



The Effects of School-Based Decision Making on Educational Outcomes in Low and Middle Income Contexts: A Systematic Review

Roy Carr-Hill, Caine Rolleston, Tejendra Pherali and Rebecca Schendel

Submitted to the Coordinating Group of:

<input type="checkbox"/>	Crime and Justice
<input checked="" type="checkbox"/>	Education
<input type="checkbox"/>	Disability
<input checked="" type="checkbox"/>	International Development
<input type="checkbox"/>	Nutrition
<input type="checkbox"/>	Social Welfare
<input type="checkbox"/>	Other:

Plans to co-register:

<input type="checkbox"/>	No		
<input checked="" type="checkbox"/>	Yes	<input type="checkbox"/> Cochrane	<input checked="" type="checkbox"/> Other IDCG
<input type="checkbox"/>	Maybe		

Date Submitted: April 14, 2014

Date Revision Submitted: June 5, 2014

Approval Date: August 13, 2014

Publication Date: September 1, 2014

Note: Campbell Collaboration Systematic Review Protocol Template version date: 24 February 2013

Contents

Background	3
Objectives	8
Methodology	9
References	27
Sources of support	32
Declarations of interest	32
Review authors	32
Roles and responsibilities	33
Preliminary timeframe	35
Plans for updating the review	35
Authors' responsibilities	35
Publication in the Campbell Library	35
Annex 1: Included country list	37
Annex 2: Coding for data extraction (review question 1)	38
Annex 3: Risk of bias coding (review question 1)	40
Annex 4: Coding for data extraction (review question 2)	42
Annex 5: Coding for quality appraisal (review question 2)	44

BACKGROUND

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, White, & Carpenter, 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).

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 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, Patrinos, & Rubio-

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.

Codina, 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 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’ includes 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, we anticipate that the available evidence will relate to the three main mechanisms outlined above: (1) devolving decision-making around *management* to the school level; (2) devolving decision-making around *funding* to the school level; and (3) devolving decision-making around *curriculum, pedagogy and other aspects of the classroom environment* to the school level.

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.

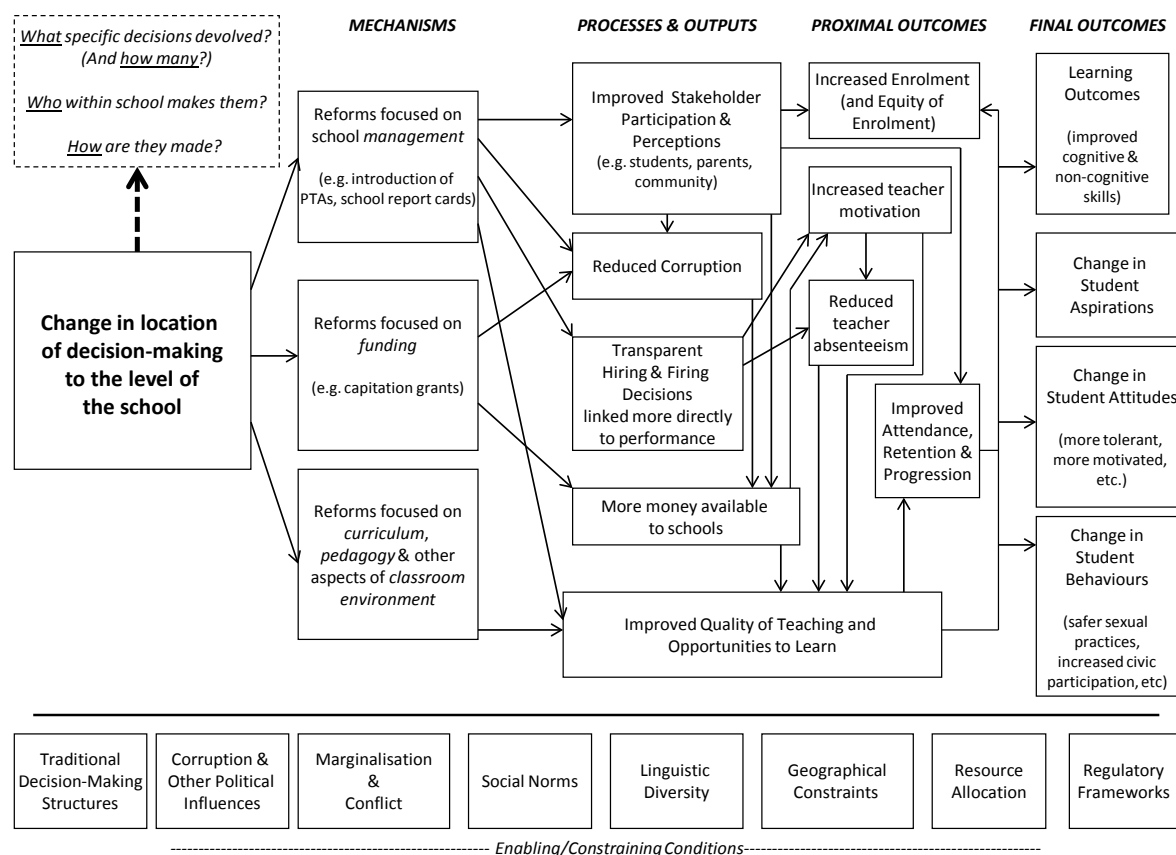
However, it is clear that devolving decision-making to the level of the school does not lead *directly* to such outcomes. Rather, school-based decision-making is likely to impact on outcomes via a number of causal pathways. Reforms that increase accountability and responsiveness to local needs are assumed 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 cost of educational provision, a proximal outcome highly 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.

Furthermore, 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, Hopkins, Lintott, & Riddell, 1999; Condy, 1998; Glassman, Naidoo & Wood, 2007; Pherali, Smith & Vaux, 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). 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. In addition to the ways in which enabling or constraining conditions and circumstances can alter the outcomes of school-based decision-making reforms, it is clear that differences in *implementation* can also affect outcomes. Those vested with the authority to make decisions on behalf of the school must have the capacity and knowledge to make such decisions, or their decisions are unlikely to have a positive impact on outcomes (World Bank, 2004). 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 and (b) the incentives provided will positively impact student learning. 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, Ilias, & Kremer, 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.

Figure 1 (below) is a visual depiction of our understanding of the causal pathways, contributing factors and underlying processes that appear to affect the impact of school-based decision-making on educational outcomes. Our conceptual framework is not presented here as a definitive map of the existing evidence. Rather, it is proposed as a ‘working hypothesis’ to help guide the implementation of this review (Oliver, Dickson & Newman, 2012, p. 68). As such, we have used the framework to generate specific review questions and define our review methodology (as recommended by Anderson et al., 2011). We plan to significantly revise, modify and potentially simplify (or disaggregate) the framework during the review process, in order to more accurately reflect the current body of evidence related to school-based decision-making in lower-income contexts. This may include articulating separate theories of change for some of the individual mechanisms, depending on the evidence available.

Figure 1: Conceptual Framework



Source: Original

Why it is Important to do the 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 (Barrera-Osorio & Linden, 2009; Jimenez & Sawada, 2003; Sawada & Ragatz, 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, Leon, Zapata, Sugimaru, & Cueto, 2012, on teacher absenteeism; Petrosino, Morgan, Fronius,

Tanner-Smith, & Boruch, 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, Filmer & Patrinos, 2012, on accountability reforms), rather than considering the full range of school-based decision-making models. The comprehensive reviews that do exist (e.g. Santibanez, 2007; World Bank, 2007) need updating, as they (a) rely on literature that is now nearly ten years out of date, (b) focus almost exclusively on Central America, referencing almost no evidence from other low- or middle-income countries, and (c) do not report the use of systematic searches, critical appraisal and statistical synthesis of study effect sizes. 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.

Furthermore, existing reviews on this topic tell us almost nothing about *why* school-based decision-making has positive or negative effects in different circumstances. The exclusive focus on evidence collected through impact evaluations and quasi-experimental designs has significantly limited the policy relevance of these reviews as this approach has (a) resulted in a very small (<60) number of studies and (b) prevented any analysis of the conditions and circumstances under which school-based decision-making models can have a positive impact.

We anticipate that the outcomes of this review will be useful for a wide range of stakeholders. In particular, policy-makers, at both the national and supranational levels, will benefit from the evidence linking decentralised decision-making processes to a wide range of potential outcomes and the analysis of underlying conditions that affect impact. 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. The timing of this review will 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/>).

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 (question 1 above). This objective will be accomplished by examining the results of causal studies (e.g. 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. Such analysis will allow us to report on all relevant quantitative measures of educational outcomes. Although we recognise that focusing on quantitative studies may preclude our ability to discuss outcomes usually considered harder-to-measure, we anticipate that the results will be useful, both for illuminating the ways in which school-based decision-making models do impact outcomes and for highlighting the current gaps in the evidence base. We also aim 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 – effective outcomes (question 2 above). This objective will be accomplished 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 question 1. Given the broader scope of this second review question, studies do not need to be causal in nature in order to be included.

In addition to examining the overall (positive and negative) effects of decentralisation processes on outcomes, we aim within this review to examine how changes in decision-making processes might impact differentially on diverse groups within societies. We are particularly concerned with groups which have historically experienced poor service delivery and/or demonstrated poor educational outcomes (e.g. marginalised or low-performing students). This will be accomplished by examining: (1) whether the interventions outlined in the included studies specifically target particular populations and (2) whether the included studies report any sub-group analysis for such populations.

METHODOLOGY

These objectives will be accomplished through the implementation of a high quality systematic review, relying on existing methodological guidance from the Campbell Collaboration and the EPPI-Centre at the Institute of Education (e.g. Becker et al., undated; Gough, Oliver & Thomas, 2012; Hammerstrom, 2009; Shadish & Myers, 2004).

As this review aims 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 have elected to conduct a mixed methods review, following the guidelines developed by Snilstveit (2012) for ‘effectiveness plus’ systematic reviews in international development. As such, we will use our conceptual framework 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 will also consider 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 will need 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, will be reviewed in response to the second sub-question, as other methods are likely to be particularly useful for clarifying which external conditions and/or implementation factors may substantially affect outcomes.

Criteria for including and excluding studies

Studies will be included in the review if they meet the following selection criteria.

Types of Participants/Settings

We will be looking exclusively at evidence related to primary and secondary schools in LMICS. Studies of both public and private sector provision will be included. In order to be included, studies must 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 will exclude evidence collected in LMICs located within Central and Eastern Europe or the former USSR.

Types of Interventions

We have defined ‘school-based decision-making’ quite broadly for the purposes of this review. There were two reasons for this decision: (1) As impact evaluation has been used only sparingly in the literature, we felt it important to use a broad definition in order to capture adequate breadth of literature to respond to the review questions; and (2) By constraining our search to only particular models of school-based decision-making, we thought it likely that we would miss potentially common features across models which may be found to have a significant impact on particular outcomes. Given the need for breadth, we have elected to include any study that empirically investigates an intervention utilising at least one of the three school-based decision-making mechanisms outlined in the conceptual framework (i.e. school management reforms, funding reforms, or curricular/pedagogical reforms). This is likely to include a long list of particular interventions, such as school management committees, school ‘report cards’, and capitation grants (both school and individual). An exhaustive list of intervention models has not been developed *a priori*, so as to allow for the broader possible range of potentially includable studies.

In reference to the first review question, we are likely to find comparisons 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. We may also find comparisons between groups in which *different* school-based decision-making reforms have been attempted (e.g. funding reforms versus school management reforms). Both will be included, although they will be distinguished from one another during synthesis. Comparisons must be contemporaneous (i.e. in comparisons between interventions, the interventions must have been implemented during the same time period, and, in comparisons between a reform group and a non-reform group, data must reflect the same time period).

Comparison groups are not a prerequisite for inclusion in relation to the second review question.

Types of Outcome Measures

As school-based models of decision-making can yield a wide range of outcomes (both positive and negative), we will not be excluding studies on the basis of a pre-determined list of outcomes. However, for inclusion in reference to both review questions, studies must empirically investigate 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).

Types of Study Designs

In order to be included for synthesis in relation to the first review question, studies must rely on an explicit comparison or adopt an appropriate empirical strategy to identify causal effects. Eligible study designs for inclusion in this phase of the review include:

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 including of interrupted time-series with at least 3 data collection points both before and after the intervention²).

Studies are eligible which analyse data at the level of the child or at community or sub-national (e.g. district) level, as well as the level of the school. Studies reporting analyses based on these different methods and levels of data will be separated in the synthesis (see below). Studies will be excluded in relation to this question which do not present quantitative information on proximal or final outcomes, or which present comparison groups at country level or higher. Given the wide diversity of studies likely to be included in the review, we will assess the validity of all included studies prior to synthesis using risk of bias categories (see below).

Non-causal studies are likely to be of use in responding to the second review question, so we will include a broader range of empirical study designs for synthesis in relation to the second question. Eligible study designs for this component of the review include:

1. Process evaluations and/or project completion reports of any of the school-based decision-making mechanisms evaluated in reference to the first review question
2. Empirical studies (employing quantitative, qualitative or mixed methods of analysis) which report evidence of change, as a result of a school-based decision-making reform, on at least one educational outcome (either proximal or final) and which offer primary data on either (1) factors which have been found to affect the implementation of the school-based decision-making mechanism or (2) conditions/circumstances which have been found to affect the impact of the mechanism on the outcome(s).

Included studies in reference to the second review question will need 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) in order to be included for synthesis.

Duration of Follow-Up

Studies of any follow-up duration and studies with multiple follow-ups will be included. However, during coding, the specific time-lag will be captured for each included study, so that we can consider temporal differences that are likely to affect synthesis.

² Cochrane Effective Practice and Organisation of Care Group (2014).

Other Exclusion Criteria

Language: As members of our team are fluent in English, French, Spanish and Portuguese, we intend to include studies written in any of these languages. Studies written in other languages will be excluded, unless English translations are available.

Publication Status: We will include journal articles, books, conference papers and institutional grey literature, including reports and process evaluations, in the review. Unpublished papers showing empirical evidence (such as dissertations and theses, empirical studies showing null and/or negative results and the like) will also be included.

Search Strategy

Our search strategy involves 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 that are likely to contain information relevant to the review
3. Hand-searching of relevant journals
4. Contacting experts involved in research on school-based management
5. Reference snow-balling

The first four search strategies will be conducted at the start of the review process.

Existing systematic reviews will first be identified through the 3ie Database of Systematic Reviews, the EPPI-Centre Database of Education Research, and the Campbell Collaboration Library. The reference lists for any potentially relevant reviews will be screened for potentially includable studies.

We will then conduct detailed searches, with the support of our colleagues at the EPPI Centre, in the following databases and websites:

Multidisciplinary databases and catalogues

- ASSIA (Applied Social Science Index and Abstracts)
- IBSS (International Bibliography of the Social Sciences)
- IDEAS RePEc (Research Papers in Economics)
- SIGLE (Open Grey)
- UNBISNET (United Nations Bibliographic Information System)

Education databases

- AEI (Australian Education Index)
- BEI (British Education Index)
- CREATE (Consortium for Research on Educational Access, Transitions and Equity)

- ERIC (Education Resources Information Centre)
- IEA (International Association for the Evaluation of Educational Achievement)
- Review and Advice on Education and Training (NORRAG)

Other disciplinary databases in the social sciences

- EconLit (American Economic Association)
- LaborDoc (International Labor Organisation)
- National Bureau for Economic Research (NBER)
- Policy Pointers

Bibliographic databases and catalogues specifically related to international development

- Africa Bibliography
- Africa Women Bibliographic Database
- Africa-Wide (EBSCO)
- AILISS Africa (Internet Library Sub-Saharan Africa)
- AJOL (African Journals Online)
- Asia Journals Online
- BLDS (British Library of Development Studies)
- Connecting Africa
- IDRIS (International Development Research Centre Development Research Information System)
- International Development Abstracts
- LAMJOL (Latin American Journals Online)
- Network for Policy Research
- Quarterly Index of Africa Periodical Literature

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)
- ADEA (Association for the Development of Education in Africa)
- African Development Bank Evaluation Reports
- Asian Development Bank Evaluation Reports
- CEGA (Centre for Effective Global Action)
- Center for Population Development and Activities
- DEREc (DAC Evaluation Resource Centre)
- DFID (Research for Development)
- DIME (Development Impact Evaluation Initiative)
- GDN (Global Development Network)
- Global Partnership for Education
- Inter-American Development Bank Evaluation Reports
- IE2 Impact Evaluation Repository (World Bank)
- IIEP (International Institute of Educational Planning)
- IISD (International Institute for Sustainable Development)

- IPA (Yale University Innovations for Poverty Action Center)
- JOLIS (World Bank and IMF Library Catalogue)
- OECD (Organisation for Economic Co-Operation and Development library)
- SACMEQ (Southern and East African Consortium for Monitoring Educational Quality)
- SIDA (Swedish International Development Agency: Unit for Research Cooperation)
- UNESCdoc (United Nations Educational, Scientific and Cultural Organisation)
- USAID (Development Experience Clearinghouse)

These resources have been selected because they are likely to yield evidence that is relevant to the review questions while also representing a wide range of disciplinary perspectives. We have also made an explicit effort to include resources that are likely to help us identify grey literature and literature published within LMIC contexts.

In addition, we will search 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*. We will also reach 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 studies that have been completed but are not yet published.

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 will therefore limit electronic searches and journal hand searches to 2000 onwards. However, we will search reference lists of existing literature reviews (e.g. Santibanez, 2007 and World Bank, 2007) and systematic reviews (e.g. Petrosino et al., 2012) to identify additional relevant literature, including studies published before 2000.

Once the initial search has been completed, all potential titles and abstracts will be imported into EPPI-Reviewer, and a duplicate check will be completed. We will then begin the process of screening and coding studies (described in more detail below). Once we have decided on our list of studies for quality appraisal, we will complete our final search strategy by checking the reference lists of all included studies – and consulting the Web of Science, Google Scholar and Scopus to track citations of our included studies – in order to identify any key sources that we might have missed during the initial search. If any such sources are identified, they will be included prior to quality appraisal.

Our search rests on two main ‘concepts’, each of which consists of a large number of potential search terms:

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

Relying on the expertise of the EPPI Centre, we have assembled a list of controlled terms which tend to be used in the main electronic databases in reference to Concept2. The list of search terms involved in Concept 1 has been 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.

Our final search will be conducted using the following search concepts/terms. In order to be captured by the search, studies must reference at least one term from each concept in either the title or the abstract. In databases allowing for complex Boolean searches, controlled terms will also be searched (designated as SU terms in the search concept below). These terms vary by database, so the terms included here will act as stem terms in each database's individual thesaurus. All potential relevant thesaurus terms will be searched, and a separate record of controlled terms used will be saved for each database.

Concept 1 = School-based decision making

TI, AB(((decentral* OR devolv* OR governance) n2 school) OR ((decentral* OR devolv* OR governance) n2 education) OR ("school based management" OR SBM) OR "shared decision making" OR "school management committee*" OR (accountability n2 school) OR (accountability n2 education) OR "school report cards" OR "principal leadership" OR "School level planning" OR "school autonomy" OR "parent-teacher association" OR ("community participation" n2 school) OR ("community participation" n2 education) OR "community based management" OR ((decentral* OR devolv* OR governance) n2 budget*) OR ("resource allocation" n2 school) OR (resource allocation" n2 education) OR ("capitation grant*") OR ("block grant*" n2 school) OR ("block grant*" n2 education) OR ((decentral* OR devolv* OR governance) n2 curriculum) OR ((decentral* OR devolv* OR governance) n2 pedagog*) OR "contract teachers" OR "supply teachers" OR (curriculum n2 local) OR (pedagog* n2 local) OR "teacher allocation" OR "teacher distribution") OR SU("decentralization/decentralisation" OR "School Based Management" OR "School Autonomy" OR "School Management" OR "Principal Leadership" OR "Educational/School Accountability" OR "School Boards" OR "School Data/Statistics" OR "School Funding" OR "Teacher Distribution" OR "Teacher Allocation" OR "School Report Cards")

--
AND

Concept 2 = Low-or Middle-Income Countries

(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)); This term list will also include the names of all LMICs outside of Europe and Central Asia. A full list of the included countries is attached as Annex 1 to this protocol.

Note: When searching in databases and websites that are likely to include a number of French, Spanish or Portuguese sources, we will also include a number of core terms in these languages (e.g. "gestión escolar autónoma").

Selection of Studies

There will be two screening phases in the review: (1) Screening on Title and Abstract, and (2) Screening on Full Text. EPPI-Reviewer will be used throughout the review to manage and record the review process. This specialist software maintains a detailed search log of every decision made during the importing, screening and coding phases, allowing for future replication of the review process.

During both screening phases, studies will be reviewed and assessed against the review's inclusion/exclusion criteria (outlined above). As we anticipate a very large number of potential studies, we will not be able to double-screen every study. However, a moderation exercise will be conducted at the start of screening to allow for a discussion of decisions between individual team members and to resolve any inconsistencies. We will also double-screen a random sample of 10% of the total studies during each phase, utilising the EPPI-Reviewer functionality for resolving conflicting coding.

Description of Methods Used in Primary Research

We anticipate that our search strategy will identify primary studies such as the following:

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.

This study employed an experimental design to test the impact of report cards on test scores, private school fees, enrolment, and household-level and school-level educational

investments. The study was conducted in 112 villages in rural Pakistan, with over 800 public and private primary schools. The villages were randomly selected from a list frame of villages with at least one private school. Grade 3 children in all primary schools were tested, and then report cards were disseminated in a randomly selected 50% of the villages which contained the results of school and child test scores for all schools and tested children. Further data was gathered to assess the impact of this intervention through school and household surveys, along with a repeat of the child-level tests. This variety of study would be classified as relevant for inclusion in relation to the first review question, as it presents evidence of impact of a management reform (the provision of information to parents through school report cards) on both proximal (e.g. enrolment, private school fees and educational investment) and final (e.g. test scores) outcomes.

Eskeland, G., & Filmer, D. (2007). Autonomy, participation, and learning in Argentine schools: Findings, and their implications for decentralization. *Education Economics* 15 (1), 103–127.

This paper analyses national education data from a random cross-section of schools across Argentina. The data set includes language and maths test scores for children in the 6th and 7th grades, linked to student socio-economic background, and standardised indices for 'school autonomy' and 'parent/community participation', based on responses to questionnaires distributed to teachers and principals. Instrumental variables are used to address confounding in analysis. School autonomy was found to have a positive impact on test scores (although not at schools with low levels of participation), while participation was found to have an impact on test scores in schools with high levels of autonomy. Such a study is also relevant for inclusion in relation to the first review question, as it presents evidence of impact of a management reform on test scores (a 'final' educational outcome).

World Bank. (2010). *Project performance assessment report: Nepal, Community School Support Project* (CR. 3808). Washington, DC: World Bank.

This evaluation report summarises the findings of project files (particularly supervision reports) and a number of implementation reports related to the Community School Support Project in Nepal. It also presents data collected by a team of independent evaluators who completed direct observation in recipient schools and conducted qualitative interviews with officials, donors and beneficiaries of the project. The Community School Support Project financially supported social mobilisation and grants for school committees. Funded committees were then expected to raise additional funds, enrol out-of-school children in their communities, and be accountable to other school committees, such as parent-teacher associations and monitoring committees. The intended outcomes of the intervention are largely proximal outcomes, including increased enrolment, reduced teacher absenteeism and improved 'quality' of provision (assessed largely through changes in classroom infrastructure). The report concludes that the project was not successful in achieving its aims. The evidence suggests that the project suffered from fiduciary challenges and

resistance by teachers' unions. Deficiencies in data collection throughout the life of the project also limited the conclusions that could be drawn around the effectiveness of the intervention on the desired outcomes. Although this limitation would prevent the results of this report from being considered in relation to the first review question, the evidence presented in relation to the implementation process would be relevant for inclusion in relation to the second.

Details of study coding categories

All included studies will be coded on a number of dimensions, as suggested by the conceptual framework. The code list is attached as Annex 2. A second moderation exercise will be conducted with all participating team members prior to initial coding, and a random sample of 10% of included studies will be double-coded to check for coding reliability between team members.

At the end of the initial coding phase, the included studies will be disaggregated into two groups: (1) studies appropriate for answering the first review question, and (2) studies appropriate for answering the second review question.

Assessment of methodological quality and risk of bias

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

Review Question 1

Studies included in reference to the first review question (i.e. all 'effectiveness' studies) will be designated as being of either 'low', 'medium' or 'high' risk of bias, using the coding criteria outlined in Annex 3. All of the 'effectiveness' studies will be double-coded by two members of the review team, and values for Cohen's kappa will be reported to demonstrate inter-rater reliability. Classifications will then be made based on the study coding.

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

- 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 will be classified as at 'medium risk of bias' if either:

- a) There are moderate threats to the validity of the attribution methodology (arising from issues with the implementation of the methodology), or

- b) 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
- c) There are possible reporting biases.

All other studies will be classified as 'high risk of bias studies'. This will, therefore, include:

- a) 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
- b) Where there is clear evidence of spillovers or contamination to comparison groups from the same communities; or
- c) Where reporting biases are evident.

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

Review Question 2

Studies included in reference to the second review question (including any effectiveness studies classified as high risk of bias) will be coded for quality appraisal using the code list included as Annex 4.³ A final moderation exercise will be conducted with all participating team members prior to coding, and a random sample of 10% of included studies will be double-coded to check for coding reliability. Cohen's kappa values will be reported to demonstrate inter-rater reliability. Studies will then be classified as 'high', 'medium' or 'low' quality studies, based on this round of coding.

'High' quality studies will have received a 'High Quality' code for each of the dimensions assessed. 'Medium' quality studies will 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). Any study receiving at least one 'Low Quality' code will be classified as 'low' quality.

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

Criteria for determination of independent findings

³ 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.

In relation to review question 1 concerning intervention impact/effectiveness, selected studies are expected to report impact estimates in relation to various educational outcomes. Each study may contribute estimates on one or more outcomes, including for example estimates of impact on one or more domain (e.g. school subject) in terms of learning outcomes. Further, studies may contribute estimates for more than one time period.

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

- Where a study reports effect sizes relating to a particular intervention on more than one outcome/domain we will include 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 will include only one estimate except in the case that the study is implemented across more than one non-overlapping and independent sample (being effectively independent studies), when one effect will be included separately for each sample. The selected effect will be the effect assessed as having the lowest risk of bias in attributing impact.
- For each independent sample, only one estimate will be 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.
- Where a study reports more than one effect size for a particular outcome/domain by sub-group, where possible we will combine groups to calculate a synthetic weighted-average effect size for the entire sample included in the study, making appropriate adjustments to variances and standard errors and making necessary covariance assumptions.
- It is possible that more than one paper may analyse and report the results of the same intervention/programme using different methods and specifications but the same or a similar sample, leading to dependent results. In such cases, we will treat these papers in a way equivalent to a single study reporting multiple effect sizes as above.
- If possible based on the selected studies, we may provide pooled estimates for selected sub-groups separately, for example by gender, as an extension.

Data Extraction

Evidence extracted from the selected impact/effectiveness studies will be used to answer our first review question, and evidence from the other included studies will be used to answer our second review question.

Review Question 1

For each included study in reference to the first review question, we will extract data on the study setting, participants, methods, details of the ‘intervention’, comparison conditions, and outcomes (Annex 2), as well as the study’s risk of bias classification (e.g. ‘High’ or ‘Medium’) (Annex 3). We will also extract any reported treatment estimates in order to calculate effect sizes (including the direction and magnitude of the effect and any reported sub-group effects) and confidence intervals, and we will report the results of power analysis for each study. Computation procedures will also be extracted at this stage.

Review Question 2

Data will be extracted from the studies included in reference to the second review question with the aid of the coding list (Annex 4). These codes were generated by the research team based on our existing knowledge of themes in the decentralisation literature, but we also intend to use emergent coding if new themes can be identified in the included studies.

As with the studies included for review question 1, data will be extracted pertaining to the study setting, participants, methods, details of the ‘intervention’, outcomes, and the study’s quality classification (‘High’ or ‘Medium’) (Annex 5). In addition, data will be extracted pertaining to the intervention’s implementation process and the conditions/context in which the intervention was implemented. The extracted data will be organised in two matrices to aid synthesis, one outlining data pertaining to implementation factors and one outlining data pertaining to conditions and circumstances mediating impact.

Codes identified a posteriori will be noted as such in the report.

Statistical procedures and conventions

In relation to review question 1, we expect that for the most part, outcome variables (particularly test-scores) reported in the studies will be continuous measures. However, some outcomes are expected to be reported as dichotomous or categorical variables and among the proximal and final outcome measures there is likely to be some diversity of measures. For each intervention-type and outcome (i.e. for each intervention-outcome pair), we will select appropriate outcome variable types for use in pooling estimates based on the most common type found in the studies. Where studies reporting effects on a particular outcome employ different types of variables, appropriate methods will be employed to convert measures, for example to convert categorical into binary measures. These methods will be reported in the review.

Measurement of treatment effects

Reported effect sizes will be transformed into standardised mean differences (making appropriate statistical adjustments for small sample sizes where appropriate) for outcomes where continuous measures are employed and into odds or risk ratios where dichotomous or categorical measures are used, reporting standard errors and 95% confidence intervals in each case (as suggested by Keef and Roberts, 2004; Borenstein et al., 2009). Appropriate statistical transformations will be applied to calculate standardised mean differences and odds or risk ratios from matching and regression studies. We will discuss the interpretation of effect size measures (sign and magnitude) and their interpretation in relation to each intervention-outcome pair. If appropriate based on the results, we will translate odds/risk ratios into standardised mean differences in order to put all effect estimates onto the same metric to allow for greater comparability.

Missing data issues

Where required owing to the absence of full information in selected studies for the calculation of standardised mean differences or odds/risk ratios, appropriate formulae will be employed to extract or impute effect sizes from statistics reported (such as t, z or F statistics and p values) in the study. Where data and statistics reported in a study are inadequate to compute effect sizes on this basis, we will attempt to contact the authors of the studies to gather missing data and statistics. In the event that authors are unable to provide this information, we will calculate response ratios which require more limited information, where possible and where appropriate based on the type of outcome data. An assessment will be made of the likely effect of any adjustments made for missing data and statistics on the synthesis and will be reported in relation to each outcome.

Unit of analysis issues

For each study, an assessment of the likelihood of unit of analysis error will be conducted. Such error may be present where, for example, a decentralisation intervention takes place to shift decision-making power from districts to schools and where the measure of impact is based on pupil-level test scores in selected schools in districts in receipt of the intervention, when 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). This is likely to be common among the selected studies. Data clustering at school and district level needs to 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, since pupils within clusters are likely to be more homogenous than across clusters and hence pupil-level observations are not fully independent. In such cases, the effective sample size is reduced when compared with a simple random sample (e.g. pupil-level random assignment), depending on the level of intra-cluster correlation. Our assessment will examine whether studies have employed appropriate statistical methods to account for data clustering, such as the use of cluster fixed effects and robust standard errors. Where appropriate methods have not been employed, where possible, we will apply corrections to standard errors and confidence intervals using available formulae (and if

required using estimates of intra-cluster correlation for relevantly similar studies) to minimise the likelihood of type I error (as suggested by Higgins and Green, 2011).

Synthesis

For the first review question, we will synthesise evidence from the included studies to assess whether or not the identified interventions appear to have an impact on educational outcomes, both proximal and final. This will present a synthesis of effects across the causal chain from interventions to final outcomes.

We will begin the synthesis process by providing a summary table of all included effectiveness studies (Table 1), broken down by pairs of intervention type and outcome measure and identifying the studies included in each intervention-outcome pair.

For each pair, effect sizes will be reported using forest plots and synthesised using inverse-variance weighted meta-analysis implemented in STATA software (Stata Corporation, College Station, TX, USA). Evidence from experiments and quasi-experiments will be analysed separately from studies with an explicit comparison group and those involving other methods of identifying a counterfactual (e.g. propensity score matching or instrumental variables). Effect sizes based on bivariate and multivariate methods will also be analysed independently. If no empirical differences are found between (some of) these groups, effect sizes will be pooled across the studies identified in the corresponding study designs. Random effects meta-analysis will be employed to account for contextual heterogeneity, given that the range of studies is likely to reflect a wide diversity of contexts. This approach produces estimates of pooled effect sizes with wider confidence intervals than a fixed effects model, assuming that studies are not estimating identical treatment effects (e.g. different groups may respond differently) but that these vary around a central value (mean).

Narrative assessment of synthesis results will also be provided for each intervention-outcome pair, taking account of risk of bias assessments, sample size, magnitude of effect and power analysis (as recommended by Waddington et al., 2012), as well as the results of sensitivity analysis and sub-group analysis where available (see below). Other relevant contextual factors, including individual study conditions (such as differences in the time lag between the adoption of school-based decision-making and measured effects between studies), will also be considered.

Assessment of Heterogeneity

We will test for heterogeneity across studies within each intervention-outcome pair category using the I-squared statistic, reporting the tau-squared statistic for random-effects meta-analysis.

Investigation of Heterogeneity

If appropriate, we will investigate heterogeneity further using moderator analysis (sub-group meta-analysis and meta-regression). If this is not possible, we will narratively explore factors that may explain heterogeneity. If it is possible to conduct moderator analysis, we will conduct this separately for different moderator values, including publication status, intervention process and implementation, and types of beneficiary, particularly with regard to gender and measures of socio-economic advantage. Given the diversity of mechanisms of decentralisation and likely differences in beneficiary populations and contexts, we expect heterogeneity in effects linked to process and implementation and to the level of disadvantage of beneficiaries. Evidence of heterogeneity resulting from implementation factors and/or enabling and constraining conditions will be discussed further in relation to the second review question.

Sensitivity Analysis

Moreover, even where sub-group analysis is limited by the number of available studies, we will report sub-group analysis to examine the sensitivity of synthesis results, where possible, to examine the extent to which synthesis results may be driven by differences in types of studies, treatment effect types, duration of follow-up, types of effect-size estimates, exclusion of low quality studies or the exclusion of outliers. The approach adopted will be to conduct meta-analysis excluding selected studies based on moderator values (as above).

Depending on the degree of diversity or heterogeneity among studies in each intervention-outcome pair category (in terms of study population, intervention, implementation or methodology, for example), it may not be possible to conduct statistical meta-analysis for each category. Where it is not appropriate to conduct meta-analysis, we will present forest plots with individual effect sizes (and statistical power analysis) to give an overall picture of the evidence, without reporting a combined effect, and synthesis will be conducted on a qualitative basis using narrative synthesis.

Assessment of Publication Bias

We will examine the possible effects of publication bias through sub-group analysis of published and unpublished studies (as above), as well as through use of funnel graphs where there are 10 or more studies included and statistical testing (Egger, 1997).

Treatment of Qualitative Research

Synthesis in reference to the second review question cannot be so specifically predetermined, as it is unclear what kind of evidence will be collected through our search. We anticipate that we will identify a number of qualitative studies that investigate the contextual factors that affect the impact of school-based decision-making models in particular circumstances, but we also hope to identify process evaluations and other mixed methods studies that elaborate how the implementation of school-based management models affects outcomes.

The evidence outlined in the synthesis matrices (outlined above) will first be mapped onto the conceptual framework, in order to examine the degree to which the available body of evidence reflects the various assumed causal pathways (as suggested by Snilstveit, 2012). Evidence along each pathway will then be compared and contrasted, in order to identify recurrent and/or divergent themes in the literature. This portion of the analysis will broadly follow the principles of framework synthesis (Thomas, Harden, & Newman, 2012).

Once the evidence relating to each causal pathway has been synthesised, we will use narrative analysis to aggregate the findings, highlighting the particular implementation and contextual factors which seem to substantially affect the effectiveness of school-based decision-making models.

Integrated Synthesis

If particular implementation factors are found to be relevant during synthesis, we will then return to our initial synthesis (in relation to review question 1), in order to test for the impact of such factors on differences in effects (where data are available on those factors).

Results from the two phases of synthesis will then be compiled into a ‘Summary of Findings’ table, which will highlight: (1) the summary effect sizes pertaining to review question 1, (2) a summary of implementation factors found to substantially affect impact, (3) a list of conditions and circumstances that appear to drive positive or negative impact, and (4) an assessment of the quality of the available evidence.

The integrated data set will then be used to populate and revise our conceptual framework, using a narrative synthesis approach along the causal chain (as suggested by Noyes & Lewin, 2011). This phase will necessarily be both aggregative (as it combines the results obtained through the previous stages) and interpretive (as we will be making qualitative assessments of how the data elucidates our initial conceptual model).

At this stage, we will also complete an analysis of gaps in the existing literature.

Table 1: Example primary study summary table

Source	Study Setting & Participants	Comparison Conditions	Outcomes	Methods	Results (effect size, confidence interval, power)	‘Intervention’ Details	Contextual Factors	Quality & Design Effects
<i>Randomised Experiments and Quasi-Experiments</i>								
Author (DATE)	El Salvador	Program participants compared to similar individuals who are not participants. Formal model of program placement created to control for endogeneity.	Retention rates Repetition rates	Regression using instrumental variables	Significant positive impact found on the decision to remain in school beyond age three ($d = XX$; $p = XX$; power = 0.8) Students	<i>Which:</i> Personnel <i>Who:</i> Community Education Association <i>How:</i> Monitoring of teacher performance, as well as allocating school budget and	Not discussed	11 year time lag Sample = 311 randomly-selected primary schools High rates of internal and external emigration rates may

					in participating schools less likely to repeat grades than students in traditional schools, however, this effect is not statically significant	hiring/firing teachers		bias the results. Program placement likely endogenous. Quality = HIGH
--	--	--	--	--	--	------------------------	--	---

REFERENCES

Anderson, L.M., Petticrew, M., Rehfuss, 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.

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](https://doi.org/10.1596/1813-9450-4584). Last accessed April 14, 2014.

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 Research Working Paper (Impact Evaluation Series) 4836 (29). 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. Last accessed April 14, 2014.

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

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.

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., 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.

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.

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.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315, 629-634.

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.

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>. Last accessed April 14, 2014.

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. Last accessed April 14, 2014.

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).

Jimenez, E., & Sawada, Y. (2003). *Does community management help keep kids in schools? Evidence using panel data from El Salvador's EDUCO Program*. Discussion Paper F-236. Tokyo: Center for International Research on the Japanese Economy (CIRJE), University of Tokyo. DOI: 3a10.1086_2f674096.htm.

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). Available at: <http://www.3ieimpact.org/en/evaluation/working-papers/working-paper-20/>. Last accessed April 14, 2014.

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. Available at:

http://www.iiep.unesco.org/fileadmin/user_upload/Info_Services_Publications/pdf/2010/CL_AdG_Decentralization.pdf. Last accessed April 14, 2014.

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*. Available at: <http://cqim.cochrane.org/supplemental-handbook-guidance>. Last accessed April 14, 2014.

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.

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/>. Last accessed April 14, 2014.

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.uk/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.

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.libreriacyde.com/librospdf/DTAP-188.pdf>. Last accessed April 14, 2014.

Sawada, Y., & Ragatz, A. B. (2005). Decentralization of Education, Teacher Behavior and Outcomes. In E. Vegas (Ed.), *Incentives to Improve Teaching*. Washington, DC: World Bank.

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. Last accessed April 14, 2014.

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: Continuum.

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. Available at:

http://www.gse.pku.edu.cn/lib/gse_lib/edu-search/e_publication/e_pub/268950PAPER0WDR02004.pdf. Last accessed April 14, 2014.

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

http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079934475/547667-1145313948551/what_do_we_know_SBM.pdf. Last accessed April 14, 2014

World Bank. (2014). *Country and lending groups*. Washington, DC: World Bank. Available at: http://data.worldbank.org/about/country-classifications/country-and-lending-groups#Upper_middle_income. Last accessed April 14, 2014.

SOURCES OF SUPPORT

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

DECLARATIONS OF INTEREST

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. No team member has been involved in the development of any interventions likely to be included in the review scope, although Prof. Carr-Hill has been involved in evaluations of some interventions that we might include (e.g., a recent mid-term evaluation of the Fast Track Initiative, which included detailed discussion of the locus of decision-making).

As some of the individual team members have completed primary research into questions of decentralisation, we will structure the review process so that no reviewer screens, codes or appraises his or her own work.

REVIEW AUTHORS

Lead review author:

Name: Roy Carr-Hill

Title: Professor of Education in
Developing Countries

Affiliation: Institute of Education

Address: 20 Bedford Way

City, State, Province or County: London

Postal Code: WC1H 0AL

Country: UK

Phone: +44 (0)20 7612 6631

Email: roy.carr_hill@yahoo.com

Co-author(s):

Name: Caine Rolleston

Title: Lecturer in Education &
International Development

Affiliation: Institute of Education

Address: 20 Bedford Way

City, State, Province or County:
London

Postal Code: WC1H 0AL

Country: UK

Phone: +44 (0)20 3073 8314

Name: Tejendra Pherali

Title: Senior Lecturer in Education
& International Development

Affiliation: Institute of Education

Address: 20 Bedford Way

City, State, Province or County:
London

Postal Code: WC1H 0AL

Country: UK

Phone: +44 (0)20 7331 5201

Name: Rebecca Schendel

Title: Lecturer in Education &
International Development

Affiliation: Institute of Education

Address: 20 Bedford Way

City, State, Province or County:
London

Postal Code: WC1H 0AL

Country: UK

Phone: +44 (0)20 7612 6819

ROLES AND RESPONSIBILITIES

- Content:

Roy Carr-Hill, Team Leader for this review, brings thirty years of experience in social policy analysis to the team, having worked with a wide range of international agencies to evaluate the implementation of education programmes in a number of low- and middle-income countries. Prof. Carr-Hill has widespread knowledge of both quantitative and qualitative methodologies and brings particular expertise regarding resource allocation and problems of implementing policies in a decentralised system. He will provide the overall leadership for the project and direct the synthesis process. He will also be responsible for reviewing,

appraising and synthesising the evidence from included effectiveness studies and contributing to all review reports.

Caine Rolleston and Tejendra Pherali will act as secondary reviewers, contributing to all aspects of the review process. Having previously worked with both CREATE (the Consortium for Research on Educational Access Transitions and Equity) and Young Lives, an international study of childhood poverty based at the University of Oxford, Dr. Rolleston brings extensive experience with quantitative methods of analysis, as well as research on different administrative systems, to the team. Dr. Pherali's work focuses primarily on education in conflict and post-conflict settings, particularly in South and South-east Asia. Dr. Pherali has substantial experience assessing qualitative evidence and brings a unique perspective on the challenges of collecting and analysing data in conflict settings. Both Dr. Rolleston and Dr. Pherali will review/synthesise material in their areas of expertise and comment on all review reports.

- Systematic review methods:

Rebecca Schendel will act as the team's Review Manager, drawing on her recent experience as lead Research Officer on a DFID-commissioned rigorous review to provide day-to-day management of the review process.

We have also recruited four part-time Research Officers to assist with the screening and coding of studies. Dr. Schendel will manage these positions and work directly with the EPPI Centre at IOE during information retrieval. She will also coordinate each stage of the review process, ensuring consistency across team members; review, appraise and synthesise qualitative and mixed methods evidence; and support Prof. Carr-Hill in the writing of review reports. In addition to her experience managing rigorous reviews, Dr. Schendel has used both quantitative and qualitative methods in her own work on student learning outcomes in sub-Saharan Africa.

Dr. Sandy Oliver (of the EPPI Centre) will also advise the team on methodology and synthesis techniques, drawing on her extensive experience with a diverse range of theoretical frameworks, review methodologies and analytical techniques.

- Statistical analysis:

Both Prof. Carr-Hill and Dr. Rolleston have considerable background in quantitative methods, which will allow the team to successfully complete the secondary statistical analysis likely to feature in this review. We do not anticipate using meta-analysis in this study.

- Information retrieval:

Two Information Scientists from the EPPI Centre will work with the team to clarify search terms, identify relevant literature and retrieve abstracts for initial screening. Full text retrieval will be managed by the Research Officers.

PRELIMINARY TIMEFRAME

The scheduled date for submission of a first draft of the review findings is December 1, 2014, with a final deadline scheduled for February 15, 2015.

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.

AUTHORS' RESPONSIBILITIES

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

You accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group.

PUBLICATION IN THE CAMPBELL LIBRARY

The support of the Campbell Collaboration and the relevant Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review and subsequent updates in the Campbell Library. Concurrent publication in other journals is encouraged. However, a Campbell systematic review should be published either before, or at the same time as, its publication in other journals. Authors should not publish Campbell reviews in journals before they are ready for publication in the Campbell Library. Authors should remember to include a statement mentioning the published Campbell review in any non-Campbell publications of the review.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Rebecca Schendel

Date: April 14, 2014

ANNEX 1: INCLUDED COUNTRY LIST

Low-income economies (per capita income of \$1,035 or less)

Afghanistan	Gambia, The	Nepal
Bangladesh	Guinea	Niger
Benin	Guinea-Bissau	Rwanda
Burkina Faso	Haiti	Sierra Leone
Burundi	Kenya	Somalia
Cambodia	Korea, Dem Rep.	South Sudan
Central African Republic	Liberia	Tanzania
Chad	Madagascar	Togo
Comoros	Malawi	Uganda
Congo, Dem. Rep	Mali	Zimbabwe
Eritrea	Mozambique	
Ethiopia	Myanmar	

Lower-middle-income economies (per capita income of \$1,036 to \$4,085)

Bhutan	Lao PDR	Sri Lanka
Bolivia	Lesotho	Sudan
Cameroon	Mauritania	Swaziland
Cabo Verde	Micronesia, Fed. Sts.	Syrian Arab Republic
Congo, Rep.	Mongolia	Timor-Leste
Côte d'Ivoire	Morocco	Vanuatu
Djibouti	Nicaragua	Vietnam
Egypt, Arab Rep.	Nigeria	West Bank and Gaza
El Salvador	Pakistan	Yemen, Rep.
Ghana	Papua New Guinea	Zambia
Guatemala	Paraguay	
Guyana	Philippines	
Honduras	Samoa	
Indonesia	São Tomé and Príncipe	
India	Senegal	
Kiribati	Solomon Islands	

Upper-middle-income economies (per capita income of \$4,086 to \$12,615)

Angola	Fiji	Namibia
Algeria	Gabon	Palau
American Samoa	Grenada	Panama
Argentina	Iran, Islamic Rep.	Peru
Belize	Iraq	Seychelles
Botswana	Jamaica	South Africa
Brazil	Jordan	St. Lucia
China	Lebanon	St. Vincent/Grenadines
Colombia	Libya	Suriname
Costa Rica	Malaysia	Thailand
Cuba	Maldives	Tonga
Dominica	Marshall Islands	Tunisia
Dominican Republic	Mauritius	Tuvalu
Ecuador	Mexico	Venezuela, RB

ANNEX 2: CODING FOR DATA EXTRACTION (REVIEW QUESTION 1)

- **Regional Context (Latin America & Caribbean, Sub-Saharan Africa, North Africa, Middle East, West and South Asia, Southeast Asia, East Asia, Pacific)**
- **Country Context**
- **Level of Education (Primary, Secondary or Both)**
- **Intervention Mechanism**
 - o School-based management
 - o School-based finances
 - o School-based decisions re curriculum/pedagogy
- **Intervention Model**
 - o School-Based Management
 - o Community-Based Management
 - o School Management Committee
 - o School-Based Personnel Management
 - o Community-Based Personnel Management
 - o Contract Teachers
 - o Supply Teachers
 - o School Report Cards
 - o Open Data Reform
 - o Public-Private Partnership
 - o Individual Capitation Grants
 - o School Capitation Grants
 - o Vouchers
 - o Local curriculum reform
 - o Local pedagogical reform
- **Outcome**
 - o Proximal: Enrolment
 - o Proximal: Equity of Enrolment
 - o Proximal: Attendance
 - o Proximal: Retention
 - o Proximal: Progression
 - o Proximal: Quality of Teaching
 - o Proximal: Other Opportunities to Learn
 - o Final: Student Learning (cognitive)
 - o Final: Student Learning (non-cognitive)
 - o Final: Student Aspirations
 - o Final: Student Attitudes
 - o Final: Student Behaviours
- **Research Design**
 - o Experimental using random assignment
 - o Experimental using quasi-random assignment
 - o Quasi-Experimental: Regression Discontinuity Design
 - o Quasi-Experimental: 'Natural Experiment'
 - o Quasi-Experimental: Self-selection
 - o Non-Equivalent Comparison Group with Propensity Score Matching

- Non-Equivalent Comparison Group with Covariate Matching
- Non-Equivalent Comparison Group with Difference-in-Differences Analysis
- Non-Equivalent Comparison Group with Fixed Effects Regression
- Non-Equivalent Comparison Group with Instrumental Variables
- Non-Equivalent Comparison Group with other Regression Analysis
- Non-Equivalent Comparison Group with 6-period interrupted time series
- Other Quantitative (e.g. Cross-Sectional)
- Qualitative
- Mixed Methods
- Comparison group (if applicable)
 - Any versus no SBM reform
 - Different SBM reforms
- Unit of Analysis
 - Sub-national (e.g. District, Region)
 - Community
 - School
 - Individual
- Size of Sample (total number of participants, if applicable)
- Data Collection Method (*list to be populated by review team members, based on included studies*)
- Data Analysis Method (*list to be populated by review team members, based on included studies*)
- Time lag (time between implementation of school-based decision-making and reported outcomes)

Please note: EPPI-Reviewer will automatically capture additional publication data, such as author, publication date, and publication source.

ANNEX 3: RISK OF BIAS CODING (REVIEW QUESTION 1)⁴

- **Randomisation (if applicable)**
 - **Low Risk:** Evidence of randomisation
 - **High Risk:** Evidence of self-selection or allocation based on potentially confounding criteria
 - **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
 - **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

⁴ Based on Cochrane Effective Practice and Organisation of Care Group (2014) 'Suggested risk of bias criteria for EPOC reviews', with additional questions suggested by Hombrados and Waddington (2012) and He et al (2007).

- 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

ANNEX 4: CODING FOR DATA EXTRACTION (REVIEW QUESTION 2)

- Study Setting
 - Urban
 - Rural
 - City
 - Slum
 - Town
 - Village
- Study Participants
- Quality Classification
 - High Quality
 - Medium Quality
- Implementation Factors
 - Who makes decisions
 - Principal/head master
 - Principal/ headmaster in consultation with SBM committee chair
 - School management committee within a school
 - School management committee including community members
 - Regional/ district level education officers
 - Central government
 - How decisions made
 - Top-down management within school
 - Management within school but including teachers
 - Management within school, including students
 - Management within school, including parents
 - Provision of information to community members through report cards or other means
 - Provision of training to management committee members
 - Community participation in decision-making
- Context & Conditions
 - Political Environment
 - Evidence of corruption (including rent-seeking and patronage)
 - Political interference
 - Informal governance
 - Evidence of democratic practices at national level
 - Evidence of democratic practices at local level
 - Inclusive democracy
 - Social norms
 - Decisions traditionally made autocratically by those in power
 - Decisions traditionally made by dominant social groups
 - Decisions traditionally made through local consensus
 - Education traditionally valued
 - Education not traditionally valued
 - Traditional attitudes towards female involvement in decision-making
 - Progressive attitudes towards female involvement in decision-making

- Traditional attitudes towards involvement of marginalised social groups in decision making
- Conflict
 - Conflict-affected area
 - Fragile environment
 - Non-conflict-affected area
- Linguistic Context
 - Linguistic monopoly
 - Linguistic diversity
- Geographic Constraints
 - Average distance to school
 - Evidence of constraints in physical access to school
 - Evidence of school safety
 - Evidence of lack of school safety
 - Evidence of natural hazards
 - Evidence of risks for girls and vulnerable children
- Resource Environment
 - Severe constraints on national education budget
 - Adequate national education budget
 - Education funding not available in local area
 - Funding available at local level
- Educational choice
 - Availability of private provision
 - English medium curricula
 - Dominant local language as a medium of instruction
 - Non-dominant local language as a medium of instruction
 - Shadow education
- Regulatory Framework
 - Clear regulatory framework
 - Evidence of confusion around roles and responsibilities
 - Evidence of weak implementation of regulatory framework

ANNEX 5: CODING FOR QUALITY APPRAISAL (REVIEW QUESTION 2)⁵

Transparency

- Research Question
 - o High: Study has a clear research question
 - o Low: Study does not have a clear research question
- Transparency of Research Design
 - o High: Study clearly states the design and methods
 - o Low: Study does not state clearly the design and methods
- Transparency of Data Source
 - o High: Study clearly reference which data were used and where they came from (source and/or how collected)
 - o 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
 - o High: Research design is appropriate for the research question
 - o Low: Research design is not appropriate for the research question
- Appropriateness of Sampling Population
 - o High: Population from which sample was drawn appropriate to research question and design
 - o Low: Population from which sample was drawn inappropriate for research question and design
 - o Unclear: Sampling method unclear
- Appropriateness of Sampling Method
 - o High: Sampling method appropriate for research question and design
 - o Low: Sampling method inappropriate for research question and design
 - o Unclear: Sampling method unclear
- Appropriateness of Sample Size
 - o High: Final sample size appropriate for analytical method
 - o Low: Final sample size inappropriate for analytical method
 - o Unclear: Sample size unclear
- Appropriateness of Sample
 - o High: Sample representative of the population and/or pertinent to the purpose
 - o Low: Final sample not representative of the population and/or pertinent to the purpose
 - o Unclear: Sample characteristics unclear
- Appropriateness of Data Collection Methods
 - o High: Data collection methods appropriate for the research design
 - o Low: Methods inappropriate for the research design

⁵ Based on DFID (2014).

- 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
 - 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
- **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

- **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

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