curriculum: Studies investigating contexts in which schools have been given authority over curriculum decisions
        v. Pedagogy: Studies investigating contexts in which schools have been given authority over pedagogical decisions
        vi. Language of instruction: Studies investigating contexts in which schools have been given authority over decisions about language of instruction

5. Decision-making authority

a. Code ONE option between i and iv; v can also be added if appropriate
    i. Head teacher: Studies investigating contexts in which the majority of the decision-making authority has been given to the head teacher
    ii. Teachers: Studies investigating contexts in which the majority of the decision-making authority has been given to the teachers
    iii. Community: Studies investigating contexts in which the majority of the decision-making authority has been given to the community (e.g. parents)
    iv. Shared: Studies investigating contexts in which decision-making authority is shared between school officials and community members
    v. Students: Studies investigating contexts in which students have been given decision-making authority

6. Specific intervention model
    a. Code AS MANY options as are relevant:
        i. School Management Committee
        ii. Contract or Supply Teachers
        iii. School Report Cards/Social Audit
        iv. Public-Private Partnership
        v. School Capitation Grants
        vi. Other model
7. Type of education
    a. Exclude Not About Primary or Secondary Education:
        i. Study is not about education (e.g. studies of decentralisation within the health sector), OR
        ii. Study is about another level of education (e.g. pre-primary, tertiary or adult education)
            1. *<u>Note</u>: If a study should be excluded on these grounds, no further coding is necessary*
    b. Otherwise, CODE AS:
        i. Basic/Primary Education
        ii. Secondary Education
        iii. Both Primary & Secondary Education
8. Outcome
    a. Exclude No Outcomes: Studies that exclusively investigate impact on processes or outputs, instead of outcomes, including:
        i. Studies investigating a change in stakeholder *perceptions* about the decentralisation process
        ii. Studies investigating a change in stakeholder *participation*
        iii. Studies investigating a change in the *transparency of decisions* made as a result of the SBDM intervention
        iv. Studies investigating a change in *local fundraising for school activities* as a result of the SBDM intervention
            1. *<u>Note</u>: If a study should be excluded on these grounds, no further coding is necessary*
    b. Otherwise, CODE AS MANY as are relevant (*Note: All of these changes can be positive or negative)*
        i. Enrolment: Studies investigating changes in absolute enrolment levels

ii. Equity of Enrolment: Studies investigating changes in the enrolment of particular groups as a result of the SBDM intervention
iii. Teacher absenteeism: Studies investigating a change in teacher absenteeism as a result of the SBDM intervention
iv. Attendance/Retention/Progression: Studies investigating changes in student attendance, retention or progression as a result of the SBDM intervention
v. Opportunities to learn: Studies investigating a change in the quality of student opportunities to learn (e.g. infrastructure, textbooks, teaching, etc.) as a result of the SBDM intervention
vi. Cognitive Learning Outcomes: Studies investigating changes in cognitive learning outcomes (e.g. reading, math) as a result of the SBDM intervention
vii. Non-cognitive Learning Outcomes: Studies investigating changes in cognitive learning outcomes as a result of the SBDM intervention
viii. Student aspirations/attitudes/behaviours: Studies investigating changes in student aspirations, attitudes or behaviours as a result of the SBDM intervention

9. Date data collection
    a. Exclude Date Data Collection: Any study in which all data collected prior to 1990
        i. <u>Note</u>*: If a study should be excluded on these grounds, no further coding is necessary*
    b. Otherwise CODE exact date of data collection (if data collected since 1990) or as Unknown (if date of data collection cannot be identified)
10. Date intervention
    a. Exclude Context: Any study about a context that was *not classified as a LMIC at the time of the intervention/reform*
        i. <u>Note</u>*: If a study should be excluded on these grounds, no further coding is necessary*
    b. Otherwise, CODE exact date of intervention/reform or as Unknown (if date of intervention/reform cannot be identified)
11. Time lag
    a. CODE length of time between intervention and data collection or as Unknown (if date of either intervention/reform or study cannot be identified)
12. Comparisons
    a. CODE AS one of the following:
        i. Comparison yes-and-no: Studies in which a contemporaneous comparison has been made between groups in which *no* school-based decision-making reform has been attempted and groups in which *some* school-based decision-making reform has been attempted
        ii. Comparison different reforms: Studies in which a contemporaneous comparison has been made between groups in which *different* school-based decision-making reforms have been attempted (e.g. funding reforms versus school management reforms)
            1. *Note: Studies coded as contemporaneous different reforms must discuss interventions implemented during the same time period*

iii. Non-contemporaneous: Studies in which a comparison has been made but the comparison was not contemporaneous (i.e. data from the groups do not reflect the same time period)
iv. No comparison

13. Level of analysis
    a. CODE AS one of the following:
        i. Child: Data analysed at the level of the child
        ii. Teacher: Data analysed at the level of the teacher/head teacher
        iii. School: Data analysed at the level of the school/community
        iv. Sub-national: Data analysed at another sub-national (e.g. district) level
        v. Country: Data analysed at country-level (or higher)
14. Final classification
    a. Include Review Question 1: Any study following one of the includable study designs (quantitative studies options i-xii), in which a contemporaneous comparison has been made between appropriate comparison groups and in which the level of analysis has been at a local or sub-national level
    b. Include Review Question 2: Any other includable study

8.4.4. Risk of bias coding (for Research Question 1 studies)[37]

- Randomisation (if applicable)
    - Low Risk: Evidence of randomisation
    - High Risk: Evidence of self-selection or allocation based on potentially confounding criteria
        - *<u>Note:</u> Studies should not be coded as using random assignment unless the case is clear that the haphazard mechanism was random in practice. When doubt exists, studies should be coded as non-random*
    - Unclear Risk: Allocation unclear in paper
- Baseline Characteristics
    - Low Risk: Baseline characteristics across groups are reported and similar OR Differences identified but appropriate adjustments made during analysis
    - High Risk: No report of characteristics OR report of differences across groups (not adjusted for during analysis)
    - Unclear Risk: Not clear in paper if differences identified between groups OR Not clear if baseline taken
- Blind Assessment
    - Low Risk: Authors explicitly state that primary outcome variables (as defined by the authors) were assessed blindly
    - High Risk: Outcomes not assessed blindly across comparison groups
    - Unclear Risk: Not specified in the paper
- Attrition
    - Low Risk: Evidence that no random attrition occurred during the study period OR Any non-random attrition adjusted for during analysis
    - High Risk: Evidence of non-random attrition not adjusted for in analysis

[37] Based on 'Suggested risk of bias criteria for EPOC reviews', with additional questions suggested by Hombrados and Waddington (2012) and He et al. (2007)

    - Unclear Risk: No evidence of non-random attrition but not explicitly discussed
- Similarity in data collection over time
    - Low Risk: If sources and methods of data collection were the same before and after the intervention
    - High Risk: If sources and methods of data collection before and after the intervention were dissimilar
    - Unclear Risk: No discussion of similarities/differences in data collection before and after the intervention
- Missing Data
    - Low Risk: Any missing outcome measures unlikely to bias the results (e.g. the proportion of missing data was similar in the pre- and post- intervention periods or the proportion of missing data was small relative to the effect size i.e. unlikely to overturn the study result)
    - High Risk: Any missing outcome data likely to bias the results
    - Unclear Risk: Not specified in the paper
- Confounding factors
    - Low Risk: There are compelling arguments that the intervention occurred independently of other changes over time and that the outcome was not influenced by other confounding variables/events during the study period
    - High Risk: Evidence that intervention was not independent of other changes (likely that outcome was influenced by other confounding variables)
    - Unclear Risk: Other changes may have affected results but no clear evidence either way
- Clustering (if applicable)
    - Low Risk: Evidence that authors control for external cluster-level factors that might confound the results
    - High Risk: Evidence that authors have not controlled for external cluster-level factors that might confound the results
    - Unclear Risk: Potential for external cluster-level confounding factors; unclear if controlled for in analysis
- Motivation Bias
    - Low Risk: Differences in outcomes across groups unlikely to be influenced by participant motivation as a result of programme implementation and/or monitoring
    - High Risk: Differences in outcomes across groups likely to have been influenced by participant motivation as a result of programme implementation and/or monitoring
    - Unclear risk: Unclear if differences in outcomes across groups have been influenced by participant motivation
- Other Validity Threats
    - Low Risk: Results of the study unlikely to have been affected by recall bias, researcher bias, social desirability bias or other threats to validity
    - High Risk: Results of the study likely to have been affected by recall bias, researcher bias, social desirability bias or other threats to validity
- Data Mining

    - Low Risk: The study does not suggest the existence of biased exploratory research methods (e.g. multiple sub-groups not specified in protocol or theory)
    - High Risk: Authors appear to have used biased exploratory research methods
- Spill-overs/Contamination
    - Low Risk: Unlikely that comparison group affected by the intervention
    - High Risk: Likely that the comparison group was affected by the intervention
    - Unclear Risk: Spill-over effects may have occurred but not clear in paper
- Risk of Selective Outcome Reporting
    - Low Risk: No evidence that outcomes were selectively reported
    - High Risk: Some important outcomes listed in methods section are omitted from the results
    - Unclear Risk: Not specified in the paper
- Other Risk of Bias
    - Low Risk: No evidence of other risk of biases (including uncorrected unit of analysis error, evidence of heterogeneity between sub-groups, insignificance due to lack of power, and/or evidence of unaccounted for heteroschedasticity)
    - High Risk: Evidence of other risk of biases
- Final assessment
    - Low Risk: The study
        - Demonstrates clear measurement of and control for confounding, including selection bias, and has no suspected sources of unobserved confounding;
        - Adequately describes the reform/intervention and comparison groups;
        - Has low risk of spillovers or contamination; and,
        - Demonstrates low risk of reporting biases and other sources of bias.
    - Medium Risk:
        - There are moderate threats to the validity of the attribution methodology (arising from issues with the implementation of the methodology), or
        - There are either likely risks of spillovers or contamination (arising from inadequate description of the intervention or comparison groups) or possibilities for interaction between groups (e.g. drawn from the same community), or
        - There are possible reporting biases.
    - High Risk
        - Studies where the study design is of questionable causal validity, such as those where comparison groups are not matched on observables, differences in covariates are not accounted for in multivariate analysis, or where there are serious threats to the validity of the statistical procedure used to deal with attribution; or
        - Where there is clear evidence of spillovers or contamination to comparison groups from the same communities; or
        - Where reporting biases are evident.
- Include/Exclude
    - Include for RQ1 synthesis: Studies classified as Low or Medium Risk
    - Quality appraisal for RQ2: Studies classified as High Risk

8.4.5. Coding for quality appraisal (for Research Question 2 studies)[38]

Transparency

- Research Question
    - High Transparency: Study has a clear research question
    - Low Transparency: Study does not have a clear research question
- Transparency of Research Design
    - High: Study clearly states the design and methods
    - Low: Study does not state clearly the design and methods
- Transparency of Data Source
    - High: Study clearly references which data were used and where they came from (source and/or how collected)
    - Low: Study does not clearly reference which data were used and where they came from (source and/or how collected)

Appropriateness

- Appropriateness of Research Design
    - High: Research design is appropriate for the research question
    - Low: Research design is not appropriate for the research question
- Appropriateness of Sampling Method
    - High: Sampling method appropriate for research question and design
    - Low: Sampling method inappropriate for research question and design
    - Unclear: Sampling method unclear
- Appropriateness of Sample Size
    - High: Final sample size appropriate for analytical method
    - Low: Final sample size inappropriate for analytical method
    - Unclear: Sample size unclear
- Appropriateness of Sample
    - High: Sample representative of the population and/or pertinent to the purpose
    - Low: Final sample not representative of the population and/or pertinent to the purpose
    - Unclear: Sample characteristics unclear
- Appropriateness of Data Collection Methods
    - High: Data collection methods appropriate for the research design
    - Low: Methods inappropriate for the research design
    - Unclear: Details of data collection methods not provided
- Appropriateness of Analytical Methods
    - High: Analytical techniques appropriate for the research design
    - Low: Analytical techniques inappropriate for the research design
    - Unclear: Details of data analysis not provided
- Appropriateness of Unit of Analysis
    - High: Unit of analysis equivalent to unit of intervention OR unit of analysis not equivalent to unit of intervention, but clustering taken into account in analysis

[38] Based on DFID (2014)

    - Low: Unit of analysis not equivalent to unit of intervention and clustering not taken into account in analysis
    - Unclear: Unit of analysis not equivalent to unit of intervention but unclear if clustering was taken into account in analysis
    - N/A: Studies which do not need to take clustering into account (e.g. qualitative studies)
- Recruitment Ethics
    - High: Recruitment methods appropriate and ethical
    - Low: Recruitment methods inappropriate and/or unethical
    - Unclear: Recruitment methods not clear
    - Not Applicable (no participants)
- Other Ethical Considerations
    - High: Ethics clearly considering during study implementation; no ethical concerns
    - Low: Ethical concerns
    - Unclear: Ethics not discussed

Rigour

- Validity of Data
    - High: Indicators/data suited to concept in question
    - Low: Indicators/data not suited to concept in question
- Validity of Methods
    - High: Data collection method able to validly measure the indicators/data
    - Low: Data collection method not a valid measure of indicators/data
    - Unclear: Details of data collection methods not provided
- Execution of Analytical Methods
    - High: Analytical techniques adequately executed
    - Low: Analytical techniques inadequately executed
    - Unclear: Details of data analysis not provided
- Internal Validity
    - High: Analysis satisfactorily and credibly answers the question (i.e. study takes into account other possible factors, causes or explanations)
    - Low: Analysis does not satisfactorily or credibly answer the question (does not take into account other possible factors, causes or explanations)
- External Validity
    - High: The results can be generalised to the extent advocated by the author; sampling method valid and consistent with conclusions
    - Low: The author makes claims beyond the scope supported by the data; sampling method invalid and/or inconsistent with conclusions
    - Unclear: Sampling method unclear
- Replicability
    - High: Evidence of consistency in analysis (likely to be replicated or confirmed)
    - Low: Evidence of inconsistencies in analysis
    - Unclear: Details of analysis not provided
- Reliability Testing

    - High: Study includes evidence of testing for reliability (at pilot or main study phase)
    - Low: No evidence of testing for reliability during study
- Supported Conclusions
    - High: Conclusions clearly backed up by data and findings
    - Low: Conclusions not backed up by data and findings
    - Unclear: Sampling method unclear

Cogency

- Consistency of Implementation
    - High: Data collection appears to be consistent across the study (i.e. same methods used with all participants)
    - Low: Evidence of inconsistencies in data collection
    - Unclear: Details of data collection not provided
- Consistency of Argument
    - High: Clear argument runs through the entire paper, linking the conceptual frame to the results
    - Low: Logical inconsistencies in argument of the paper OR no conceptual or theoretical grounding to paper (including no justification for methods used)

- Overall Assessment
    - 'High' quality: Studies which have received a 'High Quality' code for each of the dimensions assessed.
    - 'Medium' quality: Studies which have received 'High Quality' designations for all transparency indicators, for all indicators related to the appropriateness of the research design, for all validity indicators and for evidence of supported conclusions but may have received a designation of 'Unclear' for some of the methodological indicators (e.g. details of data collection or analysis).
    - 'Low' quality: Any study receiving at least one 'Low Quality' code
- Include/Exclude
    - Exclude Low Quality: All studies classified as Low Quality
    - Include for Synthesis: All studies classified as High or Medium Quality

8.4.6. Coding for Meta-Analysis

Geographic Region
1 = Latin America
2 = MENA
3 = SSA
4 = South West Asia
5 = East Asia

Country
1 = Brazil
2 = Columbia
3 = El Salvador
4 = Guatemala
5 = Honduras

6 = India
7 = Indonesia
8 = Kenya
9 = Madagascar
10 = Mexico
11 = Nicaragua
12 = Niger
13 = Pakistan
14 = Philippines
15 = Uganda

Income Level
1 = Low income
2 = Low middle income
3 = higher middle income

Follow up time (months)
-Coded as number of months
-99 no follow up

School Level
1 = Pre-school
2 = Primary level
3 = Secondary school
4 = Other

Analysis by sub groups included?
1 = Included
2 = Not included

Study design (RCT or quasi-experimental)
1 = RCT
2 = Quasi-Experimental (e.g. DID, propensity score matching)
3 = Other studies rated as of Medium quality (e.g. IV)

Unit of Analysis (level)
1 = School
2 = Child
3 =Other
4 = Teacher
5 = Classroom
6 = Parents

Outcome
1 = drop-out
2 = repetition
3 = failure
4 = absence
5 = language score (L2)
6 = maths score
7 = science score
8 = aggregate test score
9 = enrolment
10 = grade progression
11 = presence/attendance

12 = teacher presence/attendance
13 = teacher absenteeism
14 = teacher retention
15 = teacher activity
16 = language (L1)
17 = literacy

8.4.7. Coding for qualitative synthesis

Specific name of intervention

1. Unnamed government reform (multiple countries)
2. Unnamed government reform (Madagascar)
3. EDUCO (El Salvador)
4. PROHECO (Honduras)
5. Extra Teacher Program (Kenya)
6. PDE (Brazil)
7. Rural Education Program (Colombia)
8. Whole School Development
9. Quality Schools Program - PEC (Mexico)
10. Support to School Management - AGE (Mexico)
11. Third Elementary Education Project - TEEP (Philippines)
12. School Autonomy Reform (Nicaragua)
13. Sarva Siksha Aviyan (SSA) (India)
14. Unnamed government reform (Indonesia)
15. Democratic School leadership (Philippines)
16. ESDFP (Sri Lanka)
17. School Based Management (Philippines)

Level of decentralization

1. Very decentralized (e.g. most decisions devolved to school/community level, including the hiring/firing of teachers)
2. Somewhat decentralized (e.g. some decisions devolved to school/community level – typically financial/management and not personnel)
3. Not very decentralized (e.g. some decisions devolved to school/community level – e.g. development of school improvement plans but without any financial decision-making authority, except over community contributions)

Primary decision makers at local level

1. School (head and/or teachers)
2. Community/Parents
3. Shared (SMC includes mix of school and community reps with no clear majority)

*Decisions devolved to community level (*de jure *decision making authority)*

1. Personnel (yes/no)
2. Financial (yes/no)
3. Other management, such as school building maintenance, development of school improvement plans, etc. (yes/no) – If yes, please specify: _____
4. Pedagogy (yes/no)
5. Curriculum (yes/no)
6. School admissions (yes/no)
7. Language of instruction (yes/no)

*Decisions actually taken by community level (*de facto *decision making authority)*

1. Personnel (yes/no)
2. Financial (yes/no)
3. Other management, such as school building maintenance, development of school improvement plans, etc. (yes/no) – If yes, please specify: _____
4. Pedagogy (yes/no)
5. Curriculum (yes/no)
6. School admissions (yes/no)
7. Language of instruction (yes/no)

Implementation factors

1. Capitation grant provided to school (yes/no)
2. SMC members elected (yes/no)
3. SMC members trained (yes/no)
4. Linkages established (yes/no)
5. Use of report card or other information-sharing mechanism (yes/no)

8.5. RISK OF BIAS ANALYSIS

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
Bando (2010)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	None	None	Low Risk of Bias
Barr et al. (2012)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	None	None	Low Risk of Bias
Beasley & Huillery (2014)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	None	None	Low Risk of Bias

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
Blimpo & Evans (2011)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	None	None	Low Risk of Bias
Bold et al. (2013)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	None	None	Low Risk of Bias
Carnoy et al. (2008)	Baseline Characteristics; Blind Assessment; Similarity in data collection over time; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	Attrition; Missing Data; Confounding factors	Randomisation	Medium Risk of Bias
Cueto et al (2008)	Baseline Characteristics; Blind Assessment; Similarity in data collection over time; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	Attrition; Missing Data;	None	Randomisation	High Risk of Bias

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
Di Gropello & Marshall (2005)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	Confounding factors	Randomisation	Medium Risk of Bias
Duflo et al. (2012)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	None	None	Low Risk of Bias
Gertler et al. (2012)	Baseline Characteristics; Attrition; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	Blind Assessment; Confounding factors	Randomisation	Medium Risk of Bias
Glewwe & Maïga (2011)	Randomisation; Baseline characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Risk of Selective Outcome Reporting; Other Risk of Bias	Confounding factors; Spill-overs/contamination	None	None	None	Medium Risk of Bias

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
Jimenez & Sawada (1999)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Confounding factors; Clustering; Data Mining; Missing Data; Risk of Selective Outcome Reporting	None	Other Validity Threats; Other Risk of Bias	Motivation Bias; Spill-overs/contamination	Randomisation	Medium Risk of Bias
Jimenez & Sawada (2003)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Confounding factors; Clustering; Data Mining; Risk of Selective Outcome Reporting	None	Missing Data; Other Validity Threats; Other Risk of Bias	Motivation Bias; Spill-overs/contamination	Randomisation	Medium Risk of Bias
Khattri et al. (2010)	Baseline Characteristics; Blind Assessment; Similarity in data collection over time; Clustering; Motivation Bias; Data Mining; Risk of Selective Outcome Reporting	Confounding factors; Spill-overs/contamination	Other Validity Threats; Other Risk of Bias	Attrition; Missing Data	Randomisation	Medium Risk of Bias
King & Ozler (2005)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Risk of Selective Outcome Reporting; Other Risk of Bias	Confounding factors; Spill-overs/contamination	None	Missing Data	Randomisation	Medium Risk of Bias
Lassibille et al. (2010)	Randomisation; Baseline characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Risk of Selective Outcome Reporting; Other Risk of Bias	Confounding factors; Spill-overs/contamination	None	Missing Data	None	Medium Risk of Bias

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
Murnane et al. (2006)	Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Motivation Bias; Other Validity Threats; Data Mining; Risk of Selective Outcome Reporting; Other Risk of Bias	Baseline characteristics; Confounding factors; Clustering; Spill-overs/contamination	None	None	Randomisation	Medium Risk of Bias
Paes de Barros & Mendonca (1998)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	Confounding factors	Clustering	Randomisation	High Risk of Bias
Parker (2005)	Baseline characteristics; Blind Assessment; Attrition; Confounding; Similarity in data collection over time; Missing Data; Data Mining; Risk of Selective Outcome Reporting; Other Risk of Bias	None	Other Validity Threats	Clustering; Motivation Bias; Spill-overs/contamination	Randomisation	Medium Risk of Bias
Pradhan et al. (2011)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	None	None	Low Risk of Bias

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
Rodriguez et al. (2010)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Other Validity Threats; Data Mining; Risk of Selective Outcome Reporting; Other Risk of Bias	Spill-overs/contamination	Motivation Bias	Confounding factors; Clustering	Randomisation	Medium Risk of Bias
San Antonio (2008)	Randomisation; Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Spill-overs/Contamination; Risk of Selective Outcome Reporting	Confounding factors	Data mining; Other Risk of Bias	None	None	Medium Risk of Bias
Santibanez et al. (2014)	Baseline Characteristics; Blind Assessment; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Data Mining	None	Other Risk of Bias	Attrition; Confounding factors; Spill-overs/contamination ; Risk of Selective Outcome Reporting	Randomisation	Medium Risk of Bias
Sawada & Ragatz (2005)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Clustering; Spill-overs/Contamination; Risk of Selective Outcome Reporting	None	Other Validity Threats; Data Mining; Other Risk of Bias	Confounding factors; Motivation bias	Randomisation	Medium Risk of Bias
Skoufias & Shapiro (2006)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Confounding factors; Clustering; Motivation Bias	Spill-overs/contamination	Other Validity Threats; Data Mining; Risk of Selective Outcome Reporting	None	Randomisation	Medium Risk of Bias

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
World Bank (2011)	Randomisation; Baseline Characteristics; Blind Assessment; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	Attrition; Confounding factors	None	Medium Risk of Bias
World Bank (2013)	Randomisation; Attrition; Baseline Characteristics; Blind Assessment; Missing Data; Clustering; Motivation Bias; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	Other Validity Threats	Similarity in data collection over time; Confounding factors	None	Medium Risk of Bias
Yamauchi (2014)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Risk of Selective Outcome Reporting; Other Risk of Bias	Confounding factors; Spill-overs/contamination	None	None	Randomisation	Medium Risk of Bias
Yamauchi & Liu (2012)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Clustering; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	None	Confounding factors	Randomisation	Medium Risk of Bias
Cueto et al. (2008)	Baseline Characteristics; Blind Assessment; Similarity in data collection over time; Confounding factors; Clustering; Motivation	None	Attrition; Missing Data;	None	Randomisation	High Risk of Bias*

Citations	Areas of Low Risk	Areas of Medium Risk	Areas of High Risk	Areas of Unclear Risk	Categories not applicable	Final Assessment
	Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias					
De Umanzor et al. (1997)	Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Data Mining; Spill-overs/contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	Other Validity Threats	Baseline Characteristics; Confounding factors; Clustering; Motivation Bias	Randomisation	High Risk of Bias*
Paes de Barros & Mendonca (1998)	Baseline Characteristics; Blind Assessment; Attrition; Similarity in data collection over time; Missing Data; Motivation Bias; Other Validity Threats; Data Mining; Spill-overs/Contamination; Risk of Selective Outcome Reporting; Other Risk of Bias	None	Confounding factors	Clustering	Randomisation	High Risk of Bias*

Note: * High risk of bias studies excluded from meta-analysis.

8.6. QUALITY APPRAISAL OF INCLUDED AND EXCLUDED NON-CAUSAL STUDIES

Included studies

Full Citation	Results of quality appraisal
Bandur, A. (2008). *A study of the implementation of school-based management in Flores primary schools in Indonesia*. Unpublished PhD thesis.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection and analysis; High internal validity; High replicability; Well-supported conclusions; Consistent argument
Bjork, C. (2003). 'Local Responses to Decentralization Policy in Indonesia.' *Comparative Education Review*, 47 (2): 184-216.	High transparency; Appropriate design and unit of analysis; Sampling methodology unclear; Evidence of consistency in data collection and analysis; High internal validity; Unclear how well study could be replicated; Well-supported conclusions; Consistent argument
de Umanzor S., Soriano, I., Vega, M.R., Jimenez, E., Rawlings, L., & Steele, D. (1997). *El Salvador's EDUCO Program: A First Report on Parents' Participation in School-Based Management*. Working Paper Series on Impact of Education Reforms, Paper No. 4. Washington, DC: World Bank.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection and analysis; High internal validity; High replicability; Well-supported conclusions; Consistent argument
Fuller B. & Rivarola, M. (1998). *Nicaragua's Experiment to decentralize schools: views of parents, teachers and directors*. Working Paper Series on Impact of Education Reforms, Paper No. 5. Washington, DC: World Bank.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection and analysis; High internal validity; High replicability; Well-supported conclusions; Consistent argument
Gershberg, A.I. & Meade, B. (2005). 'Parental Contributions, School-Level Finances and Decentralization: An Analysis of Nicaraguan Autonomous School Budgets.' *Comparative Education*, 41 (3): 291-308.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection and analysis; High internal validity; High replicability; Well-supported conclusions; Consistent argument
Gunnarsson V., Orazem P.F., Sanchez M.A., & Verdisco, A. (2008). *Does Local School Control Raise Student Outcomes?: Theory and Evidence on the Roles of School Autonomy and Community Participation*. Working Paper No. 09012. Ames, IA: Iowa State University.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection and analysis; High internal validity; High replicability; Well-supported conclusions; Consistent argument
Hanushek, E.A., Link, S., & Woessmann, L. (2011). *Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA*. NBER Working Paper No. 17591. Washington, DC: National Bureau of Economic Research.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection and analysis; High internal validity; High replicability; Well-supported conclusions; Consistent argument
Reimers, F. & Cardenas, S. (2007). 'Who Benefits from School-Based Management in Mexico?' Prospects: Quarterly Review of *Comparative Education*, 37 (1): 37-56.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection and analysis; High internal validity; High replicability; Well-supported conclusions;

Full Citation	Results of quality appraisal
	Consistent argument
Vernez, G., Karam, R., & Marshall. J.H. (2012). *Implementation of School-Based Management in Indonesia*. Monograph. Santa Monica, CA: RAND Corporation.	High transparency; Appropriate design and unit of analysis; Evidence of consistency in data collection; Some aspects of data analysis unclear; High internal validity; High replicability; Well-supported conclusions; Consistent argument

Excluded studies

Full Citation	Reason for low quality assessment
(2013). *Interim Support to Education Programme (INSTEP) Project Completion Review*. London: DFID.	No clear research question; Lack of transparency regarding research design; Inappropriate unit of analysis; Low internal validity; Low external validity
Abdinoor, A. (2008). 'Community Assumes the Role of State in Education in Stateless Somalia'. *International Education* 37(2): 43-61.	No clear research question; Unclear sampling method; Unclear analytical methods
Akyeampong, K. (2011). *(Re)Assessing the Impact of School Capitation Grants on Educational Access in Ghana*. CREATE Pathways to Access Research Monograph No. 71. Brighton: University of Sussex.	Inappropriate research design; Low internal validity; No evidence of reliability testing of instruments; Unclear sampling method/sample size/sample characteristics; Unclear execution of analytical methods; Unclear if conclusions supported
Amirrachman, A., Syafi'i, S. and Welch, A. (2008). 'Decentralising Indonesian education: the promise and the price'. *World Studies in Education* 9(1): 31-53.	No clear research question; Lack of transparency regarding research design; Lack of transparency regarding data source; Low internal validity; Unsupported conclusions
Chowdhury, M.D., Al-Mahmood, A., Bashar, M.A., and Ahmed, J.U. (2011). *Localization of Digital Content for Use in Secondary Schools of Bangladesh*.	Lack of transparency regarding research design; Inappropriate analytical methods; Unclear data collection methods
Condy, A. (1998). *Improving the Quality of Teaching and Learning Through Community Participation: Achievements, Limitations and Risks: Early lessons from the Schooling Improvement Fund in Ghana*. Social Development Working Paper No. 2. London: DFID.	Lack of transparency regarding research design; Lack of transparency regarding data source
Cossou, M. (2000). *Recherche opérationnelle sur la coopération en éducation de base dans les pays francophones d'Afrique de l'Ouest : cas du Bénin* (Operational research on cooperation in basic education in the francophone countries of West Africa: the case of Benin). Montreal: Fondation Paul Gérin-Lajoie; Ottawa: International Development Research Centre.	Inappropriate research design; Inappropriate analytical methods; Low internal validity; Unclear sampling method; Unclear sample size; Unclear data collection methods; Unclear if conclusions supported

Full Citation	Reason for low quality assessment
Dowd, A. and Namathaka, L. (2007). 'Malawi, 1994–2003: Training on a National Scale'. In D. Glassman, J. Naidoo, and F. Woods (eds), *Community schools in africa: Reaching the unreached*. New York: Springer.	Lack of transparency regarding research design; Lack of transparency regarding data source
Ekosiswoyo, R., Evans, D.P., Thair, M., and Wello, M.B. (2007). *Final evaluation: Managing Basic Education (MBE) Project*. Washington, DC: The Mitchell Group.	No clear research question; Lack of transparency regarding research design; Lack of transparency regarding data source
Holger, D. (2007). *School decentralization in the context of globalizing governance: international comparison of grassroots responses*. Dordrecht: Springer.	Lack of transparency regarding research design; Lack of transparency regarding data source
Jones, A. (2005). 'Conflict, development and community participation in education: Pakistan and Yemen'. *Internationales Asienforum* 36(3-4): 289-310.	No clear research question; Lack of transparency regarding research design; Lack of transparency regarding data source
Pailwar, V.K., and Mahajan, V. (2005). 'Janshala in Jharkhand: An Experiment with Community Involvement in Education'. *International Education Journal* 6(3): 373-385.	No clear research question; Lack of transparency regarding research design; Low replicability; Unclear if analytical methods are appropriate
Tate, S, and Amedie, W.Y. (2011). *Mid-term evaluation of the USAID community-school partnership program for education and health*.	Inappropriate research design; Inappropriate sample size; Inappropriate data collection methods; Low internal validity; Low external validity; Unsupported conclusions
Updadhaya, H., Dubey, N., and Shrestha, O. (2007). *Understanding School Autonomy: A Study on Enabling Conditions for School Effectiveness*. Kathmandu: Research Centre for Educational Innovation and Development.	Invalid methods; Low internal validity; Unclear sampling method/sample size/sample characteristics; Unclear execution of analytical methods; Unclear if conclusions supported
Vasquez, W.F. (2012). 'Supply-Side Interventions and Student Learning in Guatemala'. *International Review of Education* 58(1): 9-33.	Inappropriate unit of analysis; Low internal validity; Unsupported conclusions; Unclear data collection methods
Wadesango, N. (2012). 'The influence of teacher participation in decision-making on their occupational morale.' *Journal of social sciences* 31(3): 361-369.	Lack of transparency regarding research design; Low replicability; Unclear sampling method/sample size/sample characteristics; Unclear execution of methods; Unclear if conclusions are supported
Wanzare, Z. (2012). 'Instructional Supervision in Public Secondary Schools in Kenya'. *Educational Management Administration & Leadership* 40(2): 188-216.	Inappropriate unit of analysis; Low internal validity; Low replicability; Unclear sampling method and sample characteristics; Unclear if conclusions are supported
Yousuf, M.I., Alam, M.T., Sajjad, M.L, and Imran, M. (2010). 'Amelioration of Educational Conditions through School Management Committees.' *Journal of College*	Inappropriate research design; Inappropriate unit of analysis; Low internal validity; Unsupported conclusions; Inconsistent argument; Unclear sampling method/sample size/sample

Full Citation	Reason for low quality assessment
Teaching & Learning 7(9): 47-52.	characteristics; Unclear execution of analytical methods
Yuki, T., Mizuno, K., Ogawa, K., and Mihoko, S. (2013). 'Promoting gender parity in basic education: lessons from a technical cooperation project in Yemen'. *International Review of Education* 59(1): 47-66.	Inappropriate analytical methods; Unclear sampling method/sample size/sample characteristics; Unclear replicability; Unclear if conclusions supported

Supplement 1: Details of included impact studies

Citation	Region	Country	Income Level	School Level	Study Design	Outcomes included in meta-analysis	Intervention	Level of Decentralisation	Publication Status
Bando (2010)	Latin America	Mexico	Upper Middle	Secondary	Quasi-Experiment and Other	Drop-out; Test scores (language, math)	PEC	Medium	PhD thesis
Barr et al. (2012)	Sub-Saharan Africa	Uganda	Low	Primary	RCT	Teacher Attendance; Test scores (aggregate)	Unnamed	Medium	CSAE Working Paper
Beasley and Huillery	Sub-Saharan	Niger	Low	Primary	RCT	Drop-out; Teacher Attendance;	Unnamed	Medium	PAL Working Paper

Citation	Region	Country	Income Level	School Level	Study Design	Outcomes included in meta-analysis	Intervention	Level of Decentralisation	Publication Status
(2014)	Africa					Test scores (language, math)			
Blimpo and Evans (2011)	Sub-Saharan Africa	Gambia	Low	Primary	RCT	Teacher Attendance; Test scores (language, math)	Whole School Development	Medium	World Bank Report
Bold et al. (2013)	Sub-Saharan Africa	Kenya	Low	Primary	RCT	Test scores (aggregate)	Extra Teacher Program	High	CSAE Working Paper
Di Gropello and Marshall (2005)	Latin America	Honduras	Lower Middle	Primary	Quasi-Experiment and Other	Test scores (language, math)	PROHECO	High	Chapter in World Bank Publication
Duflo et al. (2012)	Sub-Saharan Africa	Kenya	Low	Primary	RCT	Teacher Attendance; Test scores (language,	Extra Teacher Program	High	NBER Working Paper

Citation	Region	Country	Income Level	School Level	Study Design	Outcomes included in meta-analysis	Intervention	Level of Decentralisation	Publication Status
						math)			
Gertler et al. (2012)	Latin America	Mexico	Upper Middle	Primary	Quasi-Experiment and Other	Drop-out; Repetition	AGE	Medium	Journal Article
Glewwe & Maïga (2011)	Sub-Saharan Africa	Madagascar	Low	Primary	RCT	Test scores (aggregate)	AGEMAD	Medium	PAL Working Paper
Jimenez and Sawada (1999)	Latin America	El Salvador	Lower Middle	Primary	Quasi-Experiment and Other	Test scores (language, math)	EDUCO	High	World Bank Published Journal Article
Jimenez and Sawada (2003)	Latin America	El Salvador	Lower Middle	Primary	Quasi-Experiment and Other	Repetition	EDUCO	High	CIRJE Discussion Paper
Khattri et al. (2010)	South-East Asia	Philippines	Lower Middle	Primary	Quasi-Experiment	Test scores (language,	TEEP	Medium	World Bank Working

Citation	Region	Country	Income Level	School Level	Study Design	Outcomes included in meta-analysis	Intervention	Level of Decentralisation	Publication Status
					and Other	math)			Paper
King and Ozler (2005)	Latin America	Nicaragua	Low	Multiple*	Quasi-Experiment and Other	Test scores (language, math)	Autonomous Schools	High	KU Discussion Paper
Lassibille et al. (2010)	Sub-Saharan Africa	Madagascar	Low	Primary	RCT	Drop-out; Repetition; Teacher Attendance; Test scores (language math)	AGEMAD	Medium	World Bank Published Journal Article
Murnane et al. (2006)	Latin America	Mexico	Upper Middle	Primary	Quasi-Experiment and Other	Drop-out	PEC	Medium	Working Paper
Parker (2005)	Latin America	Nicaragua	Lower Middle	Primary*	Quasi-Experiment and Other	Test scores (language, math)	Autonomous Schools	High	Chapter in World Bank Publication

Citation	Region	Country	Income Level	School Level	Study Design	Outcomes included in meta-analysis	Intervention	Level of Decentralisation	Publication Status
Pradhan et al. (2011)	South-East Asia	Indonesia	Lower Middle	Primary	RCT	Drop-out; Repetition; Test scores (language, math)	Unnamed	Medium	World Bank Working Paper
Rodriguez et al. (2010)	Latin America	Colombia	Lower Middle	Multiple	Quasi-Experiment and Other	Drop-out; Test scores (language, math)	Rural Education Program	Low	Journal Article
San Antonio (2008)	South-East Asia	Philippines	Lower Middle	Primary	RCT	Test scores (aggregate)	Democratic School leadership	Medium	Journal Article
Santibanez et al. (2014)	Latin America	Mexico	Upper Middle	Primary*	Quasi-Experiment and Other	Drop-out; Test scores (language, math)	PEC-FIDE	Medium	Journal Article
Sawada and Ragatz	Latin	El Salvador	Lower	Primary	Quasi-Experiment	Teacher Attendance;	EDUCO	High	Chapter in World Bank

Citation	Region	Country	Income Level	School Level	Study Design	Outcomes included in meta-analysis	Intervention	Level of Decentralisation	Publication Status
(2005)	America		Middle		and Other	Test scores (language, math)			Publication
Skoufias and Shapiro (2006)	Latin America	Mexico	Upper Middle	Primary	Quasi-Experiment and Other	Drop-out; Repetition	Quality Schools Program - PEC	Medium	World Bank Working Paper
World Bank (2011)	South Asia	Sri Lanka	Lower Middle	Primary	RCT	Teacher Attendance; Test scores (language, math)	PSI	Medium	World Bank Working Paper
World Bank (2013)	South-East Asia	Philippines	Lower Middle	Primary	Quasi-Experiment and Other	Test scores (math, aggregate)	School Based Management	Medium	World Bank Report
Yamauchi (2014)	South-East Asia	Philippines	Lower Middle	Primary	Quasi-Experiment and Other	Test scores (aggregate)	School Based Management	Medium	World Bank Working Paper

Citation	Region	Country	Income Level	School Level	Study Design	Outcomes included in meta-analysis	Intervention	Level of Decentralisation	Publication Status
Yamauchi and Liu (2012)	South-East Asia	Philippines	Lower Middle	Primary	Quasi-Experiment and Other	Test scores (math, aggregate)	TEEP	Medium	World Bank Working Paper

Supplement 2: Effect size data computed

Study	Study ID	Country	Outcome	Follow up time	Intervention /Model	SMD	SE(SMD)	Type Data (ES Method)	Notes	Sample Size (N)
Skoufias and Shapiro (2006)	1	Mexico	Drop-out	36	LLR-PSM, Time T	-.0690791	.0263021	Regression co-efficient formula		67224
Pradhan et al. (2011)	5	Indonesia	Drop-out	21	Linkage + Election ITT	1.85073	1.58634	Regression co-efficient formula	SD estimated (see 1 below table)	11463
Murnane et al. (2006)	7	Mexico	Drop-out	36		-.0675147	.0344473	Regression co-efficient formula	Use p-value (**)	15014
Bando (2010)	16	Mexico	Drop-out	12	5-year exposure	-.0450883	.020076	Means and standard deviations formula		516474
Beasley and Huillery (2014)	17	Niger	Drop-out	12		-.055903	.0520028	Regression co-efficient formula		748
Gertler et al. (2012)	20	Mexico	Drop-out	0	Model 4	.0224559	.0224559	Regression co-efficient formula		6027
Lassibille et	21	Madagascar	Drop-out	12		-.027	.08124	Means and	Use	Not

al. (2010)								standard errors	standard error of mean difference	reported
Rodriguez et al. (2010)	24	Colombia	Drop-out	36		-.2315436	.0217072	Use t-test method, equal samples	No SDs	17297
Santibanez et al. (2014a)	26	Mexico	Drop-out	12	No previous PEC, Grade 3	-.0201791	.1834709	Regression co-efficient formula		3673
Santibanez et al. (2014b)	31	Mexico	Drop-out	12	No previous PEC, Grade 6	.0101232	.1809194	Regression co-efficient formula		3575
Skoufias and Shapiro (2006)	1	Mexico	Repetition	36	LLR-PSM, Time T	-.1040007	.0225944	Regression co-efficient formula		67224
Pradhan et al. (2011)	5	Indonesia	Repetition	21	Linkage + Election ITT	.7768427	.5548876	Regression co-efficient formula	SD estimated (see 1 below table)	11463
Jimenez and Sawada (2003)	12	El Salvador	Repetition	24		-.0387	.086	Use t-test method, unequal samples	No SDs	820
Gertler et al. (2012)	20	Mexico	Repetition	0	Model A	-.0547634	.0365089	Regression co-efficient formula		6027
Lassibille et al. (2010)	21	Madagascar	Repetition	21		-.1626	.08137	Means and standard errors	Use standard error of mean difference	Not reported
King and Ozler (2005a)	3	Nicaragua	Language (Spanish)	0	T5 De jure, primary G4	.148	.274	Regression co-efficient formula		967

Khattri et al. (2010)	4	Philippines	Language (English)	24	G6	.097171	.0437707	Regression co-efficient formula		5167
Pradhan et al. (2011)	5	Indonesia	Language (Bahasa Indonesia)	21	Linkage + Election ITT, G6	.2151595	.0926381	Regression co-efficient formula		11463
Di Gropello and Marshall (2005)	6	Honduras	Language (Spanish)	0	2002 control, predicted , G3	.4543711	.721224	Regression co-efficient formula		1257
Sawada and Ragatz (2005)	8	El Salvador	Language (Spanish)	0	G3	.0122757	.1515524	Regression co-efficient formula		606
World Bank (2011)	9	Sri Lanka	Language (English)	30	PSI only, G4	.2299696	.0724506	Regression co-efficient formula		4727
Bando (2010)	16	Mexico	Language (Spanish)	12	5-year exposure	.0652496	.0295898	Means and standard deviations formula		114722
Beasley and Huillery (2014)	17	Niger	Language (French)	6	G1,4,6	-.0438553	.0581408	Regression co-efficient formula		739
Blimpo and Evans (2011)	18	Gambia	Language (English)	36	WSD, G4,6	-.0941602	.2118605	Regression co-efficient formula		814
Duflo et al. (2012)	19	Kenya	Language (English)	15	Literacy ETP + SBM + CT, G1	.256	.108	Regression co-efficient formula		6533
Lassibille et al. (2010)	21	Madagascar	Language (French)	21	G4	.0009	.01908	Means and standard errors	Use standard error of mean difference	Not reported
Parker (2005a)	23	Nicaragua	Language (Spanish)	0	Grade 3, G3	.0538	.0768	Use t-test method, unequal samples	No SDs	686

Rodriguez et al. (2010)	24	Colombia	Language (Spanish)	36	G5	.1015	.0387	Use t-test method, equal samples	No SDs	2762
Santibanez et al. (2014a)	26	Mexico	Language (Spanish)	12	No previous PEC, G3	.4808173	.1459382	Regression co-efficient formula		3673
King and Ozler (2005b)	30	Nicaragua	Language (Spanish)	0	T5 De jure , secondary	.136	.451	Regression co-efficient formula		911
Santibanez et al. (2014b)	31	Mexico	Language (Spanish)	12	No previous PEC, G6	-.2197704	.1389939	Regression co-efficient formula		3673
Parker (2005b)	32	Nicaragua	Language (Spanish)	0	G6	-.0775	.0707	Use t-test method, unequal samples	No SDs	830
King and Ozler (2005a)	3	Nicaragua	Mathematics	0	T5 De jure, primary G4	-.232	.306	Regression co-efficient formula		967
Khattri et al. (2010)	4	Philippines	Mathematics	24	G6	.1100817	.0663143	Regression co-efficient formula		5167
Pradhan et al. (2011)	5	Indonesia	Mathematics	21	Linkage + Election ITT, G6	.0688202	.0508671	Regression co-efficient formula		11463
Di Gropello and Marshall (2005)	6	Honduras	Mathematics	0	2002 control, predicted, G3	.5852292	.6160308	Regression co-efficient formula		1253
Sawada and Ragatz (2005)	8	El Salvador	Mathematics	0	G3	.0649582	.1619904	Regression co-efficient formula		606
World Bank (2011)	9	Sri Lanka	Mathematics	30	PSI only, G4	.2128106	.0741935	Regression co-efficient formula		4746
World Bank	10	Philippines	Mathematics	36	G6	.3431854	.1003466	Regression co-efficient		2406

(2013)								formula		
Yamauchi and Liu (2012)	11	Philippines	Mathematics	24	G6	.2972309	.0782506	Regression co-efficient formula		3963
Bando (2010)	16	Mexico	Mathematics	12	5 year exposure	.0805095	.0301798	Means and standard deviations formula		114722
Beasley and Huillery (2014)	17	Niger	Mathematics	6	G1,4,6	-.047653	.0583424	Regression co-efficient formula		763
Blimpo and Evans (2011)	18	Gambia	Mathematics	36	WSD, G4,6	-.1839374	.122625	Regression co-efficient formula		814
Duflo et al. (2012)	19	Kenya	Mathematics	15	ETP + SBM + CT, G1	.237	.087	Regression co-efficient formula		6533
Lassibille et al. (2010)	21	Madagascar	Mathematics	15	G4	.0051	.01908	Means and standard errors	Use standard error of mean difference	Not reported
Parker (2005a)	23	Nicaragua	Mathematics	0	G3	.1076	.0768	Use t-test method, unequal samples	No SDs	686
Rodriguez et al. (2010)	24	Colombia	Mathematics	36	G5	-.019	.03742	Use t-test method, equal samples	No SDs	2762
Santibanez et al. (2014a)	26	Mexico	Mathematics	12	No previous PEC, G3	.2822623	.1469591	Regression co-efficient formula		3763
King and Ozler (2005b)	30	Nicaragua	Mathematics	0	T5 De jure , secondary	.205	.41	Regression co-efficient formula		911
Santibanez et al.	31	Mexico	Mathematics	12	No previous PEC, G6	.0252459	.1383908	Regression co-efficient		3575

(2014b)								formula		
Parker (2005b)	32	Nicaragua	Mathematics	0	G6	-.148	.0707	Use t-test method, unequal samples	No SDs	830
World Bank (2013)	10	Philippines	Aggregate Test Score	36	Includes maths, G6	.338888	.0936155	Regression co-efficient formula		2406
Yamauchi and Liu (2012)	11	Philippines	Aggregate Test Score	24	Includes maths, G6	.2871229	.0802709	Regression co-efficient formula		3693
San Antonio (2008)	25	Philippines	Aggregate Test Score	0		.120678	.0883825	Means and standard deviations formula		513
Yamauchi (2014)	27	Philippines	Aggregate Test Score	0		.3149995	.2708547	Regression co-efficient formula		2453
Bold et al. (2013)	28	Kenya	Aggregate Test Score	17		.057	.09	Regression co-efficient formula		Not reported
Barr et al. (2012)	13	Uganda	Teacher Attendance	24		.1717	.0866	Regression co-efficient formula		534
Beasley and Huillery (2014)	17	Niger	Teacher Attendance	12		-.1319695	.0784216	Regression co-efficient formula		799
Duflo et al. (2012)	19	Kenya	Teacher Attendance	15	ETP + SBM + CT	.2558062	.0715157	Regression co-efficient formula		2240
Sawada and Ragatz (2005)	8	El Salvador	Teacher Absenteeism	0	Recoded to reverse sign	-.5997492	.3224458	Regression co-efficient formula		134
World Bank (2011)	9	Sri Lanka	Teacher Absenteeism	30	Recoded to reverse sign, PSI only	.5186579	.3509268	Regression co-efficient formula		384
Blimpo and Evans	18	Gambia	Teacher Absenteeism	36	Recoded to reverse sign,	-.2147	.12124	Regression co-efficient		274

(2011)					WSD			formula		
Lassibille et al. (2010)	21	Madagascar	Teacher Absenteeism	21	Recoded to reverse sign	-.034	.08137	Means and standard errors	Use standard error of mean difference	2024

Supplement 3: Details of included non-causal studies

Citation	Region	Country	Intervention	Publication Status
Bandur (2008)	South-east Asia	Indonesia	SBM reform	PhD thesis
Bjork (2003)	South-east Asia	Indonesia	SBM reform	Journal Article
de Umanzor et al. (1997)	Latin America	El Salvador	EDUCO	World Bank Working Paper
Fuller & Rivarola (1998)	Latin America	Nicaragua	Autonomous Schools	World Bank Working Paper
Gershberg & Meade (2005)	Latin America	Nicaragua	Autonomous Schools	Journal Article
Gunnarsson et al. (2008)	Latin America	Multi-country	N/A	Working Paper
Hanushek et al. (2011)	Multi-region	Multi-country	N/A	NBER Working Paper
Reimers & Cardenas (2007)	Latin America	Mexico	PEC	Journal Article
Vernez et al. (2012)	South-east Asia	Indonesia	SBM reform	Monograph