

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

A realist synthesis of school accountability in low- and middle-income countries

Protocol written by David Eddy Spicer, Melanie Ehren, Meena Khatwa (all London Centre for Leadership in Learning) and Mukdarut Bangpan, EPPI-Centre

EPPI-Centre
Social Science Research Unit
Institute of Education
University of London

July 2014

PROTOCOL

The authors are part of the London Centre for Leadership in Learning and EPPI-Centre and were supported by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre).

Eddy Spicer D, Ehren M, Bangpan M, Khatwa M (2014) *Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised? A realist synthesis of school accountability in low- and middle-income countries*. Protocol. London: EPPI-Centre, Social Science Research Centre, Institute of Education, University of London.

© Copyright

Authors of the systematic reviews on the EPPI-Centre website (<http://eppi.ioe.ac.uk/>) hold the copyright for the text of their reviews. The EPPI-Centre owns the copyright for all material on the website it has developed, including the contents of the databases, manuals, and keywording and data extraction systems. The centre and authors give permission for users of the site to display and print the contents of the site for their own non-commercial use, providing that the materials are not modified, copyright and other proprietary notices contained in the materials are retained, and the source of the material is cited clearly following the citation details provided. Otherwise users are not permitted to duplicate, reproduce, re-publish, distribute, or store material from this website without express written permission.

Contents

Abbreviations	iii
1. Background.....	1
1.1 Aims and rationale for the review	1
1.2 Definitional and conceptual issues	3
1.3 Research background	7
1.4 Review question	9
1.5 Authors, funders, and other users of the review	9
2. Methods used in the review.....	10
2.1 Realist synthesis methodology	10
2.2 User involvement.....	16
2.3 Identifying and describing studies	17
2.4 In-depth review	19
2.5 Deriving conclusions and implications	22
3. References	23
Appendices	27
Appendix 1.1: Authorship of this report.....	27
Appendix 2.1: Advisory Group terms of reference	29
Appendix 2.2: Timeline	31
Appendix 2.3: iterative review stages	32
Appendix 2.4: Inclusion and exclusion criteria	34
Appendix 2.5: Search strategy for electronic databases	35
Appendix 2.6: Sources	38
Appendix 2.7: Draft coding tool	40

Abbreviations

DfID	Department for International Development (UK)
EMIS	Education management and information system
HIC	High-income country
LIC	Lower-income country
LMIC	Lower-middle income country
PI	Principal Investigator

1. Background

This systematic review was commissioned by the UK Department for International Development (DfID). Its primary audience is DfID Education Advisers working in DfID's priority countries.¹ The findings of this review will also be of use to policymakers, politicians, civil servants, and educational leaders in low and middle-income countries (LMICs). Section 1.1 introduces the basic principles that are discussed in more detail in the rest of the chapter

1.1 Aims and rationale for the review

This review aims to explore the conditions under which three approaches to school accountability - inspection, monitoring and assessment - lead to improvement in schools and to positive learning outcomes for schoolchildren in low- and middle-income countries, especially the poorest and most marginalised. The review examines three key outcomes that accountability systems aim to improve: children's learning; the delivery of education within schools; and the efficiency of the system of education in achieving learning and delivery outcomes across schools.

The aim to understand the connections between particular conditions and school and system outcomes has led us to an approach to systematic review known as realist synthesis. We provide a brief overview of this approach in this background section and then elaborate on our rationale and the specific steps of carrying out a realist synthesis in subsequent sections.

Existing literature that we sketch in Section 1.3 portrays complex and varied links amongst governance context, policy, design of accountability systems, mechanisms of impact and school outcomes that make translation of conditions across studies challenging. In our rapid review of the literature, we did not identify any existing systematic reviews that attempted to link the accountability policies of low- and middle-income countries with particular school-level outcomes, perhaps in part due to the multiple and interdependent considerations that generalising across contexts entails. A focus on 'conditions' introduces a plethora of variables that range from the highly individual traits and characteristics of the professionals who are the subject of accountability systems to the wider cultural contexts of planning and implementation. This diversity is compounded by the complexity of educational systems, which typically encompass multiple levels. Those working at each of these levels all functionally interpret policy in their everyday actions.

This review, as will be made clear, is most interested in what happens at the school level as a consequence of policy implementation at the national or regional level. But the school level customarily comprises one or more sub-units, such as subject departments in a secondary school, and the school itself operates within one or more successive levels of organisation, from local authorities to regional and state directorates. To add to this diversity of contexts and complexity of organisation, the implementation of one particular element of accountability does not customarily occur in isolation from the implementation of other related policies. Some type of national assessment may accompany an inspection regime. Data used for monitoring may be used for multiple purposes throughout the system, including inspection and decision making around the consequences of poor results. Due to such overlap and multiple uses, particular outcomes cannot be easily tied to

¹ These are the countries with which DfID holds bilateral agreements. For a list of these countries, see: <https://www.gov.uk/government/organisations/department-for-international-development/about#where-we-work>. For a summary of DfID's bilateral engagement in education in 2013, see Annex 3 of the Education Position Paper (DfID, 2013a, p. 22).

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

particular accountability elements. Understanding the efficacy of inspection may entail parsing diverse contexts, multiple system layers and interaction effects among approaches to accountability that are currently implemented. In sum, trying to discern how any particular approach to accountability might yield positive outcomes for children and schools necessarily involves looking into a wide array of interactions within and across system levels that are probably highly specific to particular social, economic and cultural contexts.

Well-established approaches to the design of systematic review are unlikely to yield the kind of results that might connect conditions with outcomes for several reasons. Powerful approaches to review that seek to aggregate results, such as statistical meta-analysis, assume a known universe of comparable variables and conditions. Our initial, rapid review of existing literature confirmed what we had conjectured, that the quality of the available literature about accountability in our regions of interest, low- and middle-income countries, is highly varied, with a preponderance of descriptive case studies and reports from multilateral agencies. Our aim in this review is to look at the effect of an as yet unknown universe of potential conditions on a few clear outcomes. This put us in the realm of what Gough et al. (2012) characterise as ‘configurative’ approaches to review, theory-driven approaches that are most useful for elaborating and testing patterns drawn from a widely varied body of literature. Initially, we considered an approach known as ‘framework synthesis’, which offers useful ways of understanding the attributes of social programmes that yield particular outcomes within a given set of conditions. However, we believe that the most useful results from our review will not only characterise features of accountability interventions connected with key outcomes, but also illuminate how the intervention yields changed decision making and behaviours by teachers, school leaders and others, revealing the underlying processes that produce certain outcomes given the right conditions.

We have turned to realist synthesis (Wong et al., 2013; Pawson, 2006; Pawson et al., 2005) because of the complexity and dynamism of conditions that influence the outcome of accountability systems in LMICs, the wide variability in available literature, and our aim of providing systematic explanations of the mechanisms that are important for particular outcomes, given particular conditions. In a realist framing, the overriding question is, ‘What works for whom under what circumstances, how and why?’ (Wong et al., 2013, p. 2). The goal shifts from pinpointing features of effective interventions to explaining the mechanisms through which a given approach to accountability, operating under certain conditions, is more or less likely to cause outcomes of improved service delivery, equitable learning and, ultimately, overall system efficiency for the poorest and most marginalised children in LMICs. For example, a tightly constrained view of learning, teaching to the test, is a well-documented service delivery outcome given conditions such as low-performing schools, severe consequences for low performance and inadequately prepared teachers and leaders. For accountability interventions that include standardised assessment, this review will seek to identify mechanisms that result in teaching to the curriculum rather than just to what is assessed, yielding high-quality service delivery. These mechanisms that cause teachers and leaders to behave differently might include, for example, the existence of professional networks triggered by conditions that promote a coherent sense of professionalism within and across schools.

In this way, the review aims to help educational advisers, policy makers and educational leaders to understand the causal processes that result in certain outcomes and to identify the conditions that are necessary for those processes to have the desired outcomes. The fundamental aim is to sharpen our ability to develop programmes that reflect the complexities of implementation in LMICs in sophisticated ways that are sensitive to the most significant considerations of context. We recognise that achieving this aim is ambitious and that we are likely to be in a position to put forward only a limited range of viable configurations that explain how particular accountability elements yield desired

outcomes given particular conditions. However, even identifying a limited range of causal pathways will offer insight into areas for programme development through the elaboration of models that can be tested empirically, as well as areas for further research by identifying gaps in our ability to construct viable models.

1.2 Definitional and conceptual issues

This section intends to parse each of the key words in the research question that serves as the starting point for this systematic review. To summarise, the question draws our attention to the conditions that are necessary for the three elements of accountability systems - assessment, inspection and monitoring - to improve service delivery, system efficiency and children's learning outcomes in schools and school systems in LMICs, especially for the poorest and most marginalised children. Various interpretations exist for each of the important words in this question. Here we offer our operating definitions of 'accountability systems', the three accountability elements that are the focus of this review, and the three outcomes that are of greatest interest.

1.2.1 Accountability systems

'Accountability systems' denotes for this review the collection of administrative processes designed and deployed within a state or provincial-level educational system as instruments of school and school system governance. That is, these are system-level interventions put in place in order to hold individual schools and the wider system of which they are a part to account for efficient and effective service delivery and positive student learning outcomes. Our focus is on the educational system as delineated by organisations and administrative units connected by public sector provision and/or financing, and encompassing compulsory education, typically the primary and lower secondary phases.

Accountability systems are traditionally designed as a form of quality control intended to safeguard the legitimate use of public resources. Broad reforms to the system of schooling instituted in many countries from the 1990s saw accountability used as a means of quality improvement and decision making around resources, bringing increased focus to accountability as a mechanism for not only defining standards and monitoring individuals, schools and school systems against those standards, but also as a lever for efficient resource allocation, positive change and capacity building within each organisation and across the system of education as a whole (Rosenkvist, 2010).

In defining accountability we are acutely aware of the 'accountability problem'. Onora O'Neill (2013), drawing on her widely discussed Reith Lectures of 2002, argues that 'More accountability is not always better, and processes of holding to account can impose high costs without securing substantial benefits' (p. 4). What is required, she argued, is 'intelligent accountability', which requires 'more attention to good governance and fewer fantasies about total control' (O'Neill, 2002, p. 58). Michael Fullan (2010, p. 27) cites 'intelligent accountability' as one of seven 'big ideas for whole school reform', noting:

The failure to get accountability right plagues all reform efforts ... Intelligent accountability involves a set of policies and practices that 1) actually increases individual, and especially collective, capacity so that shared responsibility carries most of the weight of effective accountability; 2) makes internal and external accountability almost seamless; and 3) leaves external accountability to do its remaining, more manageable task of necessary intervention.

It is important to emphasise that the focus of this review is not on accountability as *outcome*, as in evaluating the degree to which different social interventions may foster or discourage greater accountability. The review is interested in three distinct elements of accountability as social interventions leading to (or diverting from) outcomes of improved service delivery, improved student learning, and ultimately system efficiency.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

Elements of accountability

The Additional Information (DfID, 2013c) that accompanied DfID’s call for proposals for this review notes the following three types of accountability:

1. Regulatory school accountability: ensuring compliance with laws and regulations: focuses on inputs and processes within the school, e.g. school inspections
2. Performance/results-based accountability to improve schools: periodic school evaluations. Mechanisms include a) standardised student testing, b) public reporting of school performance and c) rewards or sanctions. In other words, they use assessment systems or monitoring systems
3. Performance-based accountability to improve administration or management: use of monitoring data and targets to improve system efficiency and delivery.

These definitions are adapted from the OECD framework that specifies the elements listed in Table 1.1.

Table 1.1: Types of school accountability

Vertical	<i>Regulatory school accountability:</i> Compliance with laws and regulations; focuses on inputs and processes within the school.
	School performance accountability: Periodic school evaluations.
Horizontal	<i>Professional school accountability:</i> Professional standards for teachers and other educational staff.
	<i>Multiple school accountability:</i> involving students, parents and other stakeholders in formulating strategies, decision-making, and evaluation.

Source: Hooge et al. (2012, p. 9)

In Hooge et al. (2012), the OECD traces the rise of horizontal accountability through an emphasis on professional standards as an effort to establish expectations and show clear pathways towards improvement, and through stakeholder/community accountability initiatives aimed at embedding the school within a wider set of local relations. These elements of horizontal accountability have come about in response to a focus on the development of *internal* school accountability, that is, the development of shared expectations amongst students, teachers, school leaders and other local stakeholders about learning outcomes and service delivery, along with processes for monitoring whether these expectations are achieved (Elmore, 2002). This review focuses specifically on those *vertical* accountability elements characteristic of *external* accountability, with particular attention on the three accountability elements of inspection, assessment and monitoring. While the elements of horizontal accountability will not be directly addressed, they will necessarily be important in understanding the essential conditions by which external accountability elements can bring about changed decision making and behaviours that give rise to desired outcomes.

The accountability elements identified - inspection, monitoring, assessment - do not correspond tightly with the definitions of ‘regulatory’ and ‘school performance’ accountability given above. For example, school inspection typically encompasses both regulatory and performance-based accountability, but often also include an evaluation of the output of schools (aggregated test scores of students). In some systems, in which inspection is viewed as a developmental opportunity for individual teachers, for example, ‘supervisors’ assigned to newly qualified teachers, inspection may also be seen as an aspect of horizontal accountability in creating learning opportunities around normalised expectations of practice.

1.2.2 Operating definitions for accountability elements

Inspection

School inspections are external evaluations of schools, undertaken by officials outside the school with a mandate from a national/local authority. Regular visits to schools are an essential part of school inspections to collect information about the quality of the school, check compliance to legislation and/or evaluate the quality of students' work (e.g. through observations, interviews and document analysis). Inspection systems were originally introduced in a number of European countries in the nineteenth century (e.g. HMI, now OFSTED, UK, dates back to 1834) and have become complex and intricate systems, using different terminologies and playing different roles.

Inspection systems in developing countries have a substantially different mandate and make-up compared to those in developed countries. Often the term 'supervision' is used when referring to inspection, and as De Grauwe (2007) describes, supervisors' role is not only to control and evaluate (as is often the case in developed countries), but also to advise, assist and support head teachers. Sometimes supervisors also have managerial tasks and are, for example, responsible for deployment of teachers, or deciding on promotion of teachers and head teachers. We are particularly interested in the evaluative dimensions of the role and recognise that a developmental brief held by the same role holder may give rise to different mechanisms and yield distinctly different outcomes.

Assessment

This review focuses on assessment of learning that is standardised at the provincial, national or regional levels and used to provide feedback to and evaluate the performance of individual schools and school system sub-units. 'Standardised' points to consistency in 'test design, content, administration and scoring to ensure comparability of the results across students and schools' (Best, 2013, p. 2). Test content is standardised to the extent that the results aim to assess students' cognitive skills in the subjects that comprise the most common aspects of curricula - literacy, mathematics, science, civics, for example. Best's systematic review (2013) of the impact of assessment on education policy offers insights into and a useful contrast to the emphasis of this review on school-level outcomes.

Monitoring

Monitoring has proved the most difficult element to delineate clearly and encompass within a discrete set of keywords. For the purposes of this review, 'monitoring' refers to system-level processes designed to collect and use school-level information. This may include the specification of information to be collected, the processes of aggregating and analysing information, the dissemination of results, and ultimately the types of actions that are taken or not taken as a consequence. Monitoring includes formal systems of EMIS (Education Management Information Systems), and the collation of 'input' or administrative data, as well as performance monitoring that specifies and aggregates performance information (e.g., school 'report cards'). System-level specification, aggregation and dissemination of school-level information, with output intended to promote school improvement and effectiveness, stand as key definitional characteristics of any monitoring approach covered by this review.

1.2.3 Operating definitions for outcomes

Service delivery

'Service delivery' is used here to refer to school- and system-level processes of organising work that have an effect on learning outcomes. It includes the 'technical core' of schooling, the primary processes that provide the conditions for learning in the classroom,

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

as well as the wider organisational structure and environment that provide the direct and indirect conditions for classroom practice. The education system comprises myriad actions and decisions of ‘service providers’ working at the school and system levels. What providers of education know and do has a pronounced effect on the quality of learning in schools and the quality of the system. The World Bank, the African Development Bank and the African Economic Research Consortium have developed a set of indicators for schooling across Africa that aims to support national efforts to improve school accountability (World Bank, 2011). The indicators focus on three general areas: 1) inputs and infrastructure at the school level; 2) effort and knowledge of teachers; and 3) availability of resources (see Table 1.2).

Table 1.2: Service delivery indicators

Indicator	Definitions
At the school: Inputs and infrastructure	
Infrastructure (electricity, water, sanitation)	The indicator measures if primary schools have access to electricity, improved sanitation and clean water. The indicator is 1 if schools have access to all three services, and 0 if they lack one or more of them.
Children per classroom	The indicator of availability of classrooms is measured as the ratio of the number of primary school age children per available primary school classrooms.
Student/teacher ratio	The indicator of teachers’ availability is measured as the average number of students per teacher.
Textbooks per student	The indicator of learning material is measured as the overall number of books available within primary schools per student. It is calculated as the sum all books per grade, which is then summed over all grades.
Teachers: Effort and knowledge	
Absence rate	The indicator of absenteeism among frontline teaching staff is measured as the share of teachers not in schools as observed during one unannounced visit.
Time children are in school being taught	The actual time children are in school being taught per day is measured, combining data from the absenteeism survey, reported teaching hours and classroom observations.
Share of teachers with minimum knowledge	This indicator measures teacher’s knowledge and is based on mathematics and language tests covering the primary curriculum administered at the school level to all teachers of Grades 3 and 4.
Funding: Effort in the supply chain	
Education expenditure reaching primary school	The indicator of availability of resources at the primary school level assesses the amount of resources available for services to students at the school.
Delays in wages	The indicator captures the share of teachers who have wages due in excess of two months.

Source: Bold et al. (2011, pp. 57-58)

These indicators will serve as initial proxies for service delivery outcomes in our efforts to understand causal processes associated with inspection, monitoring or assessment, and the conditions under which those processes are triggered. We would expect that these indicators will also serve as important conditions and/or causal processes that themselves trigger the other two outcomes of system efficiency and student learning.

System efficiency

Many countries are facing the challenge of orchestrating educational policies that promote organisational autonomy while attempting to drive system improvement and coherence through more rigorous approaches to accountability. These ‘loose-tight’ controls make crafting a generic and operational definition of ‘system efficiency’ difficult. In general terms, system efficiency is ‘the desired level of output for the lowest cost’ (Scheerens, 2000, p. 21). Characterising ‘output’ can be viewed in the short term as the ability of schools within a system to deliver educational services that provide the best possible learning outcomes at the lowest possible cost. This technical view of system efficiency, however, does not take into consideration *societal* efficiency, the long-term effects of schools within a system on the future prospects of students (Cheng, 1993, as quoted in Scheerens, 2000, p. 22). Both technical and societal efficiency are important to consider for a holistic understanding of system efficiency. For technical efficiency, we expect to look closely at processes and conditions that enable the system to ensure that education expenditures reach the school (a service delivery indicator) and that expenditures are then used in ways that improve learning outcomes for the poorest and most marginalised students (i.e., the technical aspects of system efficiency). In terms of societal efficiency, we take the premise that the desired outcome is for the educational system to ensure access and equity by addressing entrenched societal disadvantage such as gender disparities, geographic isolation, disabilities and ethnic, religious and linguistic disadvantages (DfID, 2013a, p. 6; UNESCO, 2008). We are keenly aware that DfID programme efforts have paid particular attention to giving rural girls from the poorest families access to school and helping them stay in school (DfID, 2013a, p. 10). We intend the review to contribute to greater understanding of the effects of accountability elements on this as an important system efficiency.

Learning outcomes

Learning outcomes have a wide range of definitions, from concern with ‘quantity’, as expressed in years of schooling and used in studies on returns to education (e.g., Mincer, 1974) to the broad and aspirational qualities portrayed in the United Nation’s Universal Declaration of Human Rights. Clarifying what one means by student learning outcomes depends on the purpose coupled with identification of appropriate proxies. We anticipate that much of the research we will find for LMICs will be focused on the ‘quantity’ end of that spectrum, measuring learning outcomes in terms of children’s enrolments, attendance, retention, year repetition, survival and completion rates. We may also find recent research attempting to gauge the effects of schooling on labour market participation, which moves beyond years of schooling to focus on the acquisition of cognitive skills as expressed through student performance on standardised assessments (Vegas and Petrow, 2008, pp. 8-9). Thus we also include performance on standardised assessment as a proxy for learning outcomes.

1.3 Research background

The research question emphasises the conditions under which three elements of school accountability - monitoring, inspection and assessment - improve learning outcomes for children as well as lead to systemic improvements in education for the poorest and most marginalised in LMICs. The literature on *assessment* for accountability has focused on standardised (high stakes) assessment over nearly three decades and includes large-scale surveys, small cases studies and quantitative analyses of test scores (Stecher, 2002). Of relevance to the proposed review is a recent systematic review examining the impact of assessment programmes on the formulation, monitoring and evaluation of policy in developing countries (Best et al., 2013). The search phase, conducted in 2011, identified 1,080 studies of potential interest, one-third of which were selected for full-text retrieval. In relation to understanding conditions and mechanisms of impact, studies in high-income

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

countries (HICs) have provided descriptive taxonomies of less and more effective practices (Stecher, 2002; Haladyna et al., 1991; Popham, 1991; Mehrens and Kaminski, 1989). Our emphasis on processes in LMICs that cause outcomes and the conditions that give rise to these processes challenges the ready translation of research from high-income countries. Taxonomies may prove useful in conceptualising relationships among conditions, causal processes and outcomes. In any case, the results from these studies bolster the need for close attention to context, as studies have consistently found that most practices are neither clearly effective nor ineffective because consequences for student learning are contingent on the context in which, and the extent to which, practices occur.

Increased use of data to *monitor* administrative and management performance in schools and school systems has accompanied increased use of national and international standardised assessment worldwide. In high-income countries, relevant research has focused on how schools use data as a means of monitoring and improving school and teaching effectiveness, not primarily to monitor and develop system capacity (Schildkamp et al., 2012). In the US, case studies have also focused on improving educational quality in high-poverty schools (Kerr et al., 2006; Wayman and Stringfield, 2006). In contrast, nearly three decades of emphasis on EMIS in developing countries has resulted in a compendium of descriptive and evaluative studies of their national implementation (Scepanovic et al., 2010; De Grauwe, 2008; Powell, 2006).

The literature on *inspection* is relatively recent, but has emerged as a strong focus in a wide range of case studies, surveys and quantitative analyses of inspection results and student achievement results of inspected versus non-inspected schools. Most studies are set in Europe (particularly England and the Netherlands), but the work of De Grauwe (2001, 2007) is also situated in African countries. Many studies (e.g. De Grauwe, 2007; De Grauwe and Lugaz, 2007; Dembélé and Oviawe, 2007; De Grauwe, 2001) point to a lack of resources, inefficient management and an organisational structure not adapted to current realities when describing school inspections in developing countries. Inspectorates of Education in developing countries often face a high school/supervisor and teacher/supervisor ratio, which results in a high workload. As many inspectorates also often face a lack of financial and material resources (e.g. computers, resources to travel to schools in remote areas) and have a very demanding job description (including myriad tasks related to supervision of and support for schools and teachers and additional administrative and liaison tasks) this workload is even more difficult to manage. Management problems particularly refer to challenges in selecting, recruiting, training and career development, support and evaluation of school inspectors, according to De Grauwe (2007). In many developing countries, school inspectors are recruited from school staff and sometimes lack experience in school management; when they occupy the same grade as principals in schools, principals often do not consider school inspectors as their superiors and may refuse their advice, causing a lack of impact of school inspections. Such a situation may also occur when school inspectors lack the relevant knowledge and skills to provide effective and valuable feedback to schools (including the tone of voice when providing feedback) on the areas in the school that are most in need of improvement. As many school inspectors face a lack of opportunities for career development, they may also lack incentives to innovate and improve their working methods. According to De Grauwe (2007) and De Grauwe and Lugaz (2007), organisational problems often include a lack of structure and clarity in the inspection system, a lack of co-ordination between inspection services and other organisations supporting school development and improvement (e.g. teacher training centres) and a lack of autonomy of school inspectors to follow up on their recommendations to schools. As Dembélé and Oviawe (2007) point out, these challenges have to be identified to find school inspection models and structures that are most suitable and have the highest chance of success within the specific context of developing countries. Recent literature reviews by Klerks (2013) and Nelson and Ehren (2014), drawing on studies primarily from England and the Netherlands, summarise the effects and

side-effects of school inspections on teachers' behavioural change, school improvement and student achievement. These reviews show that the overall results of inspection research are, at present, far from conclusive (Klerks, 2013; Luginbuhl et al., 2009; Rosenthal, 2004).

Our current understanding is that extensive literature exists in all three areas, suggesting separate reviews. However, we hypothesise that some related conditions may affect all three in similar ways, particularly governance context and administrative and evaluative capacity in the education system (see Barber, 2004). However, the right set of conditions may or may not trigger similar processes that cause outcomes for different accountability elements. This requires that we pay close attention to the ways the connections between conditions, the processes that arise from those conditions and the outcomes that are caused by those processes. Realist synthesis, as we explore below, is particularly well-suited for exploring these connections and developing conceptual models that may inform the decisions of researchers, policy makers and educators.

1.4 Review question

The question that we aim to address is:

Under what conditions do the following elements of an education system improve system efficiency, service delivery and learning outcomes, especially for the poorest and most marginalised in low- and middle-income countries?

- a) Monitoring systems, including using administrative data systems (e.g. EMIS) as well as more targeted monitoring mechanisms.
- b) Inspection systems
- c) Assessment systems.

1.5 Authors, funders, and other users of the review

We anticipate working closely with DfID, the most immediate user of the proposed review. Education Advisers are the primary audience within DfID. They work at the country level, managing and overseeing DfID programmes, as well as with governmental and non-governmental experts and policy makers. DfID head office staff and educational consultants would also find the review useful in support of their evaluation of accountability policy and implementation.

This review will be of use to other agencies in the design/reform, implementation and evaluation of accountability systems. Such agencies may include bilateral and multilateral agencies and organisations working in LMICs. Other interested parties will include researchers, academics and non-governmental organisations that have interests in using, disseminating and communicating results that may inform evidence-based policy making and practice. The methodology of the review, realist synthesis, has only recently been employed in systematic reviews in education. The design of the review may serve as a model for others embarking on systematic reviews in this area.

2. Methods used in the review

2.1 Realist synthesis methodology

The review aspires to build explanatory models, or one or more ‘middle-range theories’, that trace paths across conditions, mechanisms and outcomes related to the accountability elements of inspection, monitoring and assessment. Sociologist Andrew Sayer, who has charted realism across the social sciences, explains why the ‘careful conceptualisation’ entailed in developing explanatory models is warranted for the complex problems that social science aims to understand.

Social systems are always open and usually complex and messy. Unlike some of the natural sciences, we cannot isolate out these components and examine them under controlled conditions. We therefore have to rely on abstraction and careful conceptualisation, on attempting to abstract out the various components or influences, and only when we have done this and considered how they combine and interact can we expect to return to the concrete, many-sided object and make sense of it (Sayer, 2000, p. 19).

It is the rigorous process of systematically building or testing a range of middle-range theories that marks the realist review out from other review approaches. Similar to framework synthesis, realist synthesis depends on the elaboration of an ‘initial rough theory’ and the refinement of that theory through systematic review. Realist synthesis embraces theory building and testing at a greater level of specificity than does framework synthesis, developing conceptual understanding not only of the attributes of an intervention but also by elaborating relationships amongst specific features of context, programme mechanisms and intermediate outcomes. These ‘theories of the middle-range’ (Merton, 1968) offer explanatory power by operating within an empirically specified range of generalisability to explain how specific mechanisms cause particular outcomes, given the right conditions (Wong et al., 2013, p. 2).

The emphasis on theory is grounded in programme reality. Realist approaches view social programmes, like the implementation of an inspection regime, as a set of propositions - or theories - about how change comes about. Whenever an inspector shows up at a school, she or he is enacting the theory of change that underlies the inspection programme. A programme’s theory of change is typically implicit; it is assumed that the results of and feedback from inspection will cause teachers and school leaders to make decisions and take actions that align their own practice and the school with desired educational standards. An important task of a realist synthesis is to probe the primary literature to develop clear understandings about how and why a class of programmes is found to ‘work’ to generate the outcomes of interest (Wong et al., 2013, p. 2). For example, Ehren et al. (2013) found that practitioners’ actions on inspection feedback were rare. Much more common were actions based on the anticipation of inspection; in this understanding, it was the expectation that caused behaviour to change, not the results from the inspection itself.

Realist syntheses illuminate the implicit theories of social programmes by giving reviewers a systematic way of hypothesising the conditions (C), or contextual influences, that are found to trigger relevant mechanisms (M), or causal processes, that result in the outcomes (O) of interest. C-M-O configurations explain how programme actions cause particular outcomes, given the right conditions. A programme’s espoused theory of change may or may not correspond with the way change is enacted through configurations of conditions, the mechanisms triggered by these conditions and the outcomes caused by the mechanisms. Hypothesising and testing C-M-O configurations related to desired outcomes allows reviewers to develop theories that do a better job of explaining the ways programmes act in the world or to test known theories to see if they hold up with evidence from other studies.

This review is a *theory-building* review, in that we are reviewing primary literature in a field that is under-theorised. The connections between accountability implementation and intended outcomes are most often assumed to be an inevitable result of implementation and not systematically interrogated. A ‘theory-testing’ review would be possible when a relatively limited set of theories has been adequately hypothesised and described (Westhorp et al., 2014, p. 22, fn 7). By systematically identifying C-M-O configurations for a class of programmes, we are able to hypothesise systematically and with specificity the different ways that programmes are more or less likely to realise their intended outcomes. The results of this synthesis offer guidance to educators and policy makers about altering conditions to have greater likelihood of triggering the mechanisms that cause the intended outcomes (Wong et al., 2013, p. 2).

Realist synthesis, while relatively new to systematic reviews in education, has been used in a wide range of social science research.² This review follows the publication standards for realist reviews put forward by the RAMESES (Realist And Meta-narrative Evidence Syntheses: Evolving Standards) project (Wong et al., 2013). To illustrate the benefits of this approach, we turn to a recently completed systematic review funded by DfID and exploring an important issue of contemporary educational policy, community accountability, through realist synthesis. The review by Westhorp et al. (2014) employs a theory-building realist synthesis to address the question: ‘Under what circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor?’ Its findings identify 11 mechanisms and 13 categories of features of context. Mechanisms characterise the key processes through which community accountability interventions work. In this way, the review develops and refines middle-range theories about ‘the ways in which interventions work, the contexts in which they do and do not work and the differentiated patterns of outcomes that they generate’ (Westhorp et al., 2012, p. 13).

One of the middle-range theories that Westhorp et al. (2012) hypothesise involves the conditions that lead stakeholders to take actions that enhance local responsibility for schooling, given rewards and sanctions. The synthesis of primary evidence conducted by the review team lead them to a mechanism labelled ‘carrots and sticks’ (Westhorp, 2014, p. 45). A related mechanism explains not the consequence of rewards or sanctions but the anticipatory effect that awareness of inevitable sanctions or rewards might have on actors, who sculpt their actions accordingly, a mechanism that the review labels ‘big brother is watching me’ (ibid, p. 45). The mechanism of ‘big brother is watching me’ is similar to the effect of establishing expectations from inspection found by Ehren et al. (2013).

Accompanying the mechanisms is an analysis of features of context that are essential to the operation of each mechanism. One of the review’s findings around context has relevance for this review. An important feature of context for several mechanisms was the existence of a national, high-quality system of assessment of student learning and the orientation of those systems towards ‘collective action’. As an example, they identify the following passage from one study as characteristic of the programme theory (Weiss, 1998) that underlies such approaches:

these measures will empower citizens to hold their governments accountable for improving the quality of their children’s education, and also equip them with the knowledge necessary to contribute themselves to improving their children’s learning. (Lieberman et al., p. 8, quoted in Westhorp et al., 2014, p. 64)

² See Pawson et al. (2004) for examples. For more recent resources, see the website of the RAMESES project, <http://www.ramesesproject.org/>.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

The reviewers note that they did not identify any studies that directly examined the link between the assessment system and the effectiveness of community accountability. Nonetheless, the review was able to assemble findings from two reviews, one of which examined assessment systems and student results in Mexico, the other of which looked at the use of results from an assessment system in Uruguay to support collaborative action to improve learning outcomes (Westthorp et al., forthcoming 2014).

The report then concludes with nine recommendations for policy and practice. Reviewers derive these from their elaboration of middle-range theory, identifying the conditions under which certain mechanisms cause desired outcomes. The constellation of middle-range theories is then used to return to an 'initial rough theory' developed at the start of the review and strengthen it so that it can more robustly identify the proper conditions and related mechanisms that lead to desired outcomes for community accountability and empowerment initiatives. The review also clarifies the kind of research that would appear to be most needed to build better and more durable understanding of such programmes.

In subsequent sections, we elaborate the 'initial rough theory' that we are developing and outline how we intend to use that to hypothesis middle-range theories.

2.1.1 Components of accountability

We can now begin to tease apart the generic structure of systemic elements of accountability. The term 'systemic' here indicates that the element is part of an intervention designed and deployed at a system level above that of the individual school. This may be the nation state or a region, state or province in a federal national system. Broadly, systemic accountability elements are a form of performance-based contracting (Bouckaert and Halligan, 2007). Generic phases of many approaches to accountability might be identified as:

- benchmarking - the delineation of standards, performance information, performance measurement.
- incorporation - integrating definitions into documents, procedures, discourses
- use - in what ways, if any, the output from the process of incorporation is used within the system. This may include the consequences of outputs of the process for the organisation and individuals.

We can then develop a generic hypothesis about how systemic accountability intends to influence service delivery, systemic efficiency, and learning outcomes based on the integrated open systems model of school effectiveness put forward by Scheerens (1992). At its most basic, schooling at the organisational level consists of four aspects:

- inputs of technical, human and social capital
- processes of the technical and administrative core, with 'technical' indicating classroom-level interactions amongst teacher-students-curriculum and 'administrative' the organising processes of the school
- outputs that relate to student learning
- outputs that relate to the technical efficiency of the school.

Finally, we can discern two levels of outcomes - those at the organisational level and those at the level of the educational system. At the organisational level, we expect to see increased student access to education, reflected in increases in enrolment as well as more regular student attendance; we also expect to see more time devoted to teaching in classrooms and greater allocation of education expenditure for teaching and learning as an outcome. Finally these outcomes can be translated across schools in ways that lead towards system outcomes, of technical efficiency as well as societal efficiency (Cheng,

1993) - the contributions of the school and school system to an educated, equitable society.

Within this model, we draw on and extend Ehren et al. (2013) and Hatch (2013) to highlight five hypothetical mechanisms to explain how accountability systems lead to organisational and system-level outcomes:

- setting expectations
- providing feedback/consequences
- institutionalisation of norms
- capacity development of educators
- capacity development of local stakeholders.

Each of these mechanisms operates at multiple levels within the overall system and in the relationship of the system to external stakeholders (e.g., community members, politicians, policy makers). In this review, our focus is on the organisational implications of systemic elements. A realist review intends to identify mechanisms of programme action such as these and then describe the conditions under which they do or do not yield desired outcomes. Our interest in this review is in examining those mechanisms that produce school-level outcomes, as described in Table 2.1.

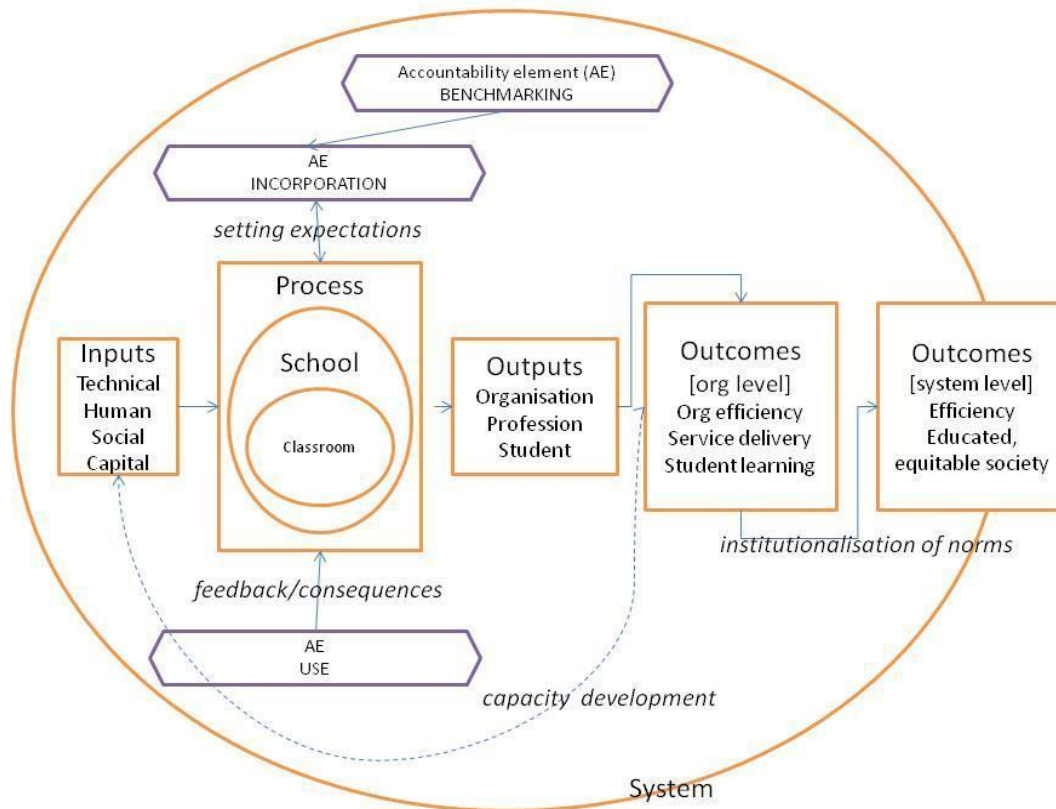
Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

Table 2.1: Provisional generic Conditions-Mechanism-Outcome Configuration (C-M-O)

Conditions	Mechanism	Outcome
<ul style="list-style-type: none"> • Effective monitoring systems • Belief that the authority holder will act on data received through monitoring system • Incentives of sufficient power • Performance can be observed 	Setting expectations	Improvements in the extent to which, or standards at which, responsible parties implement the actions required of them.
<ul style="list-style-type: none"> • Authority holder acts on performance information received through monitoring system • Effective uses of performance information for performance improvement • Incentives of sufficient power • Performance can be observed 	Providing feedback/ consequences	Improvements in the extent to which, or standards at which, responsible parties implement the actions required of them.
<ul style="list-style-type: none"> • Educators recognize value and see benefit of existing expectations • Concrete performance expectations integrated into processes of school organizing • Sustained support for development of skills and knowledge 	Institutionalisation of norms	Organisational and individual internalisation of system expectations Internal accountability with focus on meeting service delivery and learning outcomes expectations, not consequences
<ul style="list-style-type: none"> • Investment in developing high-quality teaching practice • Sustained and highly-respected opportunities to put skills into practice • Support for continued development of skills 	Capacity development of educators	Sustained improvement in service delivery Sustained improvement in student learning outcomes
<ul style="list-style-type: none"> • School leadership and staff capacities and attitudes support stakeholder engagement • Information, training and support provided to stakeholders 	Capacity development of stakeholders	Stakeholders have the skills to undertake roles expected of them Quality of stakeholder oversight of schooling Resources available for education improved

Source: Adapted from Westhorp et al. (2014, pp. 59-60)

Figure 2.1: Initial rough theory: Generic programme theory of change for accountability elements (hexagons) mapped against an open systems model of school (squares) inputs, organising processes, outputs and outcomes and system outcomes (oval)



Our aim is to treat the three elements of accountability - inspection, assessment, monitoring - as distinct and to then conduct analyses of key programme mechanisms that cause the intended outcomes and the conditions that trigger those mechanisms within and across all three elements. We anticipate that this comparative analysis will permit us to hypothesise about some middle-range theories that might operate for any accountability element - exemplified by the hypothetical configurations proposed above - as well as C-M-O configurations unique to each component, which we have not yet identified. It will also be important to identify how mechanisms of one element may act as conditions for the implementation of a companion element within the same system.

2.1.2 Risk assessment

We have proposed pursuing a realist review because of the complexity of accountability as an intervention and the promise that realist review offers of explaining how certain mechanisms cause particular outcomes given the right conditions. We believe this offers the most potential for answering the review question's stem, 'Under what conditions...' However, this approach does entail risk. In the best case, we will have adequate primary literature to elaborate several important middle-range mechanisms that operate within and across accountability elements as well as across a variety of geographical contexts. In the worst case, we will not have adequate evidence to develop more robust hypotheses. Below we outline two possible areas of risk and how we propose to address each.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

1. The review team does not have specific experience with systematic reviews using realist approaches and the question, composed of three distinct elements, is particularly complex. We believe this risk is manageable for the following reasons:
 - a. Each member of the team has relevant experience that contributes to overall strength. The Principal Investigator (PI) is knowledgeable about critical realism and realist theory. He has applied related approaches in theoretical synthesis, empirical studies and rigorous review (Eddy Spicer et al., 2014; Eddy Spicer, 2012a,b). The Co-PI has used the elucidation of programme theory of the Dutch Inspectorate to evaluate the effectiveness of school inspections (Ehren and Honingh, 2011; Ehren and Visscher, 2006; Ehren et al., 2005). She has also conducted a comparative analysis of inspection in six European countries that led to the identification of several of the key mechanisms proposed for this review (Ehren et al., 2013). The Research Officer is an experienced systematic reviewer who has conducted various types of systematic review. She has been working with the EPPI-Centre since 2005. She has trained and supported systematic reviewers, including international training. She is one of the tutors on the EPPI-Centre's MSc in Research for Public Policy and Practice. She is experienced in both qualitative and quantitative data synthesis methods.
 - b. The EPPI-Centre is interested in exploring the potential of realist reviews in education and has offered support for this approach.
 - c. The complexity of the question, as we have argued earlier, enhances rather than diminishes the value of a realist approach.
2. There may not be enough relevant data of suitable rigour from primary studies to test theories by elaborating relationships amongst specific features of context, mechanisms and outcomes.

To address this as soon as possible, we propose to consult with the EPPI-Centre support officer, DfID and our Advisory Group when we have completed the mapping of available primary studies. This consultation will include a recommendation about the further design of the review. If sufficient primary studies cannot be identified, we will propose a realist-informed framework synthesis (see Section 2.3.1).

2.2 User involvement

We envision three levels of user involvement. We are working with a small group of academic advisers, with whom we have already been in contact, in the design of the initial rough theory and in its iterative testing through the identification and verification of key mechanisms. This group of academic advisers include Professor Thomas Hatch and Dr Luis Huerta of Teachers College, Columbia University, and Anton De Grauwe, IIEP, UNESCO. We anticipate that this group will work with us throughout the development of the review.

The second level of user involvement involves our work with an Advisory Group. We are also recruiting five additional advisers to participate in the review Advisory Group (see Table A1.1 in Appendix 1.1). Invitations have gone out to potential members of the Advisory Group and we have received positive replies at the time of publication from those listed at the beginning of the protocol. The terms of reference for the Advisory Group are included as Appendix 2.1.

The third level of user engagement is around eliciting relevant sources and in disseminating initial and final findings. We will involve other stakeholders from the earliest stages by informing them of the review and asking for information about relevant

literature. These contacts will be identified through the extensive networks of the authors and the advisory group in the design and implementation of inspection, assessment and monitoring systems in LMICs. These may include contacts with academics and researchers in LMICs, those working in ministries of education, representatives of donor agencies, policy makers and education specialists.

The PI and Co-PI will present the results of the review at academic conferences, such as the American Educational Research Association and the International Congress for School Effectiveness and Improvement, as well as conferences attended by other stakeholders. Members of identified groups will be included in the distribution of the draft report as well as in the dissemination of the results.

Both the full report and the executive summary will be published online. Links to the published report will be circulated through a range of communication strategies (e.g. emails, websites and twitter) to potential users of research identified by the review team and the advisory group networks as outlined above. We will submit articles based on the findings from the review to peer-reviewed journals.

2.3 Identifying and describing studies

Identifying and describing studies occurs iteratively over three distinct phases (see timeline in Appendix 2.2):

1. theory elaboration (Section 2.3.1)
2. identification, mapping and in-depth review of primary studies of interventions that involve at least one of the three elements of accountability (inspection, assessment, monitoring) (Sections 2.3.2-2.3.6)
3. identification and in-depth review of sources of evidence that contribute to elucidating mechanisms and corresponding features of context and distinctive outcomes for each accountability element. These phases are iterative, as shown in Appendix 2.3.

These phases have several common stages, including defining relevant studies through inclusion/exclusion criteria, elaborating a search strategy to identify potential studies, conducting systematic searches to retrieve studies of interest, applying inclusion and exclusion criteria to screen studies, characterising included studies and assessing quality.

2.3.1 Defining relevant studies: Phase A - theory elaboration

The first phase has led to the development of the initial rough theory presented in the protocol, which will be used throughout the review to inform the mapping and testing of mechanisms. The scan of literature for use in developing tentative theories was undertaken by the PI and Co-PI and involved identifying relevant articles from academic journals, scholarly books and reports from multilateral and regional organisations (e.g., World Bank, IIEP/UNESCO, OECD, Brookings Institute). We read the full text of 25 articles, reports and chapters of books in the development of our initial rough theory. The PI, Co-PI and members of the review team all read Westhorp et al., (2012, 2014) and Pawson (2006), to familiarise ourselves with realist review methodology, and Wong et al. (2013), to familiarise ourselves with publication standards for realist synthesis. The PI and Co-PI also read or re-read several other sources (Pawson, 2006; Pawson et al., 2005; Pawson and Tilley, 1997), as well as methodological and training materials available on the RAMESES website.³ They developed an initial rough programme theory and identified several

³ RAMESES Project: Project Outputs, available at http://www.ramesesproject.org/index.php?pr=Project_outputs

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

hypothetical mechanisms based on this literature. The review team will seek feedback from academic advisers about the initial rough theory included in the protocol. We will refine the first draft of the initial rough theory based on this feedback.

2.3.2 Defining relevant studies: Phase B - inclusion and exclusion criteria

The second phase overlaps with the first. In this phase we are seeking to identify relevant primary studies of interventions that address each accountability element. We will include studies that meet all of the criteria listed Appendix 2.4, which relate to: type of intervention; geographical location; setting; types of studies; language; and date.

2.3.3 Identification of potential studies: Phase B - search strategy

The search strategy of Phases B and C aims to identify all studies and conceptual papers of direct relevance to the research question (see examples of search strategy in Appendix 2.5).

Search terms

Key search terms are determined by the review question and the inclusion criteria and will be developed iteratively and piloted against papers already identified through hand searching of websites and reference checking of literature identified in the scoping stage. References, citations and tracking of authors is being used in the structured searches to find studies most relevant to the initial rough theory, along with careful screening of relevant web sites

Search strings have been developed for each database using combinations of the main terms and their synonyms which denote key aspects of the review (see Table 2.2). The search uses the Boolean operator 'OR' to link each key aspect to their synonyms. Then, all key aspects are combined using 'AND' to identify relevant literature. For example, (accountability OR inspection OR monitoring OR assessment) AND (primary education OR secondary education).

A table of the key search terms used and an example of their use in a specific search can be found in Appendix 2.5.

Sources

We will search for relevant literature both published and unpublished from different sources. A detailed list can be found in Appendix 2.6.

A database system using EPPI-Reviewer (Thomas et al., 2010) will be set up to keep track of screening and coding studies found during the review. Titles and abstracts will be imported where possible, and otherwise entered manually into EPPI-Reviewer.

2.3.4 Screening studies: Phase B - applying inclusion and exclusion criteria

Inclusion and exclusion criteria will be applied successively to (i) titles and abstracts and (ii) full reports. Full reports will be obtained for those studies that appear to meet the criteria or where we have insufficient information to be sure. The inclusion and exclusion criteria will be reapplied to the full reports and those that do/did not meet these initial criteria will be excluded.

2.3.5 Characterising included studies: Phase B

The studies remaining after application of the criteria will be coded for contextual information in each study/report, for example, the country in which the study is

undertaken, types and characteristics of accountability, population, study design (if relevant), types of policy, level of systems (e.g. national, regional, sub-regional), outcomes. Data extracted at this stage will be mapped to describe the literature and refine the initial rough theory in the review (see Appendix 2.7 for a draft coding tool).

2.3.6 Identifying and describing studies: Phase B quality assurance process

The screening process will be carried out independently by all members of the review team. To ensure consistency, we will pilot the inclusion criteria with a set of studies, first with titles and abstracts and subsequently with full texts. Double screening will be carried out on a set of papers before continuing with independent screening. Any disagreements will be resolved by discussion. We will again pilot our data extraction tool and quality-assessment framework with a set of studies: two reviewers will independently assess the quality of each study, and any disagreements will be resolved by discussion.

‘EPPI-Reviewer will be used to manage the review information, for screening coding and synthesis. We will keep a record of decisions made at every stage of the review regarding which studies to include/exclude, methodological clarification and how we are adapting our search strategies.

2.3.7 Focused search: Phase C

A purposive search will be conducted during the synthesis process to identify additional documents essential for developing specific components of the theory. At the in-depth review stage, we will conduct purposive searches to identify literature that might be further helpful in refining middle-range theories (i.e., configurations of mechanisms that give rise to intended outcomes, given the right set of conditions) and addressing the review questions. After the research evidence is mapped and hypothesised, mechanisms are validated and other mechanisms identified, focused searches will be carried out to address particular questions and to seek out additional information about particular mechanisms in different contexts. This process is iterative using a snowballing approach. Search terms will be refined as understanding of literature grows.

2.3.8 Ongoing reference, citation and author tracking to identify the studies most relevant to the elaboration and testing of theory

Throughout the review, the review team will continue to search for all theory-relevant documents.

2.4 In-depth review

The conceptual papers and research evidence identified at the mapping stage will be coded and analysed. Key characteristics of accountability elements, potential mechanisms and reported outcomes in different contexts will be described. A new set of inclusion/exclusion criteria will be developed in consultation with the advisory group to identify a subset of studies for inclusion in the in-depth reviewing stage.

2.4.1 Moving from broad characterisation (mapping) to in-depth review

Realist review allows different types of literature to be employed at different review stages. It begins with the development of theory and proceed by testing and further refining that theory through the elaboration of middle-range theories (i.e., C-M-Os) and testing those theories in the light of findings from the literature. Theory building makes use of all relevant material, including not only primary studies but also theoretical pieces to elaborate theory. The systematic search carried out in the systematic map stage involves identification of primary studies that offer relevant evidence for particular middle-range theories (pertinent to inspection, monitoring, and/or assessment;

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

elucidating the relationship between system-level properties such as criteria and school-level processes and outcomes).

The second stage of the review will provide us with a description of the available primary studies. If there are sufficient primary studies of each accountability element, we will focus on developing a rich understanding of features of context that are essential for specific programme actions related to a particular accountability element to achieve distinctive outcomes. In realist terms we are mapping answers to the generic question:

What is it about this intervention that works, for whom, in what circumstances, in what respects, and why?

Translating that into the review question:

What features of context provide conditions for a specific accountability element to achieve desired outcomes, especially for the poorest and most marginalised?

Depending on the findings from the systematic map, the review may focus on particular interventions or particular accountability elements that have been used in several different contexts (e.g., the implementation of identical designs of EMIS in different countries or parts of countries). Here we will compare and contrast the ‘work’ performed by different or similar mechanisms in order to discern the effect of context on the ability of the intervention to achieve outcomes. Another way is to select a specific mechanism that causes intended outcomes in several different types of accountability interventions, such as ‘setting expectations’, involving the anticipation of reward or consequences. In this approach, we would examine the features of context across cases that allow the mechanism to work effectively.

If there is insufficient research evidence to elucidate middle-range theories for all areas, we will consult with DfID and the Advisory Group around pursuing one of three options:

1. To draw on primary studies from high-income countries (HICs) to supplement LMIC-relevant outcome studies in order to understand particular aspects of relationships amongst conditions, mechanisms and outcomes. We have considered the possibility of drawing on primary literature from other sectors (e.g., health) to derive applicable middle-range theories in LMICs. This is the recommended approach in realist reviews. However, our intended audience of DfID Education Advisers, policy makers and educators in LMICs would probably find the translation of conclusions from other sectors (e.g. health) highly problematic. Thus, we have ruled out cross-sectoral elaboration based on our assumptions around the limited face validity of this work.
2. To consult with DfID and the Advisory Group about the rigour and relevance of the available evidence for each accountability element in relation to the middle-range theories identified. This may involve recommending distinct strategies of synthesis for separate elements if the quality and/or quantity of evidence is highly varied and possibly dropping one or more elements to focus the review on those elements that hold the most promise.
3. To consult with DfID and the Advisory Group about abandoning the review and/or changing the review questions.

The strategies selected will depend on the scope and type of data available.

2.4.2 Detailed description of studies in the in-depth review

Studies included for the in-depth review will be analysed in depth using EPPI-Reviewer. A detailed coding tool will be developed to extract contextual data and assess the quality of the evidence. The first section will be designed to extract data on the name of the programme, setting, context in which the programme is carried out, mechanisms and all outcomes reported in the study. The second part contains codes for theoretical

background, aims and objectives of study, study designs, data collection and analysis approaches. A draft can be found in Appendix 2.7.

2.4.3 Assessing the quality of studies and weight of evidence for the review question

Pawson (2006) and Wong et al. (2013) suggest relevance and rigour as the means of assuring quality in a realist review. Therefore, these two quality assessment criteria will be used to assess quality of each empirical study.

Relevance in these terms is not about whether a study covers a particular topic (e.g. assessment or inspections), but whether it contributes to the elaboration of a hypothesised middle-range theory (i.e., a configuration of mechanisms that cause outcomes under specified conditions) and sufficiently explains why an intervention leads to a particular outcome; in particular specifying the features of context, including the programme actions, that trigger causal processes that lead to intended outcomes. The approach of assessing a study finding, rather than the entire study, is different from other methods that seek to establish the relevance of the entire study to the overarching review question.

Rigour refers to whether ‘a particular inference drawn by the original researcher has sufficient weight to make a methodologically credible contribution to the test of a particular intervention theory’ (Pawson, 2006, p.22). Again, this highlights rigour in relation to a discrete aspect of the study, not overall rigour of the whole study. We will adapt and use existing quality assessment criteria for assessing the rigour of components of the study relevant to the review, including the mixed Methods Appraisal Tool (MMAT) (Pluye et al., 2009), the weight of evidence framework (Gough, 2007) and DfID (2013b). Studies will be assessed according to their methodological quality using the following broad criteria:

1. Theoretical understanding (quality of the reporting of a study’s theoretical and conceptual framework, aims and rationale of research, theory of change)
2. Sampling method (steps taken to minimise selection bias and confounding)
3. The sufficiency of the strategies reported for establishing the reliability and validity of data collection methods
4. The sufficiency of the strategies reported for establishing the reliability and validity of data analysis methods

2.4.4 Synthesis of evidence

Overall approach to and process of synthesis

Information from primary studies will be summarised in a data matrix. We will include summary information (in the form of free text and short verbal descriptions) on key features of interventions and studies, including attributes of participants, settings, interventions, context and mechanisms. We will also describe the relevance and rigour of each study in order to be able to weight and value the results of each study in our summary synthesis. Quantitative results, when appropriate, will be combined statistically and other findings will be synthesised using structured qualitative matrices.

Selection of outcome data for synthesis

Outcomes of interest reported in the primary studies will be included in the analysis. Any outcomes that are not initially listed in the initial rough theories will be used to refine the theories. We are interested in both quantitative outcome data (short-term, intermediate and long-term) and qualitative outcome data (descriptive or process). The selection of outcome data for synthesis will be guided by a subset of review questions developed after the systematic mapping stage.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

Process used to combine/synthesise data

We aim to map and synthesise research evidence (both quantitative and qualitative) against each elements of C-M-Os. The quantitative studies that explore the relationship between accountability, possible mechanisms and outcomes of interest, when appropriate, will be included in statistical meta-analyses. We will calculate effect sizes where possible. This is a proposed elaboration of the publication standards (Wong et al., 2013), as realist reviews derive generalisability through the configurative approaches of theory building or theory testing, not aggregative approaches such as meta-analysis. The findings from qualitative studies will be coded and categorised into matrices. The findings will be grouped by the intervention (monitoring, assessment or inspection) and the outcomes (e.g. system efficiency, service delivery or learning outcomes), comparing the conditions of each accountability element across contexts. These results will be used to differentiate, refine and elaborate the ‘initial rough theory’ into a robust programme theory or theories that are ‘specific enough to generate propositions that can be tested’ (Wong et al., 2013, p. 11) about specific aspects of approaches to inspection, monitoring or assessment, but sufficiently abstract to be applicable across a class of accountability interventions.

2.5 Deriving conclusions and implications

We will have a series of team meetings to discuss the findings from the review and then draft initial conclusions and implications which will be circulated to the advisory group, and review users for their input. The draft review will be peer reviewed by topic experts and DfID policy advisers.

3. References

- Barakat S, Hardman F, Rohwerder B, Rzeszut K (2012) Low-cost private schools in Afghanistan and Pakistan: what evidence to support sustainable scale up. http://r4d.dfid.gov.uk/pdf/outputs/systematicreviews/Low_cost_schools_2012Barakat_protocol.pdf (accessed 8 July 2014).
- Barber M (2004) The virtue of accountability: system redesign, inspection, and incentives in the era of informed professionalism. *Journal of Education*, 85 (1): 7-38.
- Best M, Knight P, Lietz P, Lockwood C, Nugroho D, Tobin M (2013) *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries*. Final report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Bold T, Svensson J, Gauthier B, Mæstad O, Wane W (2011) *Service delivery indicators: pilot in education and health care in Africa: technical appendix*. <http://www.cmi.no/file/?1472> (accessed 7 June 2014).
- Bouckaert G, Halligan J (2007) *Managing performance: international comparisons*. London: Routledge.
- Carr SC, Leggatt-Cook C (2011) *What is the evidence of the impact of increasing salaries on improving the performance of public servants including teachers, nurses and mid-level occupations in LMIC: Is it time to give pay a chance?* http://r4d.dfid.gov.uk/PDF/Outputs/SystematicReviews/Q17_systematic_review_submitted1.pdf (accessed 8 July 2014).
- Cheng YC (1993) *Conceptualization and measurement of school effectiveness: an organizational perspective*. Atlanta, GA: AERA.
- De Grauwe A (2001) *School supervision in four African countries*. Paris: UNESCO IIEP.
- De Grauwe A (2007) Transforming school supervision into a tool for quality improvement. *International Review of Education*, 53: 709-714.
- De Grauwe A (2008) *School monitoring systems and their impact on disparities: background paper prepared for the Education for All Global Monitoring Report 2009*. Paris: UNESCO.
- De Grauwe A, Lugaz C (2007) District education offices in French-speaking West Africa: autonomy, professionalism and accountability. *Prospects*, 37 (1): 113-125.
- Dembélé M, Oviawe J (2007) Introduction: quality education in Africa: international commitments, local challenges and responses. *International Review of Education*, 53 (5-6): 473-483.
- DfID (2013a) *Education position paper: improving learning, expanding opportunities*. London: DfID. <https://www.gov.uk/government/publications/education-improving-learning-expanding-opportunities> (accessed 20 May 2014).
- DfID (2013b) *How to note: assessing the strength of evidence*. London: Department for International Development.
- DfID (2013c) Additional information on questions for bidders. London: DfID.
- Eddy Spicer, DH (2012a) Modalities of authority and the socialisation of the school in contemporary approaches to educational change. In: Daniels H (ed.) *Vygotsky and sociology*. Abingdon, New York: Routledge, pages 114-134.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

Eddy Spicer DH (2012b) The emergent logics of organisational socialisation: a review of recent literature on interorganisational networks in education in England. Paper presented at the *Annual Meeting of the American Educational Research Association*, Vancouver, 14 April.

Eddy Spicer DH, Crawford M, Earley P, James C (2014) *Effectively managing headteacher performance: final report*. Nottingham, UK: National College for Teaching and Leadership, Department for Education.

Ehren MCM, Altrichter H, McNamara G, O'Hara J (2013) Impact of school inspections on teaching and learning: describing assumptions on causal mechanisms in seven European countries. *Educational Assessment, Evaluation and Accountability*, 25 (1): 3-43.

Ehren MCM, Leeuw FL, Scheerens J (2005) On the impact of the Dutch Educational Supervision Act: analyzing assumptions concerning the inspection of primary education. *American Journal of Evaluation*, 26 (1): 60-77.

Ehren MCM, Honingh M (2011) Risk-based school inspections in the Netherlands: a critical reflection on intended effects and causal mechanisms. *Studies in Educational Evaluation*, 37 (4): 239-248.

Ehren MCM, Visscher AJ (2006) Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, 54 (1): 51-72.

Elmore RF (2002) Bridging the gap between standards and achievement: the imperative for professional development in education.

http://www.ashankerinst.org/Downloads/Bridging_Gap.pdf (accessed 4 November 2003).

Fullan M (2010) The big ideas behind whole system reform. *Education Canada*, 50: 24-27.

Gough D (2007) Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22 (2): 213-228.

Gough D, Oliver S, Thomas J (2012) *An introduction to systematic reviews*. London: SAGE.

Haladyna TM, Nolen SB, Haas NS (1991) Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20 (5): 2-7.

Hanushek EA, Woessmann L (2007) *The role of school improvement in economic development*. München: Center for Economic Studies and Institute for Economic Research.

Hatch T (2013) Beneath the surface of accountability: answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change*, 14 (2): 113-138.

Hooge E, Burns T, Wilkoszewski H (2012) *Looking beyond the numbers: stakeholders and multiple school accountability*. OECD Education Working Papers. Geneva: OECD.

Joshi A, Hossain N, Sabates-Wheeler R, Clark J, Chopra D (2011) *The impact of social accountability initiatives on improving the delivery of public services: a systematic review of four intervention types*.

<http://www.campbellcollaboration.org/lib/project/216/> (accessed 8 July 2014).

Kerr, KA, Marsh, JA, Ikemoto, GS, Darilek H, Barney H (2006) Strategies to promote data use for instructional improvement: actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112 (4): 496-520.

Klerks M (2013) The effect of school inspections: a systematic review. Paper presented at: ORD, Wageningen, The Netherlands, 20-22 June 2012. <http://schoolinspections.eu/wp-content/uploads/downloads/2013/12/ORD-paper-2012-Review-Effect-School-Inspections-MKLERKS.pdf> (accessed 7 June 2014).

- Lieberman E, Posner D, Tsai, L (2012) *Does information lead to more active citizenship? An evaluation of the impact of the Uwezo Initiative in Kenya*. Draft paper. Dar es Salaam: Twaweza.
- Luginbuhl R, Webbink D, De Wolf IF (2009) Do school inspections improve primary school performance? *Educational Evaluation and Policy Analysis*, 31 (3): 221-237.
- Mehrens WA, Kaminski J (1989) Methods for improving standardized test scores: fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice*, 8 (1): 14-22.
- Merton RK (1968) *Social theory and social structure*. New York: Free Press.
- Mincer JA (1974) *Schooling, experience, and earnings*. Cambridge, MA: NBER.
- Nelson R, Ehren M (2014) Review and synthesis of evidence on the (mechanisms of) impact of school inspections. <http://schoolinspections.eu/wp-content/uploads/downloads/2014/02/Review-and-synthesis-of-evidence-on-the-mechanisms-of-impact-of-school-inspections.pdf> (accessed 1 April 2014).
- O'Neill O (2002) *A question of trust*. Cambridge: Cambridge University Press.
- O'Neill O (2013) Intelligent accountability in education. *Oxford Review of Education*, 39(1): 4-16.
- Pawson R (2006) *Evidence-based policy: a realist perspective*. London: Sage.
- Pawson R, Greenhalgh T, Harvey G, Walshe K (2004) *Realist synthesis: an introduction*. Manchester: University of Manchester, ESRC Research Methods Programme. <http://www.ccsr.ac.uk/methods/publications/RMPmethods2.pdf> (accessed 1 April 2014).
- Pawson R, Greenhalgh T, Harvey G, Walshe K (2005) Realist review: a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy*, 10: 21-34.
- Pawson R, Tilley N (1997) *Realistic evaluation*. London: Sage.
- Petrosino A, Morgan C, Fronius T, Tanner-Smith E, Boruch R (2013) *Interventions in developing nations for improving primary and secondary school enrolment of children: a systematic review*. http://www.3ieimpact.org/media/filer/2013/10/28/interventions_in_developing_nations_for_improving_primary_and_secondary_school_enrolment_of_children_a_systematic_review.pdf (accessed 8 July 2014).
- Pluye P, Gagnon M-P, Griffiths F, Johnson-Lafleur J (2009) A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*, 46 (4): 529-546.
- Popham WJ (1991) Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice*, 10 (4): 12-15.
- Powell M (2006) Rethinking education management information systems: lessons from and options for less developed countries. InfoDev Working Paper 6. <http://www.infodev.org/articles/rethinking-education-management-information-systems-lessons-and-options-less-developed> (accessed 6 September 2013).
- Rosenkvist MA (2010) *Using student test results for accountability and improvement: a literature review*, OECD Education Working Papers, No. 54. <http://dx.doi.org/10.1787/5km4htwzvb30-en> (accessed 7 June 2014).
- Rosenthal L (2004) Do school inspections improve school quality? *Economics of Education Review*, 23 (2): 143-151.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

Sakellariou C (2006) Education policy reform, local average treatment effect and returns to schooling from instrumental variables in the Philippines. *Applied Economics*, 38 (40): 473-481.

Sayer RA (2000) *Realism and social science*. London: Sage.

Scepanovic D, Lazarevic B, Wassenmiller A (2010) Reform in progress: current trends and concerns in developing education management information systems in the south east European countries. Paper presented at: *35th Annual European Studies Conference*, Omaha, USA, 8 October.

Scheerens J (1992) *Effective schooling: research, theory and practice*. London: Cassell.

Scheerens J (2000) *Improving school effectiveness: fundamentals of educational planning*. Paris: Unesco, International Institute for Educational Planning.

Schildkamp K, Ehren M, Lai MK (2012) Editorial: data-based decision making around the world: from policy to practice to results. *School Effectiveness and School Improvement*, 23 (2): 123-131.

Stecher BM (2002) Consequences of large-scale, high-stakes testing on school and classroom practices: tests and their use in test-based accountability systems. In: Hamilton LS, Stecher BM, Klein SP (eds) *Making sense of test-based accountability in education*. Santa Monica: Rand Corporation. http://www.rand.org/pubs/monograph_reports/MR1554/ (accessed 7 June 2014).

Thomas J, Brunton J, Graziosi S (2010) *EPPI-Reviewer 4.0: software for research synthesis*. EPPI-Centre software. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Tilly C (2008) *Contentious performances*. Cambridge: Cambridge University Press.

UNESCO (2008) *Overcoming inequality: why governance matters*. Paris and Oxford: UNESCO and Oxford University Press.

Vegas E, Petrow J (2008) *Raising student learning in Latin America: the challenge for the 21st century*. Washington, DC: World Bank.

Wayman JC, Stringfield S (2006) Data use for school improvement: school practices and research perspectives. *American Journal of Education*, 112 (4): 463-468.

Weiss CH (1998) *Evaluation: methods for studying programs and policies*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Westhorp G, Walker B, Rogers P (2012) *Under what circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor? A realist synthesis*. Protocol. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Westhorp G, Walker DW, Rogers P, Overbeeke N, Ball D, Brice G (2014) *Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: a realist review*. Technical report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3450> (accessed 8 July 2014).

Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R (2013) RAMESES publication standards: realist syntheses. *BMC Medicine*, 11: 21. <http://www.biomedcentral.com/content/pdf/1741-7015-11-21.pdf> (accessed 7 June 2014).

World Bank (2011) *Learning for all: investing in people's knowledge and skills to promote development: World Bank education strategy 2020*. <http://sdindicators.org/wp-content/uploads/2013/07/World-Bank-Education-Strategy.pdf> (accessed 7 June 2014).

Appendices

Appendix 1.1: Authorship of this report

Authors

David Eddy Spicer, Melanie Ehren, and Meena Khatwa, London Centre for Leadership in Learning

Mukdarut Bangpan, EPPI-Centre, Institute of Education, University of London

Review group

Louise Banham, Ed Barnett and Jessica Vince, UK Department for International Development

Kelly Dickson, EPPI-Centre, Institute of Education, University of London

Advisory group

Anton De Grauwe, IIEP-UNESCO

Thomas Hatch, Teachers College, Columbia University

Luis Huerta, Teachers College, Columbia University

Pantalee Kapichi, UNICEF Tanzania

Dennis Shirley, Lynch School of Education, Boston College

Table A1.1: Names put forward for the Advisory Group

Name	Title	Organisation	Expertise
Pantalee Kapichi	Education Specialist	UNICEF Tanzania	Educational system, Tanzania
Charlotte Jones	Senior Education Consultant	CfBT	Mobile report cards, India
Marguerite Clarke	Senior Education Specialist	World Bank	Standardised testing in LICs
Anton de Grauwe*	Head, Technical Assistance	IIEP-UNESCO	Inspections and monitoring in LICs
Thomas Hatch* Luis Huerta*	Associate Professor Associate Professor	Teachers College, Columbia University	Accountability in HICs
Rukmini Banerji	Director	ASER Centre/Pratham	Education reform Accountability in LMICs
Michael Fullan	Emeritus Professor Author/Consultant	OISE/University of Toronto	Education reform
Barbara Bruns	Lead education economist, Latin America and the Caribbean	World Bank	Accountability in LMICs

*Academic advisers involved in validation of initial rough theory and elaboration of key mechanisms.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

This report should be cited as:

Eddy Spicer D, Ehren M, Bangpan M, Khatwa M (2014) *Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised? A realist synthesis of school accountability in low- and middle-income countries*. Protocol. London: EPPI-Centre, Social Science Research Centre, Institute of Education, University of London.

Contact details

London Centre for Leadership in Learning
Institute of Education, University of London
20 Bedford Way
London WC1H 0AL
Tel: +44(0)7612 6038
Email: d.eddyspicer@ioe.ac.uk

Conflicts of interest

None of the review team members has a financial interest in this project. Researchers have been involved in relevant primary research on the topic.

Acknowledgements

The review team gratefully acknowledges the financial support of the UK Department for International Development, and guidance from DfID staff, including Louise Banham, Ed Barnett, Laura Rivkin and Jessica Vince. At the Institute of Education, we appreciate the substantive support and input from the EPPI-Centre, especially Kelly Dickson and Sandy Oliver. We are grateful for administrative support from Chris To and Claire Phillips, London Centre for Leadership and Learning. Finally, we are indebted to Gill Westhorp of Community Matters, Australia, for her suggestions and comments on drafts of the protocol.

Appendix 2.1: Advisory Group terms of reference

Dear [name] ,

We have been commissioned by the UK Department for International Development to undertake a systematic review of school accountability in low- and middle-income countries. I'm writing to invite you on behalf of the Institute of Education to be a project advisory group member for this systematic review.

The agreed question that the review will address is:

Under what conditions do the following elements of an education system improve system efficiency, service delivery and learning outcomes in low- and middle-income countries, especially for the poorest and most marginalised?

1. Monitoring systems, including using administrative data systems (e.g. EMIS) as well as more targeted monitoring mechanisms.
2. Inspection systems
3. Assessment systems

This review aims to be of use to DfID Education Advisers as well as policymakers, politicians, civil servants, and educational leaders. You will find accompanying this letter the detailed review protocol that has been approved by DfID and is about to be published by the EPPI-Centre at the Institute of Education (<http://eppi.ioe.ac.uk/>).

You have been recommended to us by several colleagues, both within DfID and at the Institute. This systematic review would greatly benefit from your taking an advisory role in its planning, execution and dissemination of results because of your expertise in school accountability and the dynamics of system-level educational change in low-income countries.

We very much hope that you will be able to help us with this important review and would appreciate having an indication of your interest by 14th July.

By agreeing to be an advisory group member you would be committing to taking part at key points in the systematic review process. Your participation involves offering feedback in writing or by phone/Skype to documents we will send you and, depending on your availability, taking part in a group discussion by telephone or web-based conference.

The stages at which we plan to elicit your feedback would be:

1. Mapping stage - September 2014

At this stage we will provide you with a structured overview of the research literature that has been found on the topic. We will ask for your review of key aspects of the review for intensive analysis and your input on whether the three accountability elements should be considered separately and in what depth.

2. Interim findings - December 2014

At this stage we will provide an update on the systematic review and seek your feedback on what we have found so far, as well as implications for relevance in your setting or settings known to you.

3. Final report - February 2014

At this stage we will provide you with a draft of the systematic review and seek your feedback on clarity, practicality and utility of the report. We will also ask for your input on dissemination and how best to communicate our findings.

At each of these three stages, we will organise an online forum for contributing feedback.

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

We plan to hold one online meeting for each phase that will be conducted by telephone or web conferencing.

Please let us know of your interest in taking on this important role in the development of our review. We are very interested in making this review useful and practical, while maintaining high standards of rigour and quality. Your contribution to ensuring this review is carried out in a way that is useful, applicable and to a high standard is a crucial part of the success of this systematic review. We appreciate the time and energy it takes to take part in advisory groups.

Honorarium

If you are working or are self-employed we will provide you with a £150 honorarium in total to participate in 3 project consultations, which will be subject to tax and national insurance. This is £50 for each consultation in which you participate. If you are receiving benefits please alert us to this and we will ensure you are paid the maximum you are entitled to receive according to the benefit rules that apply to you.

Expenses

Due to our restricted budget, we are unable to provide payment for expenses.

Receipt

All participation fees will be made directly to you and processed after the meetings have taken place.

Being an advisory group member for this systematic review will also mean a full commitment from us to support your needs in order to ensure your full involvement and participation. We will contact you nearer the time of our first online meeting to discuss any requirements you might have.

If you have any questions please do not hesitate to contact me.

Yours sincerely

Dr David Eddy Spicer

Principal Investigator

+44(0)20 7612 6038

d.eddyspicer@ioe.ac.uk

DfID Systematic Review Team

David Eddy Spicer

Melanie Ehren

Mukdarut Bangpan

Meena Khatwa

Appendix 2.2: Timeline

	Start date	End date
2014		
Registration of title with EPPI Centre	1 Feb	1 Feb
STAGE 0 - Clarifying the review questions & approach		
Protocol - Preparation & clarification of review questions	1 Feb	28 Mar
Protocol - submit to EPPI Centre for review	2 Apr	8 Apr
Protocol - DfID and external review (six weeks)	8 Apr	21 May
STAGE 1 - Mapping primary studies to mechanisms		
Searching for primary studies - all elements	1 Mar	21 May
Screening of primary studies - all elements	17 Apr	ongoing
Elaborated protocol with map of key mechanisms per element*	20 May	15 Sep
Advisory Group - 1 st meeting	15 Aug	15 Sep
STAGE 2 - Elaborating mechanisms		
Search for additional sources for each key mechanism	1 Jul	1 Nov
STAGE 3 - Quality appraisal	1 Jul	1 Nov
STAGE 4 - Extracting data to elaborate programme theories	30 Sep	30 Nov
STAGE 5 - Synthesising data	30 Sep	31 Dec
Advisory Group - 2 nd meeting*	1 Nov	1 Dec
STAGE 6 - Dissemination		
Preparation of draft report	1 Dec	15 Jan
2015		
DfID and External review of draft report (allow six weeks)	15 Jan	28 Feb
Revision of draft report	1 Mar	30 Mar
Preparation of evidence brief for policy	1 Jan	30 Mar
Advisory Group - 3 rd consultation	1 Feb	1 Mar
Final revisions	1 Mar	30 Mar
Publication of final report and evidence brief	1 Apr	1 Apr

* Decision point: Consult DfID/AG/EPPI-Centre support teams about design of systematic review.

Appendix 2.3: iterative review stages

Dates	Stages Outputs	0 Clarifying review questions and process	1 Mapping primary studies to mechanisms	2 Elaborating middle-range theories	3 Quality appraisal	4 Data extraction	5 Data synthesis	6 Disseminating findings
		Model key programme theories and related mechanisms	Background familiarization search					
		Prioritize key mechanisms for investigation	Define incl/excl criteria					
21 May	Protocol - initial	External review of protocol						
		Refine selection of key mechanisms	Search for empirical studies to test mechanisms	Search for sources of programme theories and related mechanisms	Assessment of relevance of primary inquiry to inform model	Annotation, note-taking on candidate mechanisms		Incorporate feedback from EPPi-Centre, Advisory Group, DfID around analytic and policy focus
		Formalise model of key mechanisms and programme theories				Collation of materials from selected primary studies	Prime foci of synthesis selected and formalised	
	Protocol - elaborated	Consultation on which emerging lines of inquiry to follow OR resort to realist-informed study						
15 Sep	1 st Advisory Group				Assessment of rigour of primary data to test	Detailed reportage of evidence from	Absorbing primary materials into developing	

Dates	Stages Outputs	0 Clarifying review questions and process	1 Mapping primary studies to mechanisms	2 Elaborating middle-range theories	3 Quality appraisal	4 Data extraction	5 Data synthesis	6 Disseminating findings
	Meeting				mechanisms	each study	synthesis	
				Search for further empirical studies consequent on revisions to model	Further assessment of rigour as each study enters synthesis	Differential reportage of evidence from each study	Juxtaposing, adjudicating, reconciling, consolidating and situating further evidence	
1 Dec	2 nd Advisory Group Meeting	Preparation of draft report					Revised model of the complex and interrelated elements of programme theory	Summary theory to initiate process of 'thinking through' future implementation decisions
15 Jan 2015	Draft report	External review of draft report and policy brief						
		Incorporate feedback and revise					Refine model of the complex and interrelated elements of programme theory	
1 Mar	3 rd Advisory Group Meeting							Consultation on dissemination strategy
1 Apr	Final report	Preparation of final report						

Source: Adapted from Pawson (2006, p. 103)

Appendix 2.4: Inclusion and exclusion criteria

AT MAPPING STAGE: Initial inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
<p>Types of intervention:</p> <p>Studies or reports that investigate or explore accountability (monitoring, assessment and/or, inspection) of education system</p>	<p>Types of intervention:</p> <p>Studies or papers focusing on accountability that do NOT clearly state or make reference to a sub-national, national, regional and/or international level of an assessment, inspection or monitoring programme</p>
<p>Geographical location:</p> <p>Conducted in low- and lower-middle-income countries according to World Bank classification⁴</p>	<p>Geographical location:</p> <p>NOT conducted in low- and lower-middle-income countries according to World Bank classification</p>
<p>Setting:</p> <p>Targeting primary, secondary and/or compulsory education</p>	<p>Setting:</p> <p>NOT designed for primary, secondary, and/or compulsory education</p>
<p>Types of studies:</p> <p>All types of study designs, policy and theoretical/conceptual framework documents</p>	<p>Types of studies:</p> <p>No restriction</p>
<p>Language:</p> <p>Published in English</p>	<p>Language:</p> <p>NOT published English</p>
<p>Date:</p> <p>Published in and after 1990</p>	<p>Date:</p> <p>Published before 1990</p>

⁴ <http://data.worldbank.org/about/country-classifications> (accessed 15 February 2014)

Appendix 2.5: Search strategy for electronic databases

Table A2.5: Key search terms used in the review

Key aspects of the reviews	Search terms and synonyms
<p>Accountability, inspection, monitoring, assessment</p>	<p>accountability, educational accountability, educational quality, benchmarking, government role, quality assurance, quality control, school accounting, school-based management, standards-based accountability, quality management</p> <p>Assessment</p> <p>Alternative assessment, assessment program, educational assessment, cognitive assessment system, cognitive measurement, cognitive tests, criterion referenced tests, achievement tests, educational tests & measurements, examinations, exit examinations, high stakes tests, measurement, measures (individuals), national assessment, national competency tests, national competency-based educational tests, curriculum based assessment, performance based assessment, standardised student testing, national testing; norm referenced test, standardized assessment system, standardised tests, testing, state tests, student evaluation, teaching to the test, test coaching, test bias, testing effects, testing programs, test use, value added assessment</p> <p>Monitoring</p> <p>Administrative organization, educational monitoring, administrator evaluation, bureaucracy, database management systems, decision support systems, educational indicators, information management, information systems, information utilization, internal evaluation, management information systems, management systems, performance information, performance factors, performance management, performance indicators, program monitoring, progress monitoring, school performance, progress reporting, recordkeeping, records, school-level data, school self-evaluation, SSE, self-assessment, student evaluation of teacher performance, teacher evaluation, total quality management, database management systems, school monitoring, EMIS, school performance data, monitoring systems, school governance, school autonomy, school efficiency, national information systems</p> <p>Inspection</p> <p>Inspection, administrator evaluation, audits (verification), external evaluation, external review, inspection & review, quality control, quality review, review, school evaluation, school inspections, school inspectors, school supervision, school visitation, supervision, supervisor qualifications, supervisor- supervisee relationship, supervisors, teacher supervision, evaluation, institutional evaluation, state supervisors, inspectorate, school evaluation</p>
<p>Developing countries</p>	<p>Developing nations, low-/lower-income countries, less-developed countries, Third world countries, less-developed economies, and country names as classified by World Bank as Low and Middle income countries</p>
<p>Primary and secondary education</p>	<p>Secondary school curriculum, secondary education, secondary schools, secondary school education, secondary school students, junior high schools, high schools, elementary schools, elementary school students, elementary school education, elementary school curriculum, primary education, compulsory education, elementary education</p>

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

Example of search strategy

1. Accountability terms

a) Free text

accountability or (monitoring N1 activit*) or (monitoring N3 system*) OR (progress W1 monitoring) OR (monitoring W1 mechanism*) OR (monitoring W! process*) OR (monitoring W1 procedure*) OR (targeted W1 monitoring) OR (inspection*) OR (inspector*) OR (inspectorate) OR (supervis*) OR (EMIS) OR (education W1 management W1 information W1 system) OR (performance W1 review*) OR (financial W2 management) OR (audit*) OR (budget*) OR (education* W1 finance) OR (Total W1 quality W1 management) OR (quality W1 assurance) OR (quality W1 control) OR (information W1 management) OR (database W1 management) OR (information W1 system*) OR (decision W1 support W1 system) OR (standardised W1 test*) OR (standardized W1 test*) OR (budget W1 tracking) OR (appraisal N1 process*) OR (management N2 education) OR (competency-based W1 education) OR (competency W1 based W1 education) OR (performance W1 based) OR (result* W1 based) OR (outcome-based) OR (outcome W1 based) OR (alternative W1 assessment) OR (curriculum W1 based W1 assessment) OR (curriculum-based W1 assessment) OR (educational W1 assessment) OR (assessment N2 procedure) OR (standardised W1 assessment) OR (standardized W1 assessment) OR (informal W1 assessment) OR (assessment W1 system*) OR (assessment W1 mechanism*) OR (assessment W1 process*)

b) Indexed terms

((DE "Accountability") OR (DE "Assessment Centers (Personnel)") OR (DE "Performance Based Assessment" OR DE "Alternative Assessment") OR (DE "Audits (Verification)" OR DE "Financial Audits") OR (DE "Budgeting" OR DE "Program Budgeting" OR DE "Educational Finance") OR (DE "Curriculum Based Assessment") OR (DE "Educational Assessment") OR (DE "Performance Tests") OR (DE "Competency Based Education") OR (DE "Outcome Based Education") OR (DE "Evaluation") OR (DE "Progress Monitoring") OR (DE "Informal Assessment") OR (DE "Total Quality Management" OR DE "Quality Control" OR DE "Inspection") OR (DE "Quality Assurance") OR (DE "Supervisors" OR DE "State Supervisors" OR DE "Student Teacher Supervisors" OR DE "Supervision" OR DE "Practicum Supervision" OR DE "School Supervision" OR DE "Teacher Supervision") OR (DE "Information Management" OR DE "Knowledge Management" OR DE "Database Management Systems" OR DE "Management Information Systems" OR DE "Decision Support Systems" OR DE "Information Systems"))

2. Setting: primary and secondary education

a) Free text

(primary W3 school*) OR (elementary W1 school*) OR (high W1 school*) OR (secondary W3 School*) OR (Secondary W1 Teach*) OR (secondary W1 education) OR (primary W1 education) OR (compulsory W1 education) OR (elementary W1 education)

b) Indexed terms

DE "Grade 10" OR DE "Grade 11" OR DE "Grade 12" OR DE "Grade 9" OR DE "High Schools" OR DE "Vocational High Schools" OR DE "Junior High Schools" OR DE "Secondary School Curriculum" OR DE "Secondary School Mathematics" OR DE "Secondary School Science" OR DE "Secondary School Students" OR DE "High School Students" OR DE "Junior High School Students" OR DE "Secondary School Teachers" OR DE "Secondary Schools" OR DE "High Schools" OR DE "Junior High Schools" OR DE "Secondary Education" OR DE "College Preparation" OR DE "Compulsory Education" OR DE "Elementary Secondary Education" OR DE "Elementary Education" OR DE "Secondary Education" OR DE "Primary Education" OR DE "Elementary School Students" OR DE "Elementary School Teachers" OR DE "Grade 1" OR DE "Grade 2" OR DE "Grade 3" OR DE "Grade 4" OR DE "Grade 5" OR DE "Grade 6" OR DE "Grade 7" OR DE "Grade 8" OR DE "Intermediate Grades" OR DE "Elementary School Curriculum" OR DE "Elementary School Mathematics" OR DE "Elementary School Science" OR DE "FLES" OR DE "Elementary Education" OR DE "Adult Basic Education" OR DE "Primary Education" OR DE "Elementary Schools"

3. School and setting terms

a) Free text

(school W1 evaluation) OR (school W1 efficiency) OR (school W1 governance) OR (school W1 autonomy) OR (school N3 self-evaluation) OR (school N3 self W1 evaluation) OR (school W1

accounting) OR (school-based W1 management) OR (school W1 based W1 management) OR (school N3 monitoring) OR (school N3 assessment) OR (primary W1 education N3 monitoring) OR (primary W1 education N3 assessment) OR (elementary W1 education N3 monitoring) OR (secondary W1 education N3 assessment) OR (secondary N3 monitoring)

b) Indexed terms

DE "School Based Management" OR DE "School Accounting" OR DE "School Effectiveness"

4. Geographical locations

a) Low-income countries

(afghanistan or Bangladesh or Benin or (Burkina Faso) or Burundi or Cambodia or (Central African Republic) or Chad or Comoros or (Democratic Republic of the Congo) or Eritrea or Ethiopia or Gambia or Guinea or Guinea-Bissau or Haiti or Kenya or (the Democratic People's Republic of Korea) or (North Korea) or (Kyrgyz Republic) or Liberia or Madagascar or Malawi or Mali or Mozambique or Myanmar or Nepal or Niger or Rwanda or (Sierra Leone) or Somalia or (South Sudan) or Tajikistan or Tanzania or Togo or Uganda or Zimbabwe)

b) Lower-middle income countries

Armenia or Bhutan or Bolivia or Cameroon or (Cabo Verde) or India or Kiribati or Kosovo or Laos or (the Lao People's Democratic Republic) or Lesotho or Samoa or (the Democratic Republic of São Tomé and Príncipe) or (São Tomé and Príncipe) or Senegal or (Solomon Islands) or (Sri Lanka) or (The Democratic Republic of the Congo) or (DR Congo) or (Congo-Kinshasa) or DROC or RDC or (Ivory Coast) or (Côte d'Ivoire) or (the Republic of Côte d'Ivoire) or Djibouti or Egypt or (El Salvador) or Georgia or Ghana or Guatemala or Guyana or Honduras or Indonesia or Mauritania or (Federated States of Micronesia) or Micronesia or Moldova or Mongolia or Morocco or Nicaragua or Nigeria or Pakistan or (Papua New Guinea) or Paraguay or Philippines or Sudan or Swaziland or (the Syrian Arab Republic) or Syria or (Timor-Leste) or Ukraine or Vanuatu or Vietnam or (West Bank and Gaza) or (the Republic of Yemen) or Yemen or Zambia or (South Africa) or Palestine

c) Free text

TI or AB (developing W1 nation*) OR (developing W1 countr*) OR (developing W1 world) OR (developing W1 econom*) OR (less* W1 developed W1 countries) OR (less* W1 developed W1 nation*) or (less* W1 developed W1 world) OR (less* W1 developed W1 econom*) OR (underdeveloped W1 countr*) OR (underdeveloped W1 nation*) OR (underdeveloped W1 world) OR (underdeveloped W1 economies) OR (under W1 developed W1 nation*) OR (under W1 developed W1 world) OR (under W1 developed W1 economies) OR (low* W1 income W1 countries) OR (low* W1 income W1 nation*) OR (low* W1 income W1 econom*) OR (low* W2 middle W3 countr*) OR (LMIC) OR (LMICs) OR (LLMIC) OR (LLMICs) OR (third W1 world) OR (underserved W1 countr*) OR (underserved W1 nation*) OR (deprived W1 countr*) OR (deprived W1 nation*) OR (deprived W1 world) OR (poor* W1 countr*) OR (poor* W1 nation*)

Appendix 2.6: Sources

The following sources will be used:

1. A range of bibliographic and specialist education, social and economic databases.
 - Education Resources Information Centre (ERIC)
 - Australian Education Index (AEI)
 - British Education Index (BEI)
 - Econlit
 - International Bibliography of Social Science (IBSS)
 - Social Science Citations Index (SSCI)
 - Social Service Abstracts (SSA)
 - PsycINFO
 - Sociological Abstracts
 - 3ie Database of impact evaluations: www.3ieimpact.org/database_of_impact_evaluations.html
 - AfricaBib: Bibliography of Africana Periodical Literature Database: www.africabib.org/
 - Africa Journals Online (AJOL): www.ajol.info/
 - Bangladesh Journals Online (Bangla JOL): www.banglajol.info/
 - Bioline International: www.bioline.org.br/
 - East View Information Service Online Databases: www.eastview.com/
 - IDEAS Economics and Finance Database (RePEc): <http://ideas.repec.org/>
 - Indian Citation Index (ICI): www.indiancitationindex.com/
 - JOLIS library catalogue: <http://www.unsceb.org/content/world-bank-jolis-library-catalogue>
 - Nepal Journals online (NepJOL): www.nepjol.info/
 - OpenGrey: www.opengrey.eu/
 - Philippines Journal Online (PhilJol): www.philjol.info/philjol/index.php
 - SciDev Net (Science and Development Network): www.scidev.net/en/
 - Thai Research: <http://thesis.stks.or.th/>
 - National Technical Information Service (NTIS): www.ntis.gov/
 - Social Science Research Network (SSRN): papers.ssrn.com/sol3/DisplayAbstractSearch.cfm
 - Education Research Global Observatory: <http://www.ergobservatory.info/index.html>
 - EducatiON-Line: http://www.leeds.ac.uk/bei/COLN/COLN_default.html

2. Relevant systematic reviews

Reference lists of reviews and literature reviews: Westhorp et al. (2014); Klerks (2013); Petrosino et al. (2013); Barakat et al. (2012); Carr and Leggatt-Cook (2011); Joshi et al. (2011).

3. Key websites

- National Bureau of Economic research: <http://www.nber.org>
- UNESCO: <http://www.unesco.org>
- International Institute for Education Planning: <http://www.iiep.unesco.org>
- Asian Development Bank: <http://www.adb.org>
- African Development Bank: <http://www.afdb.org/en/>
- Association for the development of Education in Africa: <http://www.adeanet.org/portalv2/>
- Australian Council for Education Research: <http://www.acer.edu.au/>
- Inter-American Development Bank: <http://www.iadb.org/en/inter-american-development-bank,2837.html>
- UNDP: <http://www.undp.org/content/undp/en/home.html>
- World Bank: <http://www.worldbank.org/>
- USAID: <http://www.usaid.gov/>
- AusAID: <http://aid.dfat.gov.au/Pages/home.aspx>
- Overseas Development Institute: <http://www.odi.org.uk/>
- Institute for Fiscal Studies: <http://www.ifs.org.uk/>
- British Library for Development Studies: <http://blds.ids.ac.uk/>
- ELDIS: <http://www.eldis.org/>
- Poverty Action Lab: <http://www.povertyactionlab.org/>
- Institute of Development Studies: <http://www.ids.ac.uk>
- The Future of Children: <http://futureofchildren.org/>

4. Expert contacts

We will contact experts working in the area of accountability elements (assessment, inspection, and monitoring) in education to recommend any potential relevant literature in the field.

5. Google and Google Scholar

We will conduct supplementary search for relevant studies through Google and Google Scholar after searching on all other search sources has been completed. This includes forward reference checking of relevant studies.

Appendix 2.7: Draft coding tool

School accountability systematic review: Draft coding tool for mapping (v4 - 15 Jul)

Category	Codes
Publication [Study]	<p>a) Published article in peer reviewed journal <i>This could be a primary study, systematic review, or conceptual/discussion paper</i></p> <p>b) Research report <i>(e.g. working papers, research series)</i></p> <p>c) Policy-relevant document <i>(e.g. policy briefs, statistics, inception reports)</i></p> <p>d) Dissertation</p> <p>e) Conference paper</p> <p>f) Books/Book chapter</p>
Locations of implementation/discussion	<p>a) Code the country/countries in which the policy or program operates(as stated by the authors)</p> <p>b) Not applicable-(e.g. systematic reviews/literature reviews)</p>
Years of implementation Code the years in which the program or policy was implemented and the years covered by the research/evaluation if specified	<p>a) Years of policy / program implementation</p> <p>b) Not stated</p>
Policy or Program name Code the name of the program if specified	<p>a) Please specify</p>
Study approach and/or design Code the relevant features of study approach or study design. Note any significant implications of design for realist analysis (i.e., poorly aligned/well aligned)	<p>a) Randomised controlled trial <i>Each participant is randomly has the same chance of being in the intervention and comparison group</i></p> <p>b) Non-randomised controlled trial/controlled before and after study <i>Study includes intervention and comparison groups, with before and after data for both groups</i></p> <p>c) Retrospective controlled before and after study <i>Data from large repeated surveys are used to retrospectively construct intervention and comparison groups, with before and after data for both groups</i></p> <p>d) Simple comparison study <i>Intervention and comparison groups, only one data point also referred to as with and without study</i></p>

	<p>e) Before and after study <i>One group of study before and after data</i></p> <p>f) Non comparison evaluation <i>Only one data point-for example, post test only, cross-sectional study</i></p> <p>g) Modelling study <i>Based on theoretical/modelled events not real ones</i></p> <p>h) Study using qualitative data collection methods <i>For example, interviews, focus groups, observations</i></p> <p>i) Mixed methods design <i>A study employs more than one methods above (a-h)</i></p> <p>j) Not empirical paper <i>(e.g. discussion piece, policy brief, conceptual paper, statistics document)</i></p>
Accountability elements	<p>a) Accountability in general</p> <p>b) Monitoring</p> <p>c) Inspection/supervision</p> <p>d) Assessment</p> <p>d) More than one element (Please specify)</p>
Schooling level	<p>a) Primary</p> <p>b) Secondary</p> <p>c) Both</p>
School type	<p>a) Government</p> <p>b) Private</p> <p>c) Community</p> <p>d) Religious</p> <p>e) Other</p> <p>f) Not stated</p>
Outcomes assessed - Service delivery <i>Refer to school- and system-level processes of organising work that has an effect on learning outcomes. Service delivery includes the 'technical core' of schooling, the primary processes that provide the conditions for learning in the classroom, as well as the wider organisational structure and environment that provide the direct and indirect conditions for classroom</i>	<p><i>The service indicators may include, but are not limited to: Infrastructure (electricity, water, sanitation); Children per classroom; Student/teacher ratio; Textbooks per student; Teacher absence rate; Time children are in school being taught; share of teachers with minimum knowledge; Education expenditure reaching primary school; Delays in wages.</i></p> <p>Please code outcomes as described in the document, noting any correspondence with items listed above.</p>

Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?

<p><i>practice.</i></p>	
<p>Outcomes assessed - System efficiency <i>Refers to whether school and system-level processes deliver school educational services effectively and efficiently.</i></p>	<p><i>This may include, but are not limited to: Cost/expenditure; Access; Equity</i> Please code as described in the document, noting correspondence with items listed above.</p>
<p>Outcomes - Learning outcomes</p>	<p><i>This may include, but are not limited to: enrolment; attendance; retention; year repetition; completion rate; attainment; labour market participation</i> Please code as described in the document, noting correspondence with items listed above.</p>
<p>Outcomes - Other</p>	<p><i>Please note any outcomes mentioned that do not fit in categories above.</i></p>
<p>Interventions and outcomes What is/are the specific intervention(s) discussed in the document? How is the intervention supposed to work (programme theories)? How does/did it work in practice? (Consider integrity of implementation, unintended effects, etc.) In what ways, if any, was evidence presented of ways in which specific programme actions related to outcomes noted above?</p>	
<p>Key mechanisms: What are the explicit and/or implicit reasons asserted or implied for the connection or disconnection of programme actions to the outcomes of interest (system delivery, system efficiency, and learning outcomes)? Please code all descriptions reported</p>	<p>a) Setting expectation b) Providing feedback/consequences c) Institutionalisation of norms d) Capacity development of educators e) Capacity development of local stakeholders</p>

in the document	f) Others (Please code as described in the document) g) Not stated
Context - pre-existing conditions Code explicit statements by the authors that link effects of the context on the mechanisms that fired / intended to fire.	a) Please specify <i>Political, economic, cultural, power relations, participation features of intervention implementation that affected whether and how the program generated outcomes</i> b) Not stated
Summarise programme hypotheses/theories Succinctly characterise hypotheses or theories, if explicit, that underlie intervention(s) addressed. For what aspect(s) of the intervention does this document provide evidence for establishing clear hypotheses or theories (ie evidence 'to support, refute or refine elements of theory')? <i>(Include a brief summary of the hypothesis, nature of the evidence and page numbers if appropriate).</i>	
What amendments to the initial rough theory outlined in the protocol might you propose on the basis of this document?	
Comments or questions on methodology	a) Relevance b) Rigour
Other notes or comments	

First produced in 2014 by:

Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square
London WC1H 0NR

Tel: +44 (0)20 7612 6397

<http://eppi.ioe.ac.uk/>

<http://www.ioe.ac.uk/ssru/>

The **Evidence for Policy and Practice Information and Co-ordinating Centre** (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

The views expressed in this work are those of the authors and do not necessarily reflect the views of the EPPI-Centre or the funder. All errors and omissions remain those of the authors.

This document is available in a range of accessible formats including large print. Please contact the Institute of Education for assistance:

telephone: +44 (0)20 7947 9556 email: info@ioe.ac.uk