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

SYSTEMATIC REVIEW [DECEMBER 2016]

1

The authors of this report are
Dr David Eddy-Spicer, University of Virginia
Dr Melanie Ehren, UCL Institute of Education, University College London
Dr Mukdarut Bangpan, UCL Institute of Education, University College London
Dr Meena Khatwa, UCL Institute of Education, University College London
Frank Perrone, University of Virginia

Contact
David Eddy-Spicer, Associate Professor
Curry School of Education, University of Virginia
dhe5f@eservices.virginia.edu

Conflicts of interests

None

Citation
Eddy-Spicer D, Ehren M, Bangpan M, Khatwa M, Perrone F (2016) *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*. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Picture

Photo by Tanya Holden/DFID using licence https://creativecommons.org/licenses/by-nc/2.0/

METHODOLOGICAL APPROACH

This report presents the findings from a systematic review using realist methodology. The review uses transparent methods to identify, critically appraise and synthesise studies to inform the development of theories about how school accountability policies operate locally to improve school systems and children's learning outcomes in low- and middle-income countries (LMICs). The theories take the form of configurations of 'contexts', 'mechanisms' and 'outcomes' (CMO), which offer explanatory accounts of the processes through which school accountability policies work in local contexts to achieve school-level outcomes. The narrative supporting each CMO configuration provides a theoretically-informed and empirically grounded explanation for the proposed configurations, presenting a transparent argument based on the synthesised literature. As a type of systematic review, the purpose of a realist review is not only to explore whether a particular intervention does or does not work but to explain why certain outcomes arise through elaboration of the connections amongst contexts, mechanisms, and outcomes grounded in the literature. 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).

REPORT STRUCTURE

As this is the technical report of a realist review, using systematic methods, some sections of the report are necessarily detailed. Without compromising the transparency that is expected of a systematic review, we have structured this report to help those who are more concerned with the findings than the methods. Therefore, the report is organised in three sections:

1. **Systematic review summary**: An eight-page executive summary of the key findings of the review.

2. **Main technical report:** This contains the background and methods to the reviews (Chapters 1-2, an overview of the studies included in the review (Chapter 3) and the reviews findings (Chapter 4-6). The review findings outlining the CMO syntheses in full are presented in standalone chapters for each school accountability policy area: Assessment, Monitoring and Inspection. The technical report concludes with a summary of the CMO configurations, in addition to the strengths, limitations and implications of the review.

3. **Appendices**: The appendices contain additional details about the reviews search strategy, coding tools used, and further details about the studies including how the CMO's configurations were generated.

SYSTEMATIC REVIEW SUMMARY

This systematic review explores how school accountability policies operate locally to improve school systems and children's learning outcomes in low- and middle-income countries (LMICs). These policies include:

- **Assessment:** student examinations used to monitor the quality of the education system, some of which (high-stakes examinations) also carry direct consequences for performance for schools, school teachers and individual students.
- **Monitoring:** the system-level processes designed to collect, compare and report school-level information about the composition, organisation and function of schools.
- **Inspection:** formal site visits to schools by education authorities to observe classroom and management activities.

Overall, findings suggest that:

- **Desirable school-level outcomes** were associated with coherent support for meeting performance expectations and for translating information about performance into the everyday practices of teaching and learning.
- **Undesirable school-level outcomes** were associated with insufficient consideration of school leaders' and teachers' capacities to engage productively with accountability activities, whether in interpreting exam results, in making use of Educational Management and Information System (EMIS) information or in conducting school self-evaluations as part of inspection.

OUTLINE OF EVIDENCE

Overall, the findings from the three elements of school accountability suggest that:

- Assessment may improve the quality of teaching and learning when the following mechanisms are triggered by specific conditions prevalent in the local school context:
- o **Trust** in the pedagogical authority of the assessment approaches is triggered by system- and school-level support for teaching tied to assessment approaches.
- o Teachers' close attention to results in ways that improve teaching follows from **customised guidance** around interpreting results.
- o Incentives prompt teachers' **desire for reward** and improvements in teaching quality when incentives are focused on individual (not collective) performance and are perceived as high-value.
- o Parental **oversight of quality of teaching and learning** promotes student performance gains when individual student incentives are perceived by parents as being of high value.
- Key barriers to assessment activities that aim to improve teaching and learning can include:
- o School staff fearing the consequences of poor performance.
- o Lack of individual teacher incentives.
- o Lack of training and support to use and interpret assessment results effectively by school staff.
- Monitoring could lead to improvement in school management and performance when one or more of the following conditions are prevalent in the local school context:
- o **Interpreting information:** Sustained effects on school management and student attendance are seen when there is consistent and clear feedback about results that is accompanied by training to interpret the results across district, sub-district and school levels.
- o **Accuracy of information:** Timely and accurate reporting of school- and district-level information occurs when those at higher levels of the system place value on understanding system performance rather than rewarding positive results ('reality testing').

- o **Local school development planning:** This is likely to be effective when school leaders and teachers are given opportunities and the ability to learn from failure.
  - o **Acting on information:** School management committees use information effectively to improve school conditions when parents develop capacity for interpreting results and pressure schools to improve teaching quality and learning.
  - o **Parental involvement:** Service delivery and learning outcomes improve when parents participate in monitoring activities.
- Inspection generally has a limited impact on systems and school-level outcomes. Key barriers to successful inspection may include limited co-ordination between the inspectorate of education and other national stakeholders, or some specific attributes of inspection feedback (e.g. disrespectful tone of voice, or recommendations out of the school's control) and the inspectors providing the feedback (e.g. lack of credibility of inspectors).

IMPLICATIONS OF THESE FINDINGS FOR PRACTICE, POLICY AND RESEARCH

- **For educators at the system and school-levels:** A key insight of this review is the way in which development of capacity may need to occur within and across levels in order for accountability activity to yield desirable school, system and student outcomes. Our initial theory, based on existing literature, suggested the opposite, implicitly assuming that development of educators' capacities would follow from school accountability activities.
- **For policy makers**: The most salient implications of our review concern ways of resolving what we identified as a tension between the monitoring functions of accountability activity and the development functions that aim to cultivate educators' capacities (see the preceding point) around improving service delivery. Typical approaches to accountability activity assume that establishing performance standards and providing feedback based on results make expectations about performance improvement explicit at the school level. However, across our review, we found that expectations need to be accompanied with proactive and consistent guidance around improving school management and teaching practice. School-level service delivery did not change in the studies we examined when those at the local level did not have the capacity or resources to fulfil implicit or explicit demands.
- **For researchers**: Research has yet to trace clear connections between change in processes at the school level that occur as a result of accountability activity and changes in student learning outcomes, particularly for the poorest and most marginalised students.

APPROACH

These findings come from literature sought systematically to answer the following review question:

> 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?

Recognising the complexity of school systems and the importance of their context, realist methodology was chosen to outline the **mechanisms** that lead to service-delivery or school-level **outcomes** and to characterise the local school **contexts** under which those mechanisms operate. This is a necessary precursor to explaining why student-learning outcomes do or do not result from accountability activities. The concentration of our review on school-level service-delivery processes and outcomes means that implications for practice at the local level are highlighted, emphasising school management and instructional practices.

Our initial theoretical framework outlined five key proposed mechanisms potentially impacting on school-level outcomes, if accountability elements were implemented as planned, without necessarily taking into account particular conditions of the local school context. These mechanisms include: 1) setting expectations; 2) providing feedback/consequences; 3) capacity development of educators; 4) capacity development of stakeholders; and 5) institutionalisation of norms (see Figure 2.1).

The connection between descriptions of conditions in the local school context and reported outcomes (when available) was identified in the reviewed studies, guided by the initial theoretical framework. We did this by systematically identifying relevant papers, and then coding and summarising them. From each paper, we extracted information to describe the key features of each study and the accountability activities reported, including details on conditions in the local school context, outcomes and suggested or inferred mechanisms. Syntheses were then conducted for each accountability element. We elaborated the synthesis findings through additional mining of existing papers. The findings about conditions in the local school context and outcomes were then used to elaborate a more refined model of potential mechanisms, within and then across accountability elements.

We included 68 studies that investigate the three accountability elements in primary and secondary education in LMICs. We included studies published on/after 2001 and in English. The evidence base for school accountability relevant to this review is largely from sub-Saharan Africa with a smaller portion of the papers coming from South Asia, Latin America, and East Asia and Pacific.

ASSESSMENT: KEY FINDINGS

We identified 34 papers focused on standardised assessment. Nearly half were from Sub-Saharan Africa (47%); with less than a quarter (21%) from Latin America and the Caribbean. The remaining papers were from East Asia (9%), South Asia (9%) and developing countries with no country-specific context (14%). Six papers were judged to be high on rigour, 18 medium and 10 low. Eleven studies employed quantitative evaluation designs; of these four were randomised controlled trials (RCTs) of interventions related to assessment activity.

Two key types of assessment activities were identified: high-stakes examinations and low-stakes assessments.

HIGH-STAKES ASSESSMENTS

These include tests and other forms of national assessments that may be designed to monitor the education system but also carry consequences connected with the use of assessment results to evaluate the performance of schools and of school teachers as well as of individual students.

The table below outlines the key findings of the review in relation to high-stakes assessment activity. The table is organised around the categories of mechanisms proposed in our initial rough theory, discussed above. These 1) include setting expectations, 2) providing feedback, 3) capacity development of educators, and 4) capacity development of stakeholders.

Mechanisms domain	Conditions, mechanisms and outcome statements
1) Setting expectations	a) High-stakes examinations are more likely to improve the quality of teaching and learning through establishing **trust in the pedagogical authority of the examination.**

Mechanisms domain	Conditions, mechanisms and outcome statements
	Evidence from studies suggests that this is particularly the case when there is: • internal (school-level) and external support for high-quality teaching and instructional leadership, and for interpreting results, • appreciation of productive persistence, or • external and internal pressure for results (school system, parents and community, or media). b) In contrast, lack of support coupled with pressure for performance appears to trigger **fear in school leadership and teachers**, resulting in unintended consequences in terms of using short-term teaching and learning strategies, emphasising the teaching of technical compliance not content mastery, and limiting opportunities for low-performing students.
2a. Providing feedback through information	High-stakes examinations are more likely to produce teaching that promotes higher-order thinking and content mastery when the **customised guidance to schools and teachers is accompanied with coherent training for preparing examinations and interpreting results at the school and local levels** (Following-up/following through).
2b. Providing feedback through incentive	a) High-stakes examinations are more likely to increase efforts by individual teachers on exam preparation and working with lower performing students, and produce sustained increases in test results **through the desire for reward.** The evidence suggests that this is more likely to be the case when there are: • teacher-level individual incentives, • pressures from school leadership and external stakeholders for results, or • teachers' recognition that the incentive is of value and merits additional effort. b) High-stakes examinations are more likely to increase student participation and teacher attendance, and produce sustained gains in test scores through **parental ability to exert pressures to improve children's performance.** Further conditions in the evidence to support this content-mechanism-outcome (C-M-O) include the presence of: • individual awards or incentives for students, • parental recognition of value and pressure for results, • pressures from school leadership for results, or • teacher acceptance of parental input. b) High-stakes examinations could limit efforts by school staff to meet minimum requirements to earn school-level incentives for improving students' performance over

Mechanisms domain	Conditions, mechanisms and outcome statements
	time, triggered by **compliance with bureaucratic authority among** teachers and school leaders, when there are: • incentives at group or school level, or • pressures from school leadership and provincial or state education authorities for results.
3 & 4 Capacity development of educators through school-based assessment (SBA)	a) School-based high-stakes assessments could lead to negative teachers' perceptions such as perceptions of interactive pedagogies and assessment for learning as inappropriate or irrelevant, or undesirable instructional practices, resulting from **a lack of follow-up/follow-through** for interpreting examination results. The evidence suggests this may be the case when: • there is lack of provincial support, • SBA is facilitated by inexperienced educators, • The context and framing of tasks are predetermined by national design and not appropriate to the level of student understanding, or • local school administration and/or provincial and state authorities pressure teachers for results. b) School-based high-stakes assessments could lead to unintended teaching outcomes such as lack of competence in translating lesson objectives into assessment tools, teachers 'parroting' assessment tasks from previous exams, or negative teachers' perceptions on SBA as a technical procedure unrelated to professional judgement, **through a mechanism of compliance with bureaucratic authority**, when, according to the evidence: • there is inadequate teacher preparation, • professional development for teachers focuses mainly on procedural or bureaucratic aspects, • SBA emphasises summative, not formative aspects of assessment, or • local school administration and/or provincial and state authorities pressure teachers for results.

LOW-STAKES ASSESSMENTS

These are national assessments that have been introduced as a means of monitoring educational quality and that provide information not carrying direct consequences for schools, teachers and/or students.

Mechanisms domains	Conditions, mechanisms and outcome statements
1) Setting expectations	Low stakes assessments could influence teacher training, curricular change and pedagogical innovation, triggered via **trust in pedagogical authority of assessment approaches**. The evidence suggests this may be when: • there is consensus around form and process of assessment, and • reporting of results is viewed by teachers as useful means of informing daily practice.
3) Capacity development of educators	Low-stakes assessment may have an impact on school-level practices through the provision of information and guidance to teachers and school managers when detailed analyses of student responses that are directly relevant to teaching practice are available.

MONITORING: KEY FINDINGS

We identified 22 papers focusing on monitoring. Nearly half of these were from sub-Saharan Africa (n=10). An additional five papers covered multiple regions, and seven were country case studies in Latin America (n=2), East Asia (n=4) and South Asia (n=1). Three papers were judged to be of high rigour, 11 medium and 8 low. Three monitoring programme activities were identified: 1) Educational Management and Information Systems (EMIS) for school-level management decisions; 2) school report cards; and 3 EMIS for local school development planning (SDP).

EMIS FOR SCHOOL-LEVEL MANAGEMENT DECISIONS

Mechanism domains	Conditions, mechanisms and outcome statements
1) Providing feedback through EMIS	Information from EMIS and guidance and training provided to staff and school *(follow up/follow through)* at the district, sub-district and school levels could lead to improvement in school and instructional management processes (e.g. for monitoring absenteeism, tracking progress in student learning, the school supervision and monitoring system) when there are: • clarity of key organisational processes within and across classroom, school, sub-district and district levels (e.g., reporting requirements, approaches to information gathering), • support in implementing such key organisational processes with and across levels, and/or • consistent and clear feedback about implementation of workflows, or ways to improve.
2) Providing feedback through school report cards	a) School report cards are likely to improve school performance by reallocating resources, reducing school fees, or improving children's test performance, through **parental ability to exert pressure**, when there are: • decisions from the local stakeholder group (e.g., school management committee) of consequences to school personnel, • local stakeholders' capacity to use information effectively to understand school performance, and/or • school personnel capacity to work with local stakeholders. b) School report cards using participatory approaches could create a sense of ownership among school staff and parents and other community members and may lead to reductions in pupil and teacher absenteeism in schools, improve children's test scores, **through parental participation in monitoring activities,** when: • there is engagement of parents and local community members in making decisions around what information to collect and the process of collecting information, • parents and local community members have the capacity to use information effectively to understand school performance, and/or • school personnel have the capacity to work with local stakeholders.
3) Setting expectation through EMIS with SDP	a) Uses of EMIS for school development planning (SDP) could create an ownership of local education issues and may lead to the improvement of primary school enrolment among minority girls, triggered by **learning from failure**, when there is: • supplemental funding for implementation of a new national curriculum, • empowerment at school level of planning and resource allocation,

Mechanism domains	Conditions, mechanisms and outcome statements
	• school leadership training in education management and on new curricular materials, and/or • An environment for experimentation. b) Uses of EMIS for SPD could improve the quality and reliability of information for school planning and equitable allocation of resources at school and classroom levels, through a mechanism of **gathering information**, using that information and evaluating that use in order to refine how the information is gathered and used in subsequent rounds, when there are: • minimum school standards focusing on basic inputs for learning, and/or • an EMIS infrastructure for accurate and timely monitoring and reporting of school and district-level progress against standards. c) Uses of EMIS for SPD could lead to decreased ownership of local education issues by the school and community, triggered **by lack of follow-up/follow-through**, if: • there is a lack of capacity at the district level to support participatory processes at the school level, and/or • 'participatory processes' are aimed at meeting district demand for information, not local participation driven by district rather than community needs.

INSPECTION: KEY FINDINGS

We identified 22 inspection papers in this review. Of these, three were judged to be high on rigour, eight medium and 11 low. Only two studies suggest that inspection has an impact. Macpherson (2011) describes how school inspections in Timor Leste have the potential to contain the scale of corruption in the misuse of school grants; however, the study only looked at how schools are investigated and how inspectors investigate allegations of misuse, so no claims can be made about school inspections actually leading to a decrease in corruption. Brock (2009) draws on a number of medium-rigour case studies in Gansu province in China when explaining how increased power to school inspectors to report on the quality of schools and to propose changes in and support of schools led to an improvement of school development planning.

Unintended consequences occur where teachers in Ghana put on an act during inspection classroom observations (Opoku-Asare, 2006), and principals use inspections to threaten their teachers in South Africa (Mazibuko, 2007). According to Opoku-Asare (2006), school inspections are often pre-announced, thus enabling the teachers concerned to prepare adequately for the observation lessons. This enables those teachers to arm themselves with all the teaching materials they can possibly lay hands on, and sometimes, rehearse the lessons they intend to teach for the exercise (p. 112).

Our review focused on the underlying mechanisms of change and the unintended consequences of school inspections, and mechanisms that explain a lack of impact. We found evidence of school-level outcomes and traced corresponding conditions related to inspection activity under four of the six proposed categories of mechanisms: setting expectations, providing feedback, capacity development of educators and capacity development of stakeholders.

Mechanisms Domain	Conditions, mechanisms and outcome statements
1) Setting expectations	The development of standards, guidelines and frameworks to inspect schools can be an important driver for improvement as they: • inform schools of where to focus on in their improvement plans, • support school self-evaluation, and/or • ensure consistency of inspection assessments and feedback to schools.
2) Providing feedback/ consequences	Lack of impact from inspection may result from specific attributes of inspection feedback and the inspectors providing the feedback, when there is: • lack of credibility of inspectors (e.g. due to low pay scale), • disrespectful tone of voice, and/or • recommendations on administrative procedures and conditions out of the school's control
3) Capacity development of educators	An increase in school internal evaluation when used as a component of school inspections may strengthen the participation and commitment of teachers in school change and sustainable improvement, particularly when: • schools are supported in the development of their internal evaluations, and/or • school have access to guidelines and handbooks that would support their evaluation.
4) Capacity development of stakeholders	a) Alignment and co-ordination between inspectorates of education and other education service providers or stakeholders in the education system are expected to enhance the impact of school inspections, as this ensures that: • school improvement efforts across the system focus on the same standards (preventing confusion in schools), • there are consequences and follow-up on inspections, • relevant actors follow up on inspection recommendations, and/or • relevant actors (such as teacher unions) buy-in to inspection standards and recommendations. These relevant actors will then support and pressure school staff to act on inspection findings. b) In contrast, limited co-ordination between the inspectorate of education and other national stakeholders, such as teacher training or resource centres in the dissemination and use of inspection findings, potentially limits the impact of school inspections, as it leads to a lack of follow-up on school inspection visits and findings, and limited support to schools on the implementation of inspection feedback.

RESEARCH GAPS

There is a need for more robust research around what it takes to improve teaching and learning in schools. Our review highlights a paucity of high-quality studies in most areas, particularly in low-stakes assessments, and this may be one of the areas that offers promise for revealing the dynamics of change in schools and in classrooms. Moreover, the connectivity of mechanisms of change suggests that one-dimensional research approaches of looking at cause and effect of inspections, assessment and monitoring translate with difficulty when trying to explain the impact of accountability in low- and middle-income countries. Traditional methods of analysing large, longitudinal datasets to link schools' status on accountability measures to improve student achievement results (see for example Allen and Burgess, 2012; Hussain, 2012) are one kind of evidence for research in developing countries but may not provide a complete picture. New research methodologies and approaches need to be developed which specifically look at interlocking mechanisms and conditions of change, examining cyclical cause and effect relations to explain and understand impact of accountability in different settings. New approaches to conceptions of rigorous research as deeply embedded in continuous improvement of practice within and across levels of the education system may have particular salience in this regard (e.g., Bryk, et al., 2015).

The limitations of the literature that we identified made it impossible to conduct comparative analyses of accountability approaches across geographic regions or even within regions. Such work could make an important contribution to understanding how systematic variation in historical, social, organisational and cultural contexts shapes responses to system-wide accountability initiatives at the local level if it were structured to examine those contexts in depth. In a related way, intensive studies of the interdependence of accountability initiatives within a single education system would provide valuable insight into how accountability elements operate in concert to produce various outcomes.

CONTENTS

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 are intended to be of use to policy makers, politicians, civil servants and education leaders in low- and middle-income countries (LMICs). This chapter introduces the basic principles, aims and rationale for the review, the policy and research backgrounds and the review question.

1.1 AIMS AND RATIONALE FOR CURRENT REVIEW

This review explores the conditions under which three approaches to school accountability – inspection, monitoring and assessment – lead to improvements in schools and school systems as well as to positive learning outcomes for schoolchildren, especially the poorest and most marginalised in LMICs. The review emphasises the impact of accountability interventions on the quality of education delivered by schools, as this has increasingly been the focus of accountability initiatives in LMICs over the past two decades (Bruns, et al., 2011; Carr-Hill et al., 2015).

The objective of understanding 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, then elaborate on our rationale and the specific steps necessary in subsequent sections.

We have turned to realist synthesis (Pawson, 2006; Pawson et al., 2005; Wong et al., 2013) because of the complexity and dynamism of conditions that influence the outcome of accountability systems in LMICs, the wide variability in the 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 a high-stakes examination that serves as a gatekeeping function for further education, severe consequences for students and teachers for low performance in the exam, under-resourced schools, and inadequately prepared teachers and school managers. For accountability interventions that include standardised assessment, this review aims to identify mechanisms that result in teaching to the broader curriculum and teaching to gaps in students' understanding rather than just to what is assessed, yielding high-quality service delivery. The conditions that cause teachers and leaders to behave differently might include, for example, the existence of professional networks accompanied by a coherent sense of professionalism within and across schools.

In this way, the review aims to help education advisers, policy makers and education leaders to understand the causal processes that result in certain outcomes and to identify the conditions that are necessary for those

---

[1] These are the countries with which DfID holds bilateral agreements. For a list, 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).

processes to have the desired outcomes. The review intends to sharpen policy makers' and educators' abilities 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. However, this review intends to highlight important mechanisms and associated conditions in ways that might lead to insights into areas for programme development as well as areas for further research.

1.2 DEFINITIONAL AND CONCEPTUAL ISSUES

In this section, we parse each of the key concepts in the research question as the starting point for clarifying the topic of the systematic review. Various interpretations exist for each of the important words in the review 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.

GENERAL ACCOUNTABILITY

'General accountability' comprises the types of accountability approaches described in the additional information (DfID, 2013c) that accompanied DfID's call for proposals for this review:

Regulatory school accountability: Ensuring compliance with laws and regulations. This focuses on inputs and processes within the school, e.g. school inspections.

Performance/results-based accountability to improve schools: Periodic school evaluations. The mechanisms include: a) standardised student testing; b) public reporting of school performance; and c) rewards or sanctions. In other words, the use of assessment systems or monitoring systems.

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	**Regulatory school accountability:** Compliance with laws and regulations; focuses on inputs and processes within the school.
	School performance accountability: Periodic school evaluations.
Horizontal	**Professional school accountability:** Professional standards for teachers and other education staff.
	Multiple school accountability: 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 to the three accountability elements of inspection, assessment and monitoring. While the elements of horizontal accountability are not directly addressed, they are important in understanding the essential conditions by which external accountability elements could bring about changed decision making and behaviours that give rise to desired outcomes.

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.

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. Her Majesty's Inspectorate – 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, the 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 recognise that a developmental brief held by the same role holder may give rise to different mechanisms and yield distinctly different outcomes.

In this review, we were, however, particularly interested in the school-level evaluative dimensions of the role, which means that we considered inspection/supervision that has at its core an element of 'judgement', using a framework that allows for some level of comparison between schools, where the person responsible for making the judgement is external to the school (not present in the school on a day-to-day basis) or responsible for more than one school. The judgement would typically also have consequences for schools/school staff, which may be punitive or in the form of additional support for schools/head teachers. We recognised however that those consequences are often not put in place in developing countries, due to limited resources (e.g. no funding for additional monitoring), but the authority undertaking inspections/supervision needed to have a formal role/position that was authorised to enact such consequences. A judgement can include an aggregate score for the school (e.g. as failing or performing well), but may also include an overview of strengths and weaknesses. The judgement is communicated to the school, and typically also (but not necessarily) published in an inspection report, and made available to the school and the wider community.

ASSESSMENT

The review focused on standardised tests of learning that are developed and deployed at various levels – provincial, national or regional – and are scored and reported in a consistent manner that permits comparison of performance at the school level. We emphasised forms of assessment that are used to compare and rank the performance of individual students within schools, schools and groups of schools. 'Standardised' points to consistency in 'test design, content, administration and scoring to ensure comparability of the results across students and schools' (Best et al., 2013, p. 2). Test content and scoring 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. Also, standardised test results are used at the system level to make judgements about the performance of schools and/or groups of schools, and to potentially implement consequences in relation to those judgements (e.g. school closure, intensive monitoring). We were not concerned with regional or international assessments and surveys used to gauge the performance of school systems (e.g., PISA, TIMSS). We were interested in the processes of assessment as these influence school-level outcomes.

MONITORING

Monitoring encompasses the infrastructure and methods used to track school-level information collected through quantitative/empirical methods, primarily numerical information that is then used to evaluate school performance against benchmarks and/or targets in order to evaluate quality. Monitoring refers specifically to the system-level processes designed to collect, compare and report school-level information about the composition, organisation and functioning of schools. Monitoring includes formal systems of EMIS, and the collation of 'input' or administrative data, as well as data that tracks performance information (e.g., school report cards). Input and administrative data typically include data on student-teacher ratio, dropout rates, graduation rates, number of school staff, etc.

OPERATING DEFINITIONS FOR OUTCOMES

SERVICE DELIVERY

'Service delivery' was 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, and 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 to the 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 served as initial proxies for service delivery inputs in our efforts to understand causal processes associated with inspection, monitoring or assessment. These indicators were also considered as important conditions, resources or causal processes that might contribute to service delivery outcomes as well as outcomes indicative of system efficiency and student learning.

SYSTEM EFFICIENCY

Many countries are facing the challenge of orchestrating education polices 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 education 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 considered closely the processes and conditions that enabled the system to ensure that education expenditures reach the school (a service delivery indicator) and that expenditures were 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 recognised that the desired outcome is for the education 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 were 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). In this review, we aimed to gain a 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. We emphasise performance on standardised assessment as a proxy for learning outcomes. We realise that clarifying what one means by student learning outcomes depends on the purpose coupled with identification of appropriate proxies. We considered the 'quantity' end of that spectrum, measuring learning outcomes in terms of children's enrolments, attendance, retention, year repetition, survival and completion rates. We also considered longer-term outcomes, such as labour market participation. However, current approaches to school accountability overwhelmingly focus on the acquisition of cognitive skills as expressed through student performance on standardised assessments (Vegas and Petrow, 2008, pp. 8-9).

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 and low-stakes) assessment over nearly three decades

and includes large-scale surveys, small case studies and quantitative analyses of test scores (Stecher, 2002). Of relevance to the review was 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 for that study, 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 countries (HICs) have provided descriptive taxonomies of less- and more-effective practices (Haladyna et al., 1991; Popham, 1991; Mehrens and Kaminski, 1989; Stecher, 2002). Our emphasis on processes in LMICs that cause outcomes and the conditions that give rise to these processes challenged the ready translation of research from HICs. However, taxonomies may prove useful in conceptualising relationships among conditions, causal processes and outcomes. In any case, the results from these studies bolstered the need for close attention to context, as studies have consistently found that most practices were neither clearly effective nor ineffective because the consequences for student learning were contingent on the context in which, and the extent to which, practices occurred.

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 HICs, 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 (De Grauwe, 2008; Powell, 2006; Scepanovic et al., 2010).

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, 2001, 2007; De Grauwe and Lugaz, 2007; Dembélé and Oviawe, 2007) 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 accountability elements. 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. We paid close attention to the ways the connections between conditions, the processes that arise from those conditions and the outcomes that were caused by those processes. Realist synthesis 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 FUNDERS AND OTHER USERS OF THE REVIEW

We worked 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 education consultants would also find the review useful in support of their evaluation of accountability policy and implementation.

This review is useful to other agencies in the design/reform, implementation and evaluation of accountability systems. Such agencies include bilateral and multilateral agencies and organisations working in LMICs. Other interested parties are 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 serves as a model for others embarking on systematic reviews in this area.

1.5 REVIEW QUESTION

The question that we aimed 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?

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

2. METHODS USED IN THE REVIEW

2.1 REALIST SYNTHESIS

Realist synthesis aims to build explanatory models, or one or more middle-level theories, that trace paths across conditions, mechanisms and outcomes related to a programme or intervention, such as 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 middle-level 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 in our heads, 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 systematic building or testing a range of middle-level theories that marks the realist review out from other review approaches. Similar to framework analysis, realist synthesis depends on the elaboration of an 'initial rough theory' and the elaboration of and challenge to that theory through systematic review to reach a theory that aligns better with the existing evidence. Realist synthesis embraces theory building and testing at a greater level of specificity than does framework analysis, 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. Realist synthesis also offers more developed tools to plumb the complexity of social interventions that consist of a number of linkages or intervention chains. Well-known approaches to systematic reviews, or 'what works' review, rarely provide the knowledge that explains why programmes do and do not work. Several noteworthy examples of realist review helped illustrate this concretely for us, including Greenhalgh, et al. (2007) in health and Westhorp, et al. (2014) in education. The middle-level theories elaborated in these reviews 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. In our review, we intended to make these implicit programme theories explicit by articulating what we called 'pathways to impact', which we defined as the ways that programmes were designed to produce intended outcomes. An important task of a realist synthesis is to probe the primary literature to develop clear understandings about the reported or suggested ways that programmes 'work' to generate the outcomes of interest (Wong et al., 2013, p. 2). Clarifying intended pathways to impact allowed us to understand how reported or suggested mechanisms of impact described in the literature differed or aligned with what the programme intended to accomplish. For example, Ehren et al. (2013) found that practitioners' actions on inspection feedback were rare, although the intended pathway to impact for inspection is to provide performance information that will lead school managers and teachers to change their behaviours. Much more common were actions based on the

anticipation of inspection; in this understanding, the mechanism was educators' expectations that caused behaviour to change, not the results from the inspection itself.

In an analogous example, Educational Management Information Systems (EMIS) are also assumed to operate through feedback as a result of reporting of results to various stakeholders. For accountability interventions that include the implementation of EMIS, this review sought to delineate the intended pathways to impact, which included local school development planning and school monitoring report cards, and then we sought to identify the conditions and the mechanisms triggered by those conditions that related to the impact that was reported in or suggested by the literature. One of the key conditions suggested in the papers we review, for example, is the interdependent relationship between the provision of high-quality data input from schools and performance information that is meaningful to schools. In other words, papers claimed that school managers and local officials provided higher-quality data to EMIS when they understood that the performance information that came out of EMIS could help them with their day-to-day decision making and planning.

These examples highlight that 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). Realist syntheses examine how social programmes work 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. The espoused theory of change of a programme 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.

Realist synthesis, while relatively new to systematic reviews in education, has been used in a wide range of social science research.[2] This review followed the publication standards for realist reviews put forward by the RAMESES (Realist And Meta-narrative Evidence Syntheses: Evolving Standards) project (Wong et al., 2013) (see the appendix 2.1). To illustrate the benefits of this approach, we turned to a recently completed systematic review funded by DfID and exploring an important issue of contemporary education policy, community accountability, through realist synthesis. The review by Westhorp et al. (2014) employed 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 identified 11 mechanisms and 13 categories of features of context. Mechanisms characterised the key processes through which community accountability interventions worked. In this way, the review developed and refined middle-level 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-level theories that Westhorp et al. (2012) hypothesised involves the conditions that led stakeholders to take actions that enhanced local responsibility for schooling, given rewards and sanctions. The synthesis of primary evidence conducted by the review team led them to a mechanism labelled 'carrots and sticks' (Westhorp et al., 2014, p. 45). A related mechanism explained not the consequence of rewards or

---

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

sanctions but the anticipatory effect that awareness of inevitable sanctions or rewards might have on actors, who sculpted their actions accordingly, a mechanism that the review labelled 'big brother is watching me' (ibid, p. 45). The mechanism of 'big brother is watching me' was similar to the effect of establishing expectations from inspection found by Ehren et al. (2013).

Accompanying the mechanisms was an analysis of features of context that were essential to the operation of each mechanism. One of the review's findings around context was relevant to 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 identified the following passage from one study as characteristic of the programme theory (Weiss, 1998) that underlay 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., 2012, p. 8, quoted in Westhorp et al., 2014, p. 64)

The reviewers noted 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 (Westhorp et al., 2014).

The report then concluded with nine recommendations for policy and practice. The reviewers derived these from their elaboration of middle-range theory, identifying the conditions under which certain mechanisms caused desired outcomes. The constellation of middle-level theories was then used to return to an 'initial rough theory' developed at the start of the review and strengthened it so that it could more robustly identify the proper conditions and related mechanisms that led to desired outcomes for community accountability and empowerment initiatives. The review also clarified the kind of research that would appear to be most needed to build better and more durable understanding of such programmes.

Our review is also a *theory-building* review, in that we examined primary literature in a field that was under-theorised. The pathways to impact – connections between accountability implementation and intended outcomes – were 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 outline the different ways in which the accountability elements 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).

2.2 USER INVOLVEMENT

We worked with a small group of academic advisers and experts who work in the field of accountability in developing countries in the design of the initial rough theory and in its iterative testing through the identification and verification of key mechanisms. These advisers, who were identified through the extensive networks of the authors, were Thomas Hatch and Luis Huerta of Teachers College, Columbia University; Dennis Shirley, Lynch School of Education, Boston College; Pantalee Kapichi, UNICEF Tanzania; and Anton De Grauwe,

IIEP, UNESCO. These contacts were identified through the extensive networks of the authors (See Appendix 1.1)

These advisers constituted our Advisory Group and provided feedback throughout the project, particularly in finding additional relevant sources, helping us to clarify key terms and to refine the scope of the review, providing feedback on a draft report and supporting the dissemination of the initial and final findings.

The Advisory Group was contacted at the following times and for the purposes outlined:

June, 2014:

- Feedback on the protocol.
- Feedback on 'initial rough theory' (Section 2.3).
- Feedback on inclusion/exclusion criteria.
- Suggestions of additional sources/contacts.

October, 2014:

- Feedback on the searching, scoping and progress of the review (Section 2.4).
- Regional focus: Limit the focus to four regions (Latin America, Sub-Saharan Africa, East Asia, South Asia).
- Suggestions of additional sources of region/country unpublished reports/information.
- Cut-off date: 2001 agreed.
- Agreement to include all three accountability elements.

June, 2015:

- Feedback on the initial draft.

At the suggestion of the Advisory Group, researchers contacted project officers in the World Bank and staff working in national ministries of education in Mexico, Brazil and South Africa for additional literature.

2.3 IDENTIFYING AND DESCRIBING STUDIES: INITIAL ROUGH THEORY AND SYSTEMATIC MAP

Identifying and describing studies in this review was done in overlapping phases: 1) scoping of the literature and the development of initial rough theory; 2) search process; 3) screening and selection of relevant papers; and 4) characterising the included studies in a systematic map.

These phases shared several common stages, including defining relevant studies through inclusion/exclusion criteria, elaborating a search strategy to identify potential studies, retrieving relevant papers and characterising the included papers.

INITIAL ROUGH THEORY DEVELOPMENT

The first phase involved the development of the initial rough theory, which was used throughout the review to inform the mapping evidence and refining of mechanisms. The scoping of literature for use in developing the initial rough theory was undertaken by the principal and co-principal investigators 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). They read the full text of 25 articles to develop the initial rough theory of how three accountability elements lead to improved outcomes in LMICs.

We started our initial rough theory development by teasing 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, 2008). 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 then developed a generic hypothesis about how systemic accountability intended 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' indicating the organising processes of the school.
- Outputs that relate to student learning.
- Outputs that relate to the technical efficiency of the school.

We discerned two levels of outcomes – those at the organisational level and those at the level of the education system. At the organisational level, we considered increased student access to education, reflected in increases in enrolment as well as more regular student attendance; time devoted to teaching in classrooms and greater allocation of education expenditure for teaching and learning as an outcome were also included. Finally, these outcomes could be translated across schools in ways that led 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 drew on and extended Bouckaert and Halligan (2008), Ehren et al. (2013) and Hatch (2013) to highlight five hypothetical generic mechanisms to explain how accountability systems could lead to organisational and system-level outcomes. We call these 'programme pathways', as they are the mechanisms through which various accountability activities are intended to produce desired outcomes:

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

The first programme pathway is **setting expectations**. This acknowledges the fact that indicators used in accountability frameworks, such as inspection standards, testing frameworks and taxonomies, and performance indicators in monitoring systems, have a normative or standardisation purpose. Such indicators not only serve a measurement function to undertake inspections, assessment or monitoring of school quality, but they also communicate expectations about goals and about what a good school, a good lesson and good performance constitute. Schools are expected to use the criteria and descriptors set out in the accountability frameworks (e.g. inspection standards, school report cards) and in the testing frameworks for standardised

high-stakes assessments to define their own standards of a 'good school' and a 'good lesson' and to incorporate these standards into their daily work and teaching. The communication and use of standards in school accountability are expected to motivate schools to reflect on the standards, process them and adapt their goals and their practical ways of working in such a way that they come closer to the normative image of schools communicated by the accountability indicators. This behaviour in response to expectations corresponds with *answerability*, the notion that schools should be accountable for meeting agreed-upon procedures and goals (Gregory, 2003).

The second programme pathway is through the **feedback** from assessments, inspections and monitoring that is provided to schools. Such feedback may include an outline of strengths and weaknesses on school quality in inspection reports, benchmark information around a number of performance indicators in school report cards, or a comparison of the performance of specific student groups on standardised tests. Accountability systems often set targets for school performance and have consequences (e.g. sanctions and/or rewards) in place for low- and high-performing schools. Such consequences are expected to motivate schools to attend to the feedback provided. Schools are assumed to use the feedback to improve, and stakeholders are expected to take note of the feedback and hold schools accountable for their use of the feedback for improvement.

Capacity-building of educators is our third programme pathway and refers to the school's capacity to enhance the professional learning of teachers and to transform large-scale reform into accountable student-oriented teaching practices. Improvement capacity is considered to be an important condition for school development in general, as well as in response to external accountability. School accountability is expected to build a school's capacity for improvement primarily through: impact on school self-evaluation and the school's internal quality assurance systems; impact on professional development, school collaboration and external support around (improvement on) accountability indicators; and introducing new leadership roles. High-quality self-evaluation is considered to be a critical element in improvement of schools, as schools identify and correct problems in the quality of their school in preparation for, and in response to inspections, assessment and monitoring. Internal quality assurance mechanisms, together with external accountability, are seen as inseparable and integral parts of an informed and evidence-based improvement cycle that build capacity in schools to improve the teaching and learning and lead to improved student outcomes.

Capacity development of local stakeholders, as a fourth programme pathway, is about engaging a 'third' party in school accountability, providing them with the information and support to have an active role in school evaluation and improvement. Local stakeholders typically include parents and community members, as well as students and local officials. Examples of capacity development might include the public dissemination of results such as inspection reports and school monitoring report cards, as well as forms of participatory evaluations in which a school's stakeholders take an active role in the evaluation of schools, such as when stakeholders are involved in the development of inspection standards, school inspectors interview parents or school boards during school inspections, or they require the school to actively engage with community members in the process of constructing and analysing school monitoring report cards.

The inclusion of stakeholders as a 'third' party in school accountability is expected to reinforce public recognition of accountability standards and make it more likely that schools react to these standards in anticipation of the response of local stakeholders. Stakeholders may, however, become more active and raise their 'voice' in order to motivate schools to improve. If schools do not give stakeholders sufficient opportunities for participation (in that they accept some 'stakeholders' influence' or enter into 'negotiation' with them), stakeholders may retreat to the option of 'choice' or 'exit' where parents choose to enter or move their child to a higher-performing school. 'Choice' and 'exit' are expected to exert pressure on schools to conform to accountability standards through the introduction of competition between different providers,

while 'voice' alternatives allow parents to express preferences and opinions around education service delivery that would motivate schools to improve.

The final programme pathway is the **institutionalisation of norms.** When the behaviours of teachers and school leaders, as well as local stakeholders, move beyond compliance with expectations set externally, then this is an indication that the values, attitudes and beliefs implicit in systems of accountability are internalised in educators' and stakeholders' ongoing practices. This corresponds with Bouckaert and Halligan's (2008) notion of *internal consolidation* in response to performance management (p. 125). For example, when school leaders and teachers view school development planning as an integral aspect of ongoing school management practices rather than a bureaucratic procedure in response to external demand, the norms around integrated management and data use have become institutionalised in the sense that they are woven into the organisational fabric. Similarly, parents and local community members may openly question school leaders about their school's performance in comparison with neighbouring schools, which is again an indication that norms of local responsibility for schools have taken root.

Each of these programme pathways 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 was on the organisational implications of systemic elements. A realist synthesis intends to identify actual mechanisms of programme action which may or may not resemble the programme pathways we have hypothesised here. Actual mechanisms are inferred from identification of the particular conditions under which programme activities yield specific outcomes. Our interest in this review was in examining those mechanisms that produce school-level outcomes, as described in Table 2.1, with reference to our hypothesised mechanisms or programme pathways.

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

Conditions	Mechanism	Outcome
Effective monitoring systems Belief that the authority holder will act on data received through the monitoring system Incentives of sufficient power Performance can be observed	Setting expectations	Improvements in the extent to which, or standards by which, responsible parties implement the actions required of them.
Authority holder acts on performance information received through monitoring system Effective uses of performance information for performance improvement Incentives of sufficient power	Providing feedback/ consequences	Improvements in the extent to which, or standards by which, responsible parties implement the actions required of them.

Conditions	Mechanism	Outcome
Performance can be observed		
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
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 the roles expected of them Quality of stakeholder oversight of schooling Resources available for education are improved
Educators recognise the value and see the benefit of existing expectations Concrete performance expectations are integrated into processes of school organising There is sustained support for development of skills and knowledge	Institutionalisation of norms	Organisational and individual internalisation of system expectations Internal accountability with a focus on meeting service delivery and learning outcomes expectations, not consequences

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)
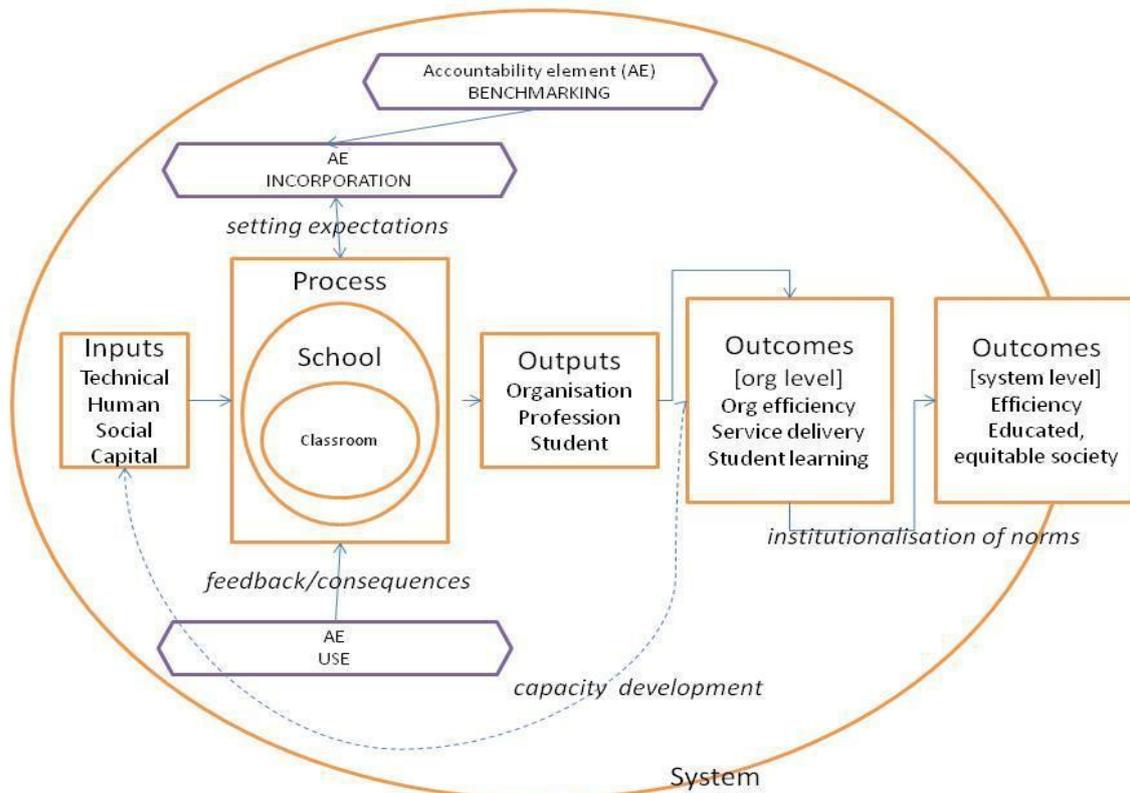
Our initial rough theory included a set of generic hypotheses, or programme pathways, about how systemic accountability is intended to influence service delivery, systemic efficiency, and learning outcomes based on the integrated open systems model of school effectiveness put forward by Scheerens (1992). We sought feedback from academic advisers about the initial rough theory included in the protocol, and subsequently refined the first draft of the initial rough theory based on this feedback (See Figure 2.1).

We analysed the three elements of accountability – inspection, assessment, monitoring – separately and then identified key programme mechanisms that caused the intended outcomes and the conditions that triggered those mechanisms within and across all three elements. This comparative analysis permitted us to hypothesise about some middle-level theories that operate for any accountability element – exemplified by the hypothetical configurations proposed above – as well as C-M-O configurations unique to each element.

IDENTIFICATION AND CHARACTERISING OF THE LITERATURE: INCLUSION AND EXCLUSION CRITERIA

The second phase overlapped with the first phase- initial rough theory development. In this phase, we identified relevant papers that addressed each accountability element. We included papers that met all of the following criteria:

- **Types of intervention**: Investigate or explore accountability (monitoring, assessment and/or, inspection) of education system
- **Geographical location**: Conducted in low- and lower-middle-income countries according to World Bank classification[3]
- **Setting**: Target primary, secondary and/or compulsory education
- **Types of studies**: All types of study designs, policy and theoretical/conceptual framework documents
- **Language**: Published in English
- **Date**: Published in and after 1990 as most accountability systems were developed after this date

The exclusion criteria are listed Appendix 2.1.

IDENTIFICATION AND CHARACTERISING OF THE LITERATURE: SEARCH STRATEGY

The search strategy that was developed aimed to identify relevant papers of direct relevance to the review question.

SEARCH TERMS

Key search terms were determined by the review question and the inclusion criteria, and were developed iteratively. At the beginning of the project, we developed the key search terms from papers already identified through hand searching of websites and reference checking of literature identified in the initial rough theory development phase.

Terms such as 'accountability' needed more clarity, as it can be defined in a broad, inter-disciplinary way, referring to different definitions and meanings covering social, community and financial concepts. In addition, different terms may be used to refer to accountability, monitoring, inspection and assessment in different contexts, for example developed versus developing countries.

We drafted initial search terms based on the literature identified through relevant reviews and websites at the theory elaboration stage. These initial terms corresponded to the three key aspects of the review: accountability elements (assessment, monitoring and inspection, as well as 'generic' terms for accountability); education level (primary and secondary education; and country (LMICs). The terms were refined through several rounds of meetings and discussions within the team.

We carried out a pilot search using the initial terms on the well-known bibliographic database for education, the Education Resources Information Centre (ERIC). We also identified more relevant search terms and key concepts through 'terms used in the database indexing and from relevant papers identified through the search during the pilot searching; we then added these additional terms to the list. The process was iterative and a final list of key terms (Table 2.3) was adapted and used in search strings for each database.

---

[3] http://data.worldbank.org/about/country-classifications (accessed 15 February 2014)

Table 2.2: Key search terms used in the review

Key aspects of the reviews	Search terms and synonyms
Accountability, inspection, monitoring, assessment	accountability, educational accountability, educational quality, benchmarking, government role, quality assurance, quality control, school accounting, school-based management, standards-based accountability, quality management Assessment 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 Monitoring 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 Inspection 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

Key aspects of the reviews	Search terms and synonyms
	relationship, supervisors, teacher supervision, evaluation, institutional evaluation, state supervisors, inspectorate, school evaluation
Developing countries	Developing nations, low-/lower-income countries, less-developed countries, third-world countries, less-developed economies, and names of countries classified by the World Bank as low- or middle-income
Primary and secondary education	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

We used combinations of the terms and their synonyms which denoted key aspects of the review. The search used the Boolean operator 'OR' to link each key aspect to their synonyms. Then, all key aspects were combined using 'AND' to identify relevant literature. For example, (accountability OR inspection OR monitoring OR assessment) AND (primary education OR secondary education).

The specific search strategy for each database can be found in Appendix 2.2.

SOURCES

We searched for both published and unpublished primary sources across an array of repositories, including nine bibliographic databases specialising in education as well as social and economic matters; references in existing systematic reviews and papers relevant to the review question, such as Klerks (2013); and websites. We also checked references and citations to find papers relevant to the initial rough theory, along with careful screening of relevant web sites. We contacted key authors and advisory group members asking for additional sources of information and relevant literature. A detailed list of the sources searched can be found in Appendix 2.3.

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

IDENTIFYING AND CHARACTERISING THE LITERATURE: APPLYING INCLUSION AND EXCLUSION CRITERIA

Inclusion and exclusion criteria were applied successively to (i) titles and abstracts and (ii) full reports. Full reports were obtained for those studies that appeared to meet the criteria or where we had insufficient

information to be sure. The inclusion and exclusion criteria were reapplied to the full reports and those that did not meet these initial criteria were excluded.

IDENTIFYING AND CHARACTERISING THE LITERATURE: CHARACTERISATION OF THE INCLUDED PAPERS

The papers remaining after application of the criteria were coded for contextual information in each study/report. We extracted key information from included studies including:

- Location.
- Publication type.
- accountability elements: accountability in general, assessment, inspection and monitoring.
- year of publication.

The main aim of the coding at this stage was to allow a rapid appraisal of the literature in the field, informing decisions for the next stage of the review (see Section 2.4.1 for further details).

IDENTIFYING AND CHARACTERISING THE LITERATURE: QUALITY ASSURANCE PROCESS

To ensure consistency, we carried out three pilot screening exercises on randomly selected titles and abstracts (a total of 133 titles/abstracts) identified through database searching and relevant websites and systematic review papers. Each team member independently applied the inclusion criteria to a set of titles/abstracts and then all met to discuss the decisions made. The team member who had more experience with systematic review explained the decision process for more moderation screening and how we were to undertake the task. Each team member explained their decision and on the whole there was a general consensus for the decisions to include or exclude each title/abstract. At the end of the first moderation exercise, the team decided to add the category of 'abstract unsure' and these papers were subsequently screened by the Principal and Co-Principal Investigators. The final screening moderation on titles and abstracts was done in pairs on a sample of 49 studies. Disagreements were resolved by group discussion before continuing with independent screening.

Two team members independently applied the inclusion criteria on a set of full-text papers (n=49). The overall outcome of the screening moderation resulted in a high agreement rate. Any disagreements were discussed and resolved before continuing with independent full-text screening of the remaining papers.

EPPI-Reviewer was used to manage the review information, for screening coding and synthesis. We kept a record of decisions made at every stage of the review regarding which papers to include/exclude, methodological clarification and how we refined our search strategies.

2.4 IN-DEPTH REVIEW

MOVING FROM BROAD CHARACTERISATION (MAPPING) TO IN-DEPTH REVIEW

The systematic search carried out during the systematic map stage involved identification of relevant literature addressing the review question (Section 2.3). The findings from the systematic map provided a basis for informed decisions about the focus of the in-depth review (See Appendix 3.1). A brief memo was sent out to the Advisory Group members presenting the findings from the systematic map and presenting different options that might be appropriate inclusion criteria for the in-depth review given the evidence identified, and

the time and resources available. Subsequently, a new set of exclusion criteria was developed in consultation with the policy advisors at DfID:

Exclude 1: Regional focus - Based on our systematic mapping of the sources we had identified, we included only papers that had a focus on or were carried out in specific regions that are most relevant to DfID priorities in improving education outcomes for the poorest and most marginalised. This entailed limiting the review to four regions that had the highest concentration of studies: East Asia and Pacific, South Asia, Sub-Saharan Africa, and Latin America. The latter region, Latin America, is not a region of focus for DfID; however, because of the number of relevant studies available and the potential for offering robust comparison and contrast with other regions, we decided to include those relevant to the Latin America region.

Exclude 2: Temporal focus - our initial searches were conducted from 1990 to the present. Systemic national and international focus on accountability policies in LMICs did not get underway until the mid-1990s, and we expected that studies from 2001 would be sufficient to capture this early period of policy sharing and national implementation.

The recommendation from the Advisory Group members is summarised in Appendix 2.4.

DETAILED DESCRIPTION OF THE STUDIES IN THE IN-DEPTH REVIEW

Studies included in the in-depth review were data extracted using a coding tool designed particularly for this review (See Appendix 2.5). The detailed coding tool was largely based on the coding tool use by previous realist review (Westhorp et al., 2014). The first section was designed to assess the relevance of studies that were included in the in-depth review. Those judged to be 'highly relevant' or 'somewhat relevant' were included in the synthesis and were subsequently coded to extract in-depth information about the aims and characteristics of the accountability elements, including theoretical assumptions, the roles of accountability elements, and programme design and implementation (see Section 2.4.3 for further details about assessing the relevance and quality of the studies). We also extracted data on the conditions under which the programme was carried out, and any explicit mention of mechanisms and outcomes reported in the study. The second part was designed to extract data on the aims and objectives of the study, study design, and data collection and analysis approaches. The last section was designed to assess the rigour of the studies included in the synthesis.

An additional purposive search was conducted during the data extraction process to identify additional papers that might be further helpful in refining middle-range theories and in addressing the review questions. This process was iterative, using a snowballing approach.

Chapter 3 presents the key characteristics of the literature identified in this review.

ASSESSING THE QUALITY OF THE STUDIES

The most common approach to quality appraisal in systematic review is to evaluate quality as a property of the research, by examining the design and conduct of the research. Realist synthesis and several forms of qualitative synthesis (e.g., meta-ethnography) add an additional dimension by emphasising quality as an emergent property throughout the process of the review. That is, appraisal of quality needs to be recursively addressed as the review proceeds. Appraisal of the quality of the papers for realist synthesis occurs together with data extraction because of the need to evaluate specific portions or 'evidential fragments' (Gough, et al., 2012, p. 177) of a study in relation to emerging understandings of configurations of context, mechanisms and outcomes. Pawson (2006) and Wong et al. (2013) argue for two dimensions of quality in realist synthesis:

relevance and rigour. Relevance aims to appraise whether and to what degree contributions from a particular paper support, weaken, modify, supplement, reinterpret or refocus the initial rough theory (Greenhalgh, 2014, p. 270). Rigour in realist terms 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).

Our implementation of quality appraisal involved an initial, holistic assessment of a paper based on rigour and relevance.

Relevance: We developed two categories to assess the relevancy of papers included in the in-depth review. We assessed whether a paper aimed or partly aimed to investigate, explore or describe accountability in general or a particular element (e.g. assessment, inspection or monitoring) **– 'Relevance Focus'.** The second aspect, '**Relevance: Theoretical contribution**', was a holistic appraisal of the potential contribution of the evidence to the elaboration and testing of our initial rough theory (i.e. a configuration of mechanisms that cause outcomes under specified conditions) and offered sufficient explanation why an intervention led to a particular outcome, in particular, specifying the particular conditions that triggered causal processes that led to the intended outcomes.

We recognised that the review included a wide range of documents and study types that was useful at different stages of the review processes, including theory development, theory refinement, causal mechanisms and empirical investigation. After several meetings, moderation exercises of the full text papers, we found that many of these included documents deemed to be 'not relevant' to our review focus because they reported only a brief description of accountability elements serving mainly for the purpose of contextual understanding. We agreed that it was important to make the review process manageable by prioritising papers judged to be 'highly relevant' and 'somewhat relevant' on the **Relevance focus** to be included in the synthesis.

Rigour: We adapted existing quality assessment criteria for assessing the methodological rigour of the whole study, including the Mixed Methods Appraisal Tool (MMAT) (Pluye et al., 2009) and DfID (2013b). Studies were assessed according to their methodological quality using the following broad criteria:

- theoretical understanding (quality of the reporting of a study's theoretical and conceptual framework, aims and rationale of the research, theory of change).
- sampling method (steps taken to minimise selection bias and confounding).
- the sufficiency of the strategies reported for establishing the reliability and validity of data collection methods.
- the sufficiency of the strategies reported for establishing the reliability and validity of data analysis methods.

Each reviewer reviewed the data extracted on the descriptive information of policy and intervention programmes, the theoretical and research backgrounds, study aim, study design, data collection, and data analysis. An overall judgement of rigour was assigned for each study in terms of the plausibility and coherence of the method/rationale used to generate data and explanation (see the quality assessment tool in Appendix 2.5.) For a non-empirical paper (e.g., papers that drew inferential claims based on reviews of literature or arrived at logical conclusions based on philosophical arguments), we assessed rigour in relation to discrete aspects of the paper according to the strength of the inferential conclusions in terms of the transparency of the premises and the quality of the underlying evidence.

2.5 SYNTHESIS OF EVIDENCE

OVERALL APPROACH TO AND PROCESS OF SYNTHESIS

Information from included papers was coded and summarised in EPPI-Reviewer 4.0. We extracted information in the form of 'line-by-line' coding or free text with short verbal descriptions (descriptive codes) of the key features of interventions and studies, including the attributes of participants, settings, interventions, outcomes, context and mechanisms. A report for each accountability element was generated that lists all data extracted from the primary studies in a tabular form. Additional sets of reports were also generated for each accountability element by region (e.g. inspection in South Asia).

Syntheses were then conducted for each accountability element following the procedure outlined in more detail below. In summary, the research team read through all data and descriptive codes, elaborating and refining the descriptive codes, and identifying convergent and divergent themes across elements. The researchers then compared and contrasted related features of C-M-O configurations across all elements to identify convergent and divergent features. Relationships and patterns in the data were explored and were considered along with level of quality. Researchers elaborated and tested the synthesis findings through additional mining of existing papers coded as 'accountability in general'. The findings were then used to develop further our initial rough theory by elaborating a more refined conceptual model of the operation of accountability elements independently and in concert.

DATA SYNTHESIS

To address the review question, we synthesised evidence about the ways in which each accountability element led to the outcomes of interest, with a particular emphasis on school-level impact as influenced by impact at the system level and generating impact on student learning.

The papers were coded and analysed in five rounds. The first four concerned each accountability element separately; the final round consisted of comparison across all elements.

In the first round, we coded all the sources for evidence of outcomes and descriptions of context, mechanisms and elements of school inspections contributing to outcomes (see Appendix 4.1). The coding included excerpts from the original text, using an inclusive approach to include both primary findings from research as well as narrative descriptions and hypothetical discussions of assessment, monitoring and school inspections by the authors of the papers, including their reference to work by others. The report of these coded texts informed our second round of analyses.

In this second round, we synthesised separately all of the data extracted for each of the three elements of accountability along with the additional category of accountability in general. A lead researcher was assigned for the synthesis of each accountability element. The researcher read through all data extracted by element, referring to the original source for clarification as necessary, as well as recoding if clarification revealed codes to be incorrectly ascribed. This phase generated a number of descriptive codes to characterise the data in more detail.

The elaboration of descriptive codes for programme characteristics led to the identification of different types of programme activities in our syntheses of Assessment and Monitoring literature. We mapped different types of programme activities to the programme pathways from our initial rough theory (described above) to identify the conditions under which programme activities produced particular outcomes. Clarifying conditions allowed us to compare how *reported or suggested* outcomes described in the literature differed or aligned with the way programmes were designed to work, that is, the intended programme pathways (e.g., high-stakes tests may be intended to work by providing feedback to students and educators to improve student

learning; however, the reported or suggested outcomes included such things as teachers' focus on short-term goals, one of the unintended consequences of high-stakes tests). Unlike assessment and monitoring, the Inspection literature emphasised only one intended type of activity – school visits by government officials. However, we identified several different types of activities relating to impact from the literature on assessment and monitoring. In assessment and monitoring, we used these types of programme activities to examine outcomes. For all accountability elements, our synthesis of outcomes proceeded by organising the findings around common themes and including, for each statement, the rigour of the study the statement was made from, and the type of evidence presented in the statement: whether it was an actual finding from the study (FI), or an inferred claim (CL). An overview of all the evidence statements with an appraisal of the quality of the evidence is provided in Appendices 4.2, 5.1 and 6.1.

In the third round, we characterised conditions that facilitated or impeded outcomes under each pathway for assessment and monitoring or as a whole for inspection.

The concluding round for each accountability element consisted of using constant comparative approaches to analyse the relationships of outcomes to conditions to make inferences about potential mechanisms that are presented in the relevant chapter. In the chapters on assessment and monitoring, we elaborated a hierarchy of impact for each type of programme activity and used this to clarify orders of impact.

These three rounds of analysis around outcomes, conditions and mechanisms correspond with the sections in each chapter on findings about types of programme activities and comprise the synthesis of each element in Chapters 4-6.

In the final round, we then conducted a comparative analysis of the three separate syntheses, identifying the ways in which proposed mechanisms for each element compared and contrasted with the programme pathways of our initial rough theory. This analysis is presented in Chapter 7. The results of this analysis along with the initial conclusions and implications were circulated to the Advisory Group and DfID reviewers for their input.

3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS

This chapter presents the results of the search, the application of the inclusion criteria and a brief description of the characteristics of the included studies in the systematic review.

3.1 STUDIES INCLUDED FROM SEARCHING AND SCREENING

At the beginning of the review process, our searches identified a total of 17,259 citations, 17,144 from bibliographic databases, and 115 from websites and key authors and experts. After removing, 7,004 duplicates, 10,255 records remained and were screened based on title and abstract. The majority of papers were excluded from the review because they were not about accountability or accountability elements (inspection, monitoring, and assessment) (6,176, 60.22%); 2,126 papers (20.73%) were excluded because they were not conducted in low- or middle-income countries (LMICs) and 1,000 (9.7%) were excluded because they were not focused on primary or secondary education. After the title and abstract screening exercise, 823 full-text papers were retrieved. Of these, 566 were excluded largely because they were not relevant to the accountability focus, and 43 papers could not be obtained within the review timescale (before September 2014).

In November 2014, we produced an initial systematic map drawing on 275 papers included in the review at the time to inform the next stage of the review. The findings from the systematic map are presented in Appendix 3.1. After discussed the findings from the systematic map with DfID in consultation with the Advisory Group, by considering the quantity of the evidence identified and the resources available, we narrowed down the scope of the review by region and publication date (see Section 2.4 for further detail). This reduced the number to 214; these are listed in Section 8.1.

We started data extraction by assessing the relevance of the 214 papers. We subsequently evaluated whether these papers: a) clearly articulated or described the focus of the papers on accountability elements (Relevance focus); and b) elaborated/contradicted some aspect of initial rough theory through theory building (C-M-O focus). A total of 68 papers judged to be 'highly' or 'somewhat' relevant on each accountability element (Relevance focus) were included in the synthesis and data were extracted for the intervention characteristics, study design, contexts, mechanisms and outcomes, and assessed for study rigour.

Figure 3.1: Flow of literature identified and included in the synthesis

Papers identified through hand searching N= 115	Papers identified through databases searching N= 17,144

Duplicate reports removed
N = 7,004

Abstract exclusions – N=9,432

EXC1: Published before 1990 N=38

EXC2: Not published in English N=8

EXC3: Conducted in HIC N=2,126

EXC4: Not on accountability elements N= 6,176

EXC5: Not primary or secondary education N =1,000

EXC6: Not clearly state to system N=84

Full text exclusions – N= 566

EXC3: Conducted in HIC N= 59

EXC4: Not on accountability elements N= 376

EXC5: Not primary or secondary education N =30

EXC6: Not clearly state to system N=58

EXC7: Not obtainable N=43

Total records screened
N = 10255

Excluded on title/abstract
N= 9,432

Full reports to be screened
N =823

Additional search (reference checking and AG contacts) N= 39

Excluded on full text
N= 566

Full reports included in the systematic map
N=296

Exclusion criteria for in-depth review N=82
EXC1: Published before 2001 N= 60
EXC2: Regional focus N= 22

Studies included for in-depth review
N =214 studies

Excluded on region and date
N= 82

Exclusion criteria for relevance assessment
EXC: Not 'highly' or 'somewhat' relevance on accountability element(s)
N = 146

Studies included in the synthesis (N= 68)
Assessment N= 34
Monitoring N = 22
Inspection N = 22

3.2 CHARACTERISTICS OF THE LITERATURE IN THE SYNTHESIS (N=68)

This section presents the main characteristics of the 68 papers that describe and explore the conditions under which school accountability may improve system efficiency, service delivery and learning outcomes. Of these, 34 papers focus on assessment, 22 on monitoring and 22 on inspection. Twenty papers focus on more than one accountability element. The next three chapters present the findings of the review on each accountability element: assessment (Chapter 4), monitoring (Chapter 5), and inspection (Chapter 6).

YEAR OF PUBLICATION

Figure 3.2 presents the numbers of papers published from 2001. The number was at its greatest at 2010. There was a sharp increase in the papers relevant to the review question from four papers in 2009 to 14 papers in 2010. Then, a smaller number of publications was identified each year in the last four years when compared with those identified in 2010, decreasing to eight in 2011, four in 2012, five in 2013, and one in 2014. Since the searches were undertaken in 2014, it is probable that the figure for that year is an underestimate.

Figure 3.2: Year of publication (N=68), code mutually exclusive
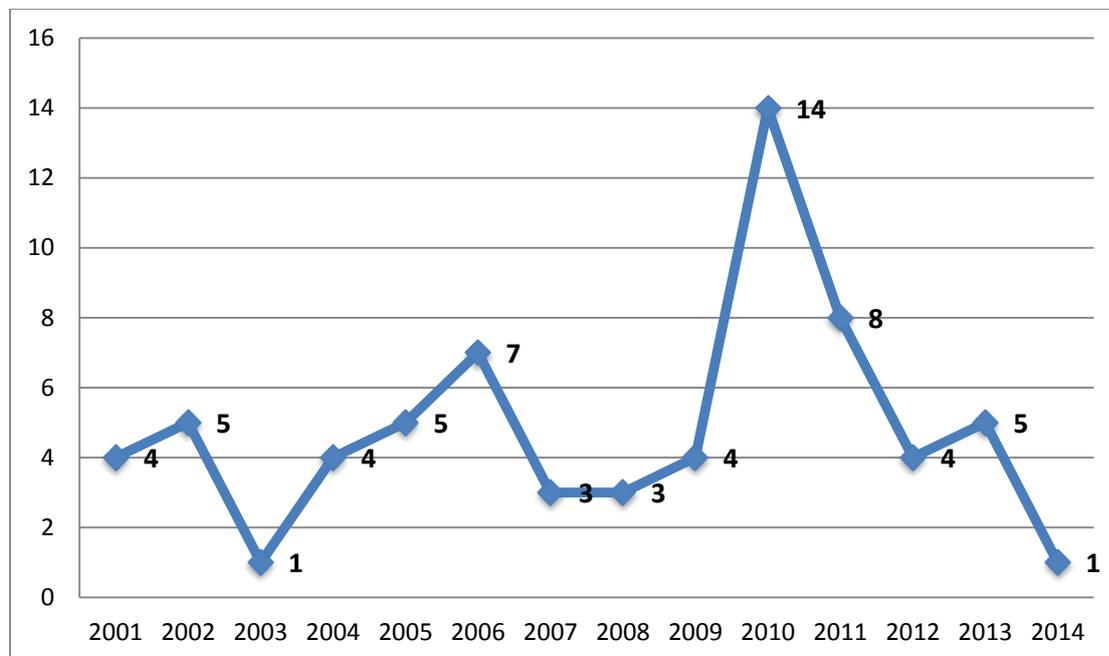
PUBLICATION TYPE

As shown in Figure 3.3, nearly half of the papers were published as peer review journal articles (n=34, 50%). A smaller number were published as research reports (n=23, 34%), with the remaining 16% consisting of policy-relevant document (n=7, 10%), dissertations (n=2, 3%), book/book chapter (n=1, 1.5%) and conference paper (n=1, 1.5%).

Figure 3.3: Breakdown of papers by publication type (n=68), code mutually exclusive
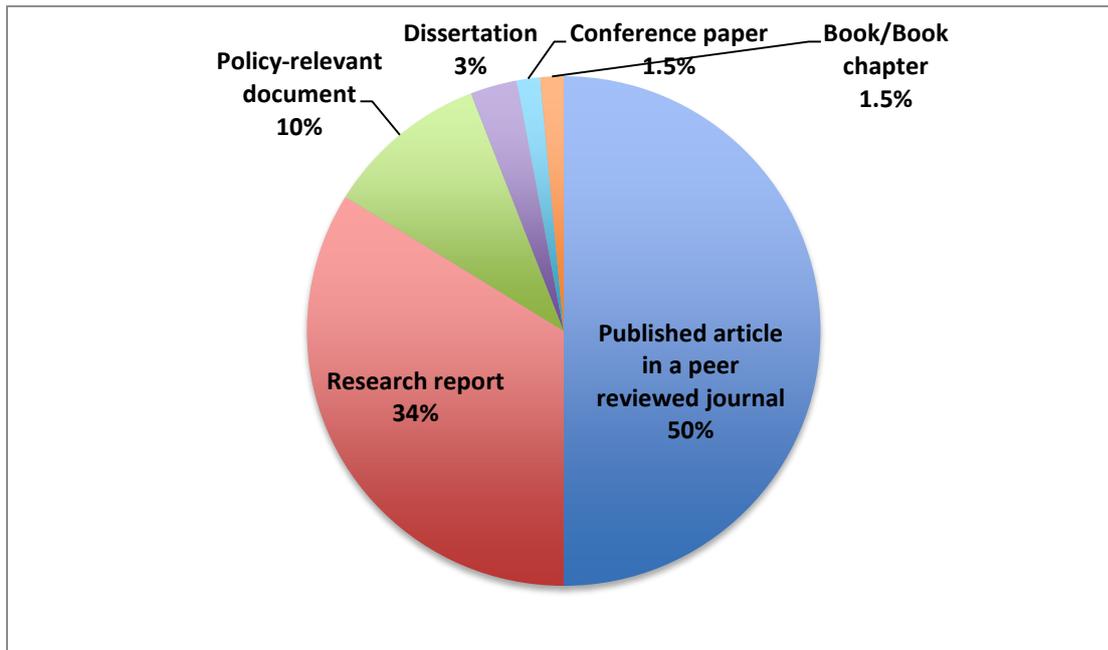
ACCOUNTABILITY ELEMENTS BY GEOGRAPHICAL LOCATION AND INCOME LEVELS

A large portion of the 68 included papers is from sub-Saharan Africa (n=33, 47%). The remainder are distributed between three regions: 9 papers (13%) from Latin America and the Caribbean, 13 (19%) from South Asia, and 8 (11%) from East Asia and the Pacific. When grouping by accountability element, the geographical distribution shows a similar pattern in all accountability elements, with the majority of papers being from sub-Saharan Africa and a smaller number of papers from the other three regions (see Figure 3.4). Eleven papers focus broadly on 'developing countries', with no focus on one particular country or region.

Figure 3.4: Spread of literature in each accountability element by region (N=57)*
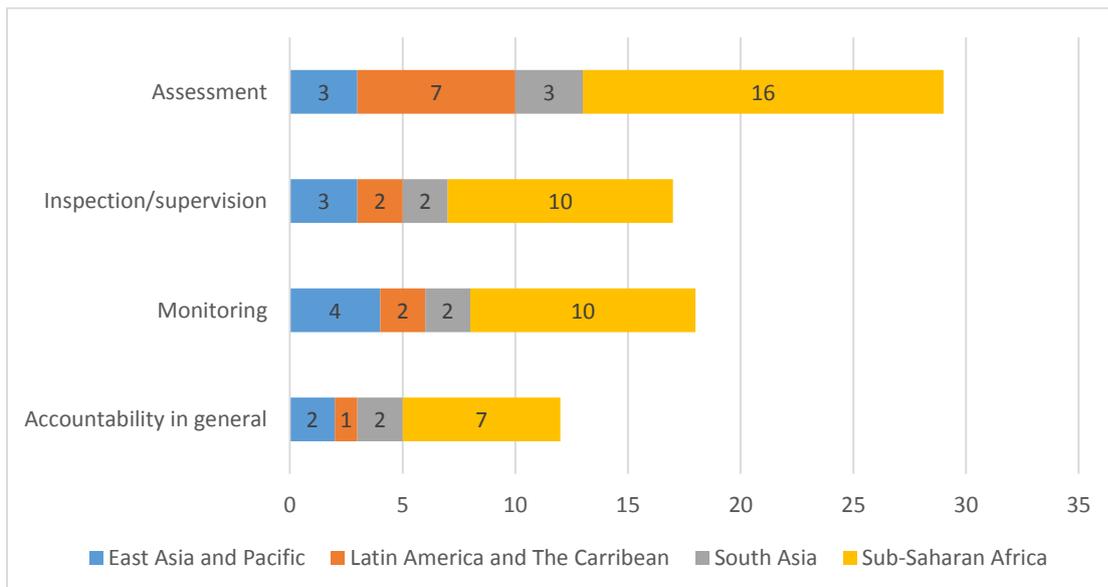
*When reported, codes not mutually exclusive

Breaking down by income level[4] when information is available (n=50), 23 of the papers identified are from upper-middle income countries, 18 from lower-middle income countries, and 12 from low-income countries. The majority of evidence on assessment is identified from middle-income country contexts (see Figure 3.5).

Figure 3.5: Spread of literature by income level (n=50)*

*Codes not mutually exclusive

---
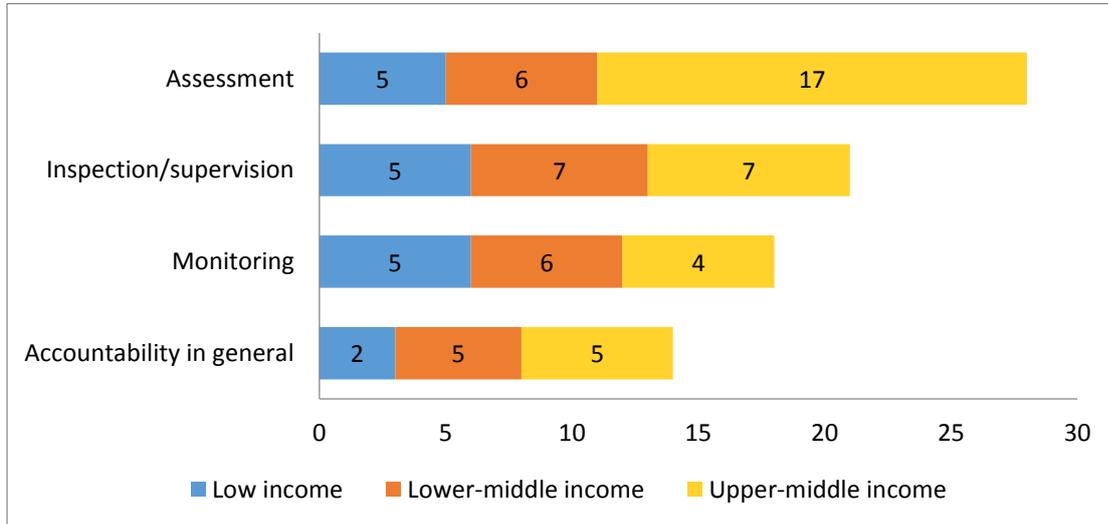
4. IN-DEPTH REVIEW: ASSESSMENT

4.1 INTRODUCTION

This chapter explores the conditions under which assessment activity leads to improvement in schools and to positive learning outcomes for schoolchildren in low- and middle-income countries (LMICS), especially the poorest and most marginalised. Thirty-four papers focusing on standardised assessment were included in the in-depth review. These papers were data extracted and assessed for their relevance and rigour as described in Chapter 2. The synthesis findings presented in this chapter were generated from papers discussing assessment in Sub-Saharan Africa (n=16) and Latin America and the Caribbean (n=7) with less than a quarter of included papers covering East Asia (n=3) and South Asia (n=3). Five papers focused on assessment activity across developing countries rather than a country-specific context.

This chapter explores standardised assessment as one particularly significant area of the wider universe of monitoring activity, which we discuss in the next chapter. The boundaries between the two, monitoring and standardised assessment, are blurred and our discussion of the use of performance information derived from standardised assessment overlaps with our discussion of monitoring. For example, Santiago et al. (2012) is a detailed narrative overview of education reform in Mexico. The authors describe uses of student- and school-level performance information, which qualifies the overview as an example of monitoring activity; however, most of the narrative concerns the development of different forms of standardised assessments, which makes it more appropriate to discuss in the context of this chapter on Assessment.

We present the synthesis of the assessment papers, reporting in six major sections:

Section 4.2. Defining assessment

Section 4.3 Quality of studies

Section 4.4 Findings about types of activities

Section 4.5 Conclusion

4.2 DEFINING ASSESSMENT

Our definition of assessment, as conceived at the beginning of our review process, appears in Chapter 1. This definition is primarily concerned with the evaluation of schools through standardised assessment of student learning outcomes. Our synthesis led us to elaborate this definition by including assessment activity that had as its aim the development of school-level service delivery, as well as the evaluation of schools. As noted in the introductory chapter, we have included the term 'standardised' in our mention of assessment in order to emphasise the connection with system-level processes. 'Standardised' refers to the close involvement of a wider system in 'test design, content, administration and scoring to ensure comparability of the results across students and schools' (Best et al., 2013, p. 2). Our initial definition concentrated on standardised assessment processes, content and scoring, with the purpose of providing school-level comparisons of aggregate student performance. Several papers in our synthesis (Kellaghan and Greaney, 2004; Lubisi and Murphy, 2002; Postlethwaite, 2004; Taylor, 2009) commented on large-scale shifts towards the use of standardised assessment results to evaluate school quality, as well as increasing use of assessment processes, in addition to results, to improve student learning through school-level improvements – improvements in teaching practices as well as school management practices.

4.3 QUALITY OF THE STUDIES

The quality of the 34 papers included in the assessment synthesis was assessed using the method described in Chapter 2. Six papers were judged as high on rigour; 18 as medium and 10 as low (See Table 4.1). Half of the papers were published as peer reviewed journal articles (n=17). The remaining half consisted of research reports (n=7), policy-relevant documents (n=8), and books/book chapters (n=2). The papers included 11 theoretical or position papers, four literature reviews related to assessment, six country or provincial case studies and two qualitative studies. Eleven employed quantitative evaluation designs; of these four were randomised controlled trials (RCTs) of interventions related to assessment activity. The papers included in the review provide narrative descriptions of assessment activity across a number of countries, particularly in Africa. The majority were from reviews of documents, surveys of education officials or small-scale case studies, often including limited descriptions of the underlying methodology and presenting self-reports of small (potentially non-representative) samples of respondents.

Table 4.1: Reviewers' judgements about rigour and relevance of each study included in the assessment synthesis

Studies (first author and date)	Rigour			Relevance: Focus		
	High	Medium	Low	High	Medium	Low
Bansilal (2011)		✓		✓		
Barrera-Osorio (2010)	✓			✓		
Barrett (2011)			✓		✓	
Beets (2011)			✓		✓	
Braun (2006)		✓			✓	
Brown (2011)	✓			✓		
Castro (2003)		✓			✓	
Chisholm (2013)			✓	✓		
Crouch (2008)			✓	✓		
De Grauwe (2008)			✓	✓		
Ferrer (2006)		✓			✓	
Glewwe (2010)	✓			✓		
Gvirtz (2002)			✓		✓	
Gvirtz (2004)		✓			✓	
Howie (2012)			✓		✓	
Kapambwe (2010)		✓		✓		
Kellaghan (2001)		✓			✓	
Kellaghan (2004)		✓		✓		
Kremer (2004)	✓			✓		
Lassibille (2010)	✓			✓		
Lubisi (2002)			✓	✓		
Luxia (2005)		✓		✓		
Mukhopadhyay (2011)			✓	✓		
Muralidharan (2011)	✓			✓		
Nsibande (2012)		✓			✓	
Ong (2010)		✓			✓	
Postlethwaite (2004)		✓			✓	

Studies (first author and date)	Rigour			Relevance: Focus		
	High	Medium	Low	High	Medium	Low
Pryor (2002)			✓	✓		
Ravela (2001)		✓			✓	
Ravela (2002)		✓			✓	
Reyneke (2010)		✓		✓		
Santiago (2012)		✓			✓	
Scherman (2011)		✓			✓	
Taylor (2009)		✓		✓		

4.4 FINDINGS ABOUT TYPES OF ACTIVITIES

We identified two distinctly different types of national assessment programmes from the literature. The first type are **high-stakes examinations**, with emphases on setting expectations, providing feedback/consequences, and capacity development of educators as the anticipated ways that examinations aim to influence students, teachers and school leaders. A more recent approach to assessment is **low-stakes assessment**, which also aims to affect what happens at the school level through setting expectations, providing feedback/consequences and capacity development of educators. However, low-stakes implies the use of indirect influence rather than the anticipation of negative consequences from poor performance to bring about change.

Usage of the names of different types of assessments and different assessment typologies is highly varied across the literature. We use the basic distinction of 'high-stakes' examination to point to any assessment activity that carries with it *consequences* for schools, school managers, teachers and/or students. 'Low-stakes' assessments are those that have no direct consequences for individuals or organisations, as is typical of national assessments that aim to gauge educational quality. National assessments and national examinations have historically had very different purposes – the former to diagnose the health of the system, the latter to gate-keep between primary, secondary and higher education. However, the more recent global emphasis on educational quality (Bruns et al., 2011) has meant that examinations are now frequently used in ways that aim to improve the quality of teaching and learning, not just provide qualifications for individual students (Kellaghan and Greaney, 2001).

High-stakes examinations include public examinations, such as South Africa's National Senior Certificate or Kenya's Primary Education Certificate, as well as national assessments that are primarily designed to monitor the education system but also carry consequences for performance for schools and school teachers as well as individual students. For example, census-based, national performance assessments are used in a number of Latin American countries to monitor the system, but also carry consequences for students in terms of passing to another level, to teachers in terms of incentives, and to schools in terms of eligibility for additional resources (Ferrer, 2006; Santiago et al., 2012). Within the frame of high-stakes examinations, the papers led us to identify four distinct programme pathways that this form of assessment aims to use: setting expectations - indirect impact through efforts to improve results (Section 4.4.1); providing feedback/consequences - direct and indirect impact through information and incentives (Section 4.4.2); and capacity development of educators - school-based performance assessment as a component of high-stakes examinations (Section 4.4.3).

Low-stakes assessments, such as national assessments, were introduced as a means of monitoring educational quality. According to Kellaghan and Greaney (2004) national assessments sought to capture 'the level of achievements, not of individual students, but of a whole education system or a clearly defined part of one' (p. xi). This may be a specific grade or a series of grades in elementary, middle or high school. National

assessments may be sample-based relying on data from sampled groups of students in schools who respond to assessment instruments and questionnaires designed nationally or regionally. They may also be census-based, collecting data from all students within a specific grade or series of grades. Data may also include responses from teachers and/or school leaders to questions deemed relevant to interpretations of their students' achievements. Within the frame of low-stakes assessment, two main categories of activity that aim to produce impact at the school-level appear in the papers: setting expectations – the establishment of curriculum standards (Section 4.4.4); and capacity development of educators - guidance to teachers and school managers (Section 4.4.5).

HIGH-STAKES EXAMINATION: SETTING EXPECTATIONS THROUGH STUDENT, TEACHER AND SCHOOL EMPHASIS ON PERFORMANCE

The literature exploring the impact of examinations on schools suggests that high-stakes examinations can influence service delivery at the school level through anticipation of consequences in both intended and unintended ways well before students sit exams.

The evidence drawn from the 13 papers we review in this section is of medium (9) and low (4) rigour, including the inferential claims of overviews of assessment practices in a number of different countries (e.g., Barrett, 2011; Ferrer, 2006; Kellaghan and Greaney, 2004) and claims made based on reviews of policy and project documents in individual country reports (e.g., Kapambwe, 2010). All papers identify unintended consequences that affect schools, teachers and students in undesirable ways. Three medium-rigour papers (Castro and Tiezzi, 2003; Ferrer, 2006; Kellaghan and Greaney, 2004) also describe examinations achieving intended outcomes to improve the quality of teaching and learning and align what is taught with what is assessed.

Unintended consequences of high-stakes examinations may appear at the level of the school as a whole or at the level of classroom practice. Narrative overviews from two medium-rigour studies (Braun et al., 2006; Kellaghan and Greaney, 2004) and one of low rigour (Mukhopadhyay and Sriprakash, 2011) of the impact of high-stakes examination on school management suggest organisational strategies and instructional practices commonly employed to manipulate reports of school performance: (a) reducing the number of low-achieving students; (b) narrowing the curriculum to focus on assessed disciplines; or (c) targeting students most likely to succeed. Five studies (four medium- and one low-rigour) report that high-stakes examinations might influence instructional practice by motivating teachers and school managers to: (a) increase classroom testing that focuses on exam preparation (Mukhopadhyay and Sriprakash 2011); (b) narrow the focus to only what is measured by the exam (Kellaghan and Greaney, 2001, 2004); (c) spend class time on coaching students on assessment instruments (Kellaghan and Greaney, 2001; Luxia, 2005; Santiago et al., 2012); and (d) increase emphasis on drilling, memorising, rehearsing, rote learning (Kellaghan and Greaney, 2004). Such impact may spread from examined years to lower, non-assessed years (e.g., the prevalence of the multiple-choice format in lower grades) (Kellaghan and Greaney, 2004).

BOX 4.1: Case examples of impacts of high-stakes assessment on schools and teachers

Brazil: Castro and Tiezzi (2003) provide an overview of ENEM, the voluntary secondary certificate in Brazil, and intimate that this examination has had a salutary impact on teacher knowledge, especially performance in relation to knowledge of the discipline:

> The ENEM has made it possible to gain a more palpable understanding of the pillars structuring secondary education reform: an interdisciplinary approach, putting learning into context and solving problems; it has allowed teachers and education specialists to visualise clearly the desired performance of young people, as is required by each of the subjects. (Castro and Tiezzi, 2003 p.14)

ENEM, they report, came to be accepted as, 'a powerful instrument to induce change insofar as it expresses what should be taught through what it assesses' (Castro and Tiezzi, 2003, p.14).

Latin America: Ferrer (2006), in a descriptive review, summarises the potential for examinations to serve as levers to improve service delivery in Latin America. His summary suggests that high-stakes tests can foster curricula and pedagogy that are centred on the student and not on the exam (Ferrer 2006):

> accreditation and incentives mechanisms do not necessarily mean that classroom efforts are reduced to 'teaching to the test.' In fact, many schools with high scores on standardized tests have made their own education and curricula proposals that place a marked emphasis on the integral, formative role of their students. (Ferrer, 2006 p.50)

Sub-Saharan Africa and Carribian: Kellaghan and Greaney (2004) also invoke the power of high-stakes examinations to influence teaching practice, again drawing on descriptive, qualitative case studies of assessment in Kenya and Trinidad and Tobago.

The available evidence suggests that if the content areas of examinations are changed (for example, if a new subject or a new component of a subject, such as essay writing, is examined), the content to which students are exposed in class will also change. (Kellaghan and Greaney, 2004 p.20)

However, they also caution that while implications for pedagogical practice may be clear, those for student learning are less so: 'The evidence regarding changes in student achievement levels and cognitive processing skills is less clear. Where improvements do occur, they are likely to be modest.' (Kellaghan and Greaney, 2004 p.20)

Several strategic responses identified above may be perceived as undesirable depending on the context. For example, school policies around increased retention or disability classification for low-performing students would most likely affect student learning and progress in undesirable ways in any situation, as would instructional practices oriented towards rote learning, especially for an assessment that aimed to measure and promote higher order cognitive skills (Kellaghan and Greaney, 2004).

Other reported impacts on schools and teachers may be beneficial, depending on the circumstances. Coaching students on assessment instruments may be highly positive if those instruments are designed to foster students' independent, higher-order thinking and teachers' abilities to monitor student learning more closely (Kapambwe, 2010) (see Box 4.2 for additional examples.)

Five papers (three of medium rigour, two low) suggested that the unintended consequences of high-stakes examinations are most likely to have the most adverse effects on the education opportunities of poor and marginalised children because of the strong association between low achievement on examinations and social disadvantage (Beets and van Louw, 2011; Gvirtz and Larripa, 2004; Howie, 2012; Kellaghan and Greaney, 2004; Reyneke et al., 2010).

One review paper suggested that the undesirable impact of high-stakes examinations had repercussions on the motivation to learn for all students. Kellaghan and Greaney (2004) summarised in a medium-quality review of assessment in sub-Saharan Africa that the impact on student learning of high-stakes examinations was predominantly undesirable. They reported that high-stakes examinations promoted students' use of short-term learning strategies, emphasised extrinsic reward, and led to students' decreased motivation for mastering content and higher-order thinking skills.

When high stakes are attached to performance, students tend to be less successful in acquiring higher-order and transferable skills; learning tasks are perceived as not inherently interesting; and, if a reward is removed, students will be less likely to engage in a task (Kellaghan and Greaney, 2004 p.23).

HIERARCHY OF IMPACT

Table 4.2 articulates the dependencies among orders of impact in six of the papers that described service-level outcomes in detail, three of medium rigour (Braun et al., 2006; Castro and Tiezzi, 2003; Kellaghan and Greaney, 2004) and three of low rigour (Howie, 2012; Kellaghan and Greaney, 2001; Mukhopadhyay and Sriprakash, 2011). To infer a hierarchy of intermediate outcomes, we begin with programme activities in the first column and proceed through first-order impact, which corresponds with direct outcomes from implementation activities; second-order impact in the next column related to the reported consequences of implementation; and third-order impact, the ultimate outcomes reported in corresponding papers. This chain of impact is what we would expect to lead to or detract from our ultimate outcomes of interest for this review, primarily service delivery, but also learning outcomes and system efficiency.

Table 4.2: Hierarchy of impact: High-stakes examinations: Setting expectations

Activity	First-order	Second-order	Third-order
High-stakes examinations – setting expectations - undesirable impact			
School emphasis on test performance (Braun et al., 2006; Howie, 2012; Kellaghan and Greaney, 2001)	Emphasis on high performing students for selection, progression Emphasis on 'borderline' students for test performance	Restricted opportunities for low-performing student progress Increased grade retention for low-performing students	School-wide barriers to academic progress and success for low-performing students School-wide devaluation of the potential of low-performing students (deficit assumptions)

Activity	First-order	Second-order	Third-order
Teacher emphasis on test performance (Kellaghan and Greaney, 2004; Mukhopadhyay and Sriprakash, 2011)	Teaching emphasis is on successful completion of test items, not student content mastery Increased use of short-term teaching strategies	Teaching emphasis is on examination results Increased use of short-term learning strategies by students	Teacher devaluation of the potential of low-performing students (deficit assumptions) Teaching focus is only on assessed subjects Student emphasis is on extrinsic reward, not content mastery Teacher emphasis is on technical compliance, not student content mastery Teacher turnover in low-performing schools Teacher alienation and burn-out, especially in low-performing schools
High-stakes examinations – setting expectations - desirable impact			
Examination emphasis on higher-order thinking and content mastery (Kellaghan and Greaney, 2004)	Increased use of teaching strategies that emphasise students' higher-order thinking and content mastery	Increased use of learning strategies by students that emphasise higher-order thinking and content mastery	Curricular emphasis on assessed content Teaching and learning emphasis on content mastery
Supportive guidance and professional development for teachers and school leaders (Kellaghan and Greaney, 2004)	Increased teacher knowledge and skills	Greater emphasis on higher-order thinking skills in classroom teaching	Student mastery of content Teacher understanding of the principles of examination

Activity	First-order	Second-order	Third-order
Acceptance of examination emphasis on higher-order thinking and content mastery (Castro and Tiezzi, 2003)	Adoption of examinations focused on higher-order thinking and content mastery by schools and/or school systems	Acceptance by schools, parents and students of the results of examinations Acceptance of examination results for admission to higher education	Legitimacy of examinations emphasising higher-order thinking skills

C-M-O CONFIGURATIONS: SETTING EXPECTATIONS

The hierarchy of impact highlights potential causal pathways from the ways that outcomes are connected. Although the first- and second-order impacts do not explain how they might generate third-order impacts, we can infer precursor conditions that are necessary for particular outcomes to arise. From these, we can infer potential mechanisms triggered by those conditions that could plausibly yield the outcomes reported or suggested. The generic pathway of setting expectations is prominent in papers that discuss the undesirable impact of high-stakes assessments. We propose one mechanism that produces undesirable impact, **fear of bureaucratic authority**, in relation to high-stakes examinations. We infer this to be triggered by a range of conditions that are mentioned repeatedly in relation to schooling in countries with severely limited resources. These conditions include school-leadership and external pressure for results coupled with a lack of external and school-based supports for teaching, leading and making sense of exam results. The lack of support coupled with pressure for performance appears to trigger fear in the school leadership and teachers, resulting in efforts to mitigate the negative consequences of low achievement for the school. Efforts move beyond mere compliance with bureaucratic authority (e.g., instrumental focus on meeting minimum thresholds) to active manipulation of the population of students sitting exams as well as examination results.

One initial condition that appears to apply across desirable outcomes is belief in the credibility of both the form of the examination and the results produced. We named this 'pedagogical authority' because the newly introduced forms of examination described in the papers were all moving towards ways of promoting higher-order thinking and away from types of questions that could be answered by memorisation and strategic preparation for testing. Given the right conditions, trust in the pedagogical authority of the examination may lead to teachers' engagement with the principles underlying the design of the examination, the development of teaching knowledge and skills that go beyond performance alone, and the cultivation of instructional leadership. In that sense, **trust in the pedagogical authority of an examination** is a mechanism because it is an aspect of reasoning that precedes teachers' and school leaders' abilities to use the exam as a means of improving student learning. However, trust is only triggered by an array of supports that includes external as well as internal features – external in the provision of relevant guidance and the availability of expertise, and internal in the forms of knowledgeable and supportive peers and school leadership. One supportive condition that may be especially important is productive persistence. This means that teachers and school leaders are willing to continually work with the examination emphasis (e.g., a shift from multiple-choice questions to essay writing) because they appreciate the potential longer-term benefit to their students' knowledge and abilities; in other words, they accept the pedagogical authority of the form of examination. There is a suggestion in

Ferrer (2006) that this developmental emphasis is most likely to be a feature of already high-achieving schools, and there is no indication that low-performing schools, those most likely to be serving the poorest and most marginalised children, are able to provide such supportive conditions in the face of pressure for exam results.

Table 4.3: C-M-O high-stakes examinations: Setting expectations

Conditions	Mechanisms	Outcomes
No internal support for teaching No instructional leadership No internal or external support for interpreting results No external support for teaching or instructional leadership External pressure for results School leadership pressure for results	Fear of bureaucratic authority	Teaching emphasis on examination results leading to unintended consequences Increased use of short-term teaching and learning strategies Teacher emphasis on technical compliance not content mastery School-level restrictions on opportunities for low-performing student progress
Internal (school-level) support for teaching and instructional leadership External support for high-quality teaching and instructional leadership Internal and external support for interpreting results School and parental emphasis on examination results Appreciation for productive persistence External pressure for results (parents and community; media) School system pressure for results	Trust in pedagogical authority of assessment approaches	Teacher understanding of principles and purposes of assessment Teaching emphasis on examination results leading to intended consequences Teaching and learning emphasis on higher-order thinking and content mastery Alignment of content taught and assessed Productive persistence and student progress

HIGH-STAKES EXAMINATIONS: PROVIDING FEEDBACK THROUGH INFORMATION AND INCENTIVES

PROVIDING FEEDBACK THROUGH INFORMATION

The use of the results of high-stakes examinations includes individualised reports to students as well as customised guidance to schools and teachers. As above, our synthesis here is largely descriptive, relying on narrative accounts in four medium-rigour papers that report the assessment practices of particular countries, including two regional, cross-country comparisons (Ferrer, 2006; Kellaghan and Greaney, 2001, 2004; Santiago et al. 2012). None of the papers we analysed reported on the direct impact of high-stakes assessment on student learning (i.e., improvement or lack of improvement in student learning as a direct consequence of feedback from examination results, the expected pathway).

Ferrer (2006) highlights how Latin American countries that use high-stakes tests have, 'devised more sophisticated reporting formats that are more useful from a pedagogical and curriculum perspective for teachers and students' (Ferrer, 2006 p.73). This is particularly true for those countries in which the examinations are used to accredit students when they leave high school or for admission to higher education (e.g., Mexico, as reported by Santiago et al. 2012). In sub-Saharan Africa, Kellaghan and Greaney (2004) discuss Uganda, Zambia, and Kenya as illustrations of countries with examination authorities that have emphasised feedback. Of Kenya, Kellaghan and Greaney (2004) comment that the feedback included analysis of the results of individual items with guidance information around addressing problems apparent from analysis of candidate responses. The information used to be disseminated to all schools through a printed newsletter, but as of 2004, was only available for purchase from the Examinations Authority.

Ferrer (2006) identifies two countries in Latin America that provide students with individualised results, the Dominican Republic and El Salvador.

- Dominican Republic: The students receive an individual report of their national test score by domain or skill, as well as by performance level, in each knowledge area. They are also given a narrative analysis of their main achievements and difficulties, and a table that compares the results to both the national average and the average of schools at the same socio-economic level or in the same geographic area. Finally, they receive a series of specific recommendations on how to improve their academic performance (Ferrer, 2006 p.32)
- El Salvador: Learning and Aptitude Test for High School Students (PAES) 'provide every student with information on the skills tested in each subject, the extent to which those skills have been attained, the overall performance average in each area, and a conceptual explanation of the levels of attainment reached' (Ferrer, 2006 p.32) The test counts for 20 percent of the passing grade in each of the main subjects.

No specific impact is mentioned; however, Ferrer claims that the information may be of use to failing students who are required to retake the assessment.

Two of the review papers (Kellaghan and Greaney, 2001, 2004) summarised information about the use of assessment guidance and results in teacher training as well as continuing professional development in Uganda, Swaziland and Chile. Kellaghan and Greaney (2001) claim that continuing support for teachers is necessary as teachers seek to implement changes that the assessment aims to introduce. The authors provide an example from Uganda, in which the unit responsible for overseeing the national examination worked with teachers, teacher trainers and school inspectors in workshops and seminars:

> Implications for teaching were considered, as well as how to use information to adapt the instructional process to improve learning. Teachers were expected to shift their emphasis in the way they presented curriculum material, and to pay more attention to areas of knowledge and skill that the national assessment identified as being relatively weak. (Kellaghan and Greaney, 2001 p.57)

> The workshops involved all teachers not only those in the examination years in an effort to develop organizational capacity for strengthening performance (Kellaghan and Greaney, 2001 p.57)

The papers claimed that the combination of teacher guidance with customised student reports has the potential for desirable impact. Kellaghan and Greaney cite their own and others' work (Rollnick, 1998) in asserting that 'Guidance provided to teachers in the preparation of students for examination, coupled with the development of teacher understanding of the demands of examinations, can lead to greater emphasis on the classroom teaching of higher-level skills' (2004 p.18).

PROVIDING FEEDBACK THROUGH INCENTIVES

We synthesise results from five empirical studies: one case study judged to be of low rigour (Mukhopadhyay and Sriprakash, 2011); and four experimental studies judged to be of high rigour (Barrera-Osario and Raju, 2010; Glewwe et al., 2010; Kremer et al., 2004; Muralidharan and Sundararaman, 2011). Two experimental studies found positive results for sustained student test score gains through the use of two different types of individual incentives, merit scholarships for girls in rural Kenya (Kremer et al., 2004) and individual teacher bonuses in rural India (Muralidharan and Sundararaman, 2011). Two other experimental studies were of group incentives by school in Kenya (Glewwe et al., 2010) and by school and year-level teaching group in Pakistan (Barrera-Osario and Raju, 2010); neither showed sustained gains in student test performance, a result similar to the Indian study mentioned earlier (Muralidharan and Sundararaman, 2011), which compared individual and school-level incentive programmes. The low-rigour study (Mukhopadhyay and Sriprakash, 2011) described unintended consequences from an examination programme in India that offered school-level and individual teacher and student incentives. In sum, these papers propose that individual incentives are more likely than group incentives to result in sustained test score gains for students.

Performance incentives are bonuses offered to students, teachers or groups of teachers within a school for exemplary performance on a public examination or census-based standardised assessment. In our review, we include incentives as part of the high-stakes examination pathway of 'providing feedback/consequences' because incentives intend to amplify feedback about results through the desirable consequence of reward or the negative consequence of forfeit for schools, teachers or students. While these incentives are thought to be powerful ways of influencing individual and organisational behaviour, the findings of impact in the high-rigour studies are mixed (Barrera-Osario and Raju, 2010; Glewwe et al., 2010; Kremer et al., 2004; Muralidharan and Sundararaman, 2011). Two high-rigour studies reported the mixed impact on schools, teachers and students were those that involved school-level incentives (Barrera-Osario and Raju, 2010; Glewwe et al., 2010), with one high-rigour study in Kenya (Glewwe et al., 2010) suggesting that the anticipation of consequences led to undesirable, unintended instructional practices (e.g., teaching to the test, narrowing of the curriculum). Two RCTs looked at individual incentives, one for teachers in rural areas of India (Muralidharan and Sundararaman,

2011) and the other for female students in rural areas of Kenya (Kremer et al., 2004). Both of these studies of individual incentives reported significant positive results with limited unintended consequences.

The quasi-experimental study of Barrera-Osario and Raju (2010) investigated a programme that aimed to expand opportunities for schooling for children from low-income households, as well as to influence the quality of education offered by the high numbers of private schools serving low-income families in Pakistan. To benefit from the subsidy, the school had to achieve a minimum pass rate in the Quality Assurance Test (QAT) and agree to waive tuition and other fees for all students. Once in the program, schools and groups of teachers within schools were eligible for bonuses, including group bonuses for teachers whose students achieved high QAT pass rates and bonuses for the highest-ranking schools in each programme district (Barrera-Osario and Raju, 2010).

The availability of the programme induced large learning gains for students in borderline schools, those marginally failing to meet the minimum pass rate in the previous year. However, once qualified for the programme, the prospect of additional bonuses for teachers based on student achievement and test participation did not result in any additional learning gains. The authors concluded that schools and teachers did whatever was necessary to meet the minimum pass rates to qualify for incentives but did not seek additional incentives by continuing to improve student learning in subsequent years:

> 'apart from the pressure from below to maintain a minimum level of learning for program participation, program schools do not face any effective incentives to continuously raise learning' (Barrera-Osario and Raju, 2010 p.38).

Glewwe et al. (2010) was an RCT that examined a programme run by a Dutch NGO in Western Kenya that offered prizes to schools based on the mean performance in all tested subjects in the annual district exams. Schools could compete for one of two awards: 'top-scoring schools' or 'most-improved schools' (p. 16). The incentives intended to improve overall teaching performance, including reducing teacher absenteeism (p. 17) by promoting mutual teacher accountability. It was anticipated that colleagues would hold one another to account for student performance. However, the results of the RCT suggested that teacher behaviour was not affected by the incentive programme in terms of teacher attendance, pedagogy practices or homework assignment. Moreover, it was evident that schools in the incentive programme conducted more test preparation sessions than comparison schools, and devoted more time explicitly to exam preparation. The findings are consistent with the assumption that the incentive programme could affect short-term school and teaching behaviours but not students' learning. The findings on student outcomes showed that there was no difference in test scores between students in the schools participating in the programme compared to those in the comparison schools after the programme period. Furthermore, while the programme increased student participation in exams, it did not influence dropout and repetition rates (Glewwe et al., 2010).

A case study of a voluntary assessment initiative for elementary schools in the Indian state of Karnataka (Mukhopadhyay and Sriprakash, 2011) also looked at school-level incentives. Schools that opted into the programme needed to meet specified targets for 'learning achievements'. High-performing schools were rewarded with cash incentives and there were also individual rewards for high-performing students and teachers. The authors concluded from conversations and non-systematic observations that low-performing schools made every effort, including malpractice and stringent regular testing, to boost their scores, while adequately performing schools did not strive towards higher performance. Low student performance triggered a state requirement to provide remedial teaching, which was perceived by school leaders and teachers as an undesirable consequence. The study authors claimed that the association of the need for remedial teaching with low performance for some students exacerbated teachers' existing deficit assumptions about lower-caste

students who had a history of lower achievement on standardised tests (Mukhopadhyay and Sriprakash, 2011).

The remaining two studies, Muralidharan and Sundararaman (2011) and Kremer et al. (2004), examined individual teacher and student incentives in rural regions of India and Kenya, respectively. Muralidharan and Sundararaman (2011) reported on a field experiment in the state of Andhra Pradesh, India, comparing two different types of incentives, both based on student performance in public examinations. The incentives rewarded teachers either as a group or as individuals. Teachers in the individual incentive schools received bonus payments based on average improvement of mathematics and language test scores for students in their classrooms. Every teacher in the group incentive schools received the same bonus based on average improvement in test scores across the school. Comparison schools were randomly allocated resources that consisted of either an extra contract teacher or a block grant of cash, regardless of test scores. The findings showed significant gains for students in both group and individual incentive schools, with both groups performing equally well in the first year. However, at the end of the second year, the authors reported that the individual incentive schools outperformed the group incentive schools: 'At the end of 2 years, the average treatment effect was 0.28 SD in the individual incentive schools compared to 0.15 SD in the group incentive schools, with this difference being significant at the 10 percent level' (Muralidharan and Sundararaman, 2011).

Unlike Glewwe et al. (2010), teachers in these incentive schools did not appear to be tailoring classroom activities solely to improve test performance. Follow-up interviews with teachers led researchers to suggest that teachers were not using different teaching strategies but had increased the intensity of their effort using existing strategies. Teachers in incentive schools were more likely to increase their effort by assigning additional homework, offering extra lessons and providing extra support to weaker students (Muralidharan and Sundararaman, 2011).

Kremer and others (2004) reported on results from an RCT trial of the Girls Scholarship Program in two rural districts of western Kenya. The programme consisted of awards to top-performing sixth grade girls for the subsequent two years of schooling. The awards consisted of payments to a winner's school and to her family to cover the costs of school supplies, textbooks and uniforms (US$19.20) and the recognition of award winners at a school awards assembly organised by the sponsoring NGO. The findings from the study on the impact of the incentive programme on both student and teacher outcomes are encouraging. The authors concluded: 'both student school participation and teacher school attendance increased in programme schools, test score gains remain large in the year following the competition, and there is no increase in the frequency of test preparation sessions' (Kremer et al., 2004, pp.2-3). Follow-up structured interviews with teachers provided clues that parental support may have had a role to play in both student and teacher outcomes. Researchers highlighted the comments of two teachers, one who noted that parents asked 'teachers to work hard so that [their daughters] can win more scholarships' and another who noted that 'parents visited the school more frequently to check up on teachers, and to 'encourage the pupils to put in more efforts' (Kremer et al., 2004, p.15).

HIERARCHY OF IMPACT: PROVIDING FEEDBACK/CONSEQUENCES

The table of impact hierarchies that we are able to construct for 'providing feedback/consequences through information' based on the evidence synthesised is limited by the lack of articulated intermediate outcomes in the papers. The programme area of incentives as consequences, however, has adequate levels of detail to complete the three orders of impact in our table.

Table 4.4: Hierarchy of impact: High-stakes examinations: Providing feedback/consequences

Activity	First-order	Second-order	Third-order
High-stakes examination: Providing feedback/consequences through information			
Information provision through formats tailored to stakeholder group **(Ferrer, 2006)**	[None reported]	[None reported]	Utility for teaching, refining teaching strategy and focus Utility for students, identifying strengths and gaps
Information provision through on-site teacher and school leader training and professional development **(Kellaghan and Greaney, 2001, 2004)**	Guidance around exam preparation	[None reported]	Teaching emphasis on higher-order thinking
High-stakes assessment: Providing feedback/consequences through incentives			
School- and teacher level group incentives **(Glewwe et al., 2010)**	Increase in classroom time, class work and homework devoted to exam preparation	Schools motivated to achieve minimum pass rate to stay in the programme Teachers not motivated to earn group bonuses	No effective incentives to continuously raise standard of learning No effect on teacher attendance or teaching practice

Activity	First-order	Second-order	Third-order
		Gain in student exam results	No influence on rates of student dropout or retention Short-term gains in results were not sustained after programme ended
Teacher-level individual incentives **(Muralidharan and Sundararaman, 2011)**	Increased teaching effort focused on exam preparation (homework, class work, after-hours sessions) Increased attention to lower-performing students	No increase in teacher attendance Gain in student exam results	Individual incentive schools outperform group incentive after 2 years
Student-level merit scholarships **(Kremer et al., 2004)**	No increase in frequency of test preparation sessions	Increase in student school participation Increase in teacher attendance	Large test score gains Parental pressure for results

C-M-O CONFIGURATIONS: PROVIDING FEEDBACK/CONSEQUENCES THROUGH INFORMATION

There is insufficient evidence to identify mechanisms related to information provision alone. Review papers of assessment in Latin America (Ferrer, 2006) and sub-Saharan Africa (Kellaghan and Greaney, 2004), along with a general review of assessment (Kellaghan and Greaney, 2001) provide leverage for suggesting that the provision of tailored information alone is necessary but not sufficient to produce desirable outcomes for service delivery. The provision of on-site teacher and school-leader training is also important, which leads to the identification of **'Follow-up/Follow-through'** as a key mechanism that produces teaching emphasis on strategies that promote higher-order thinking and content mastery. The lack of evidence around the impact of information provision alone is somewhat surprising, given that both high- and low-stakes assessments presumably aim to influence the education system and individual schools through the information they produce and the knock-on effects of that information. For example, while many review papers advocate the need for bespoke reporting for targeted audiences of teachers and school leaders, none connect specific types of reporting with particular service delivery outcomes.

Table 4.5: C-M-O high-stakes examinations: Providing feedback/consequences through information

Conditions	Mechanisms	Outcomes
Information provision through formats tailored to stakeholder group	[Additional evidence needed]	Use of exam results as tools for improving teaching strategy and emphasising higher-order thinking
Information provision through on-site teacher and school leader training and professional development Guidance around exam preparation	Follow-up/Follow-through	Teaching emphasis on strategies that promote higher-order thinking and content mastery

C-M-O CONFIGURATIONS: PROVIDING FEEDBACK/CONSEQUENCES THROUGH INCENTIVES

As noted above, the base of evidence for proposing configurations around the provision of feedback/consequences is especially robust for programmes that rely on incentives for performance results. We identify three distinct configurations, one producing the undesirable outcome of minimal effort in connection with group incentives, another emphasising individual ambition in relation to desirable outcomes for individual teacher incentives, and the third focusing on parental involvement as a mechanism for increasing student and teacher engagement. The first configuration depends on conditions internal and external to the school that emphasise school performance combined with group-level incentives awarded to groups of teachers within the school as well as the entire school. These conditions appear to trigger a compliance mindset among teachers and school leaders in which the goal becomes the achievement of the minimum

threshold to earn the incentive but with no additional effort made to continuously improve performance over time. Individual teacher incentives, on the other hand did lead to sustained improvement, suggesting that similar conditions around pressure for results triggered the desire for individual reward, which led to sustained improvement and increased focus on lower-performing students, along with a corresponding increase in test preparation sessions. Finally, awards for individual students yielded an array of desirable outcomes, which researchers suggest were produced at least in part by parental involvement (Kremer et al., 2004).

Table 4.6: C-M-O high-stakes examinations: Providing feedback/consequences through incentives

Conditions	Mechanisms	Outcomes
Group, not individual, incentive School leadership pressure for results External pressure for results	Compliance with bureaucratic authority	Minimal effort – meeting minimum requirements to earn incentive
Teacher-level individual incentives School leadership pressure for results Teacher recognition of value External pressure for results	Individual desire for reward	Increased teaching effort on exam preparation Increased attention to lower-performing students Sustained increase in test results
Student-level - individual merit award for girls School leadership pressure for results Teacher acceptance Parental recognition of value and pressure for results	Parental ability to exert pressure to improve child's performance	Increase in student school participation Increase in teacher attendance Sustained gains in test scores

HIGH-STAKES EXAMINATIONS: CAPACITY DEVELOPMENT OF EDUCATORS THROUGH SCHOOL-BASED PERFORMANCE ASSESSMENT

School-based assessments (SBAs) are summative evaluations of student performance on tasks that are modelled after formative or diagnostic classroom-based assessments. SBAs intend to extract one snapshot of performance for external evaluation from a continuous stream of formative classroom-based assessment. The

SBA is meant to serve as a 'dipstick' into a process that aims to provide teachers and students with the means of continuously monitoring performance and making mid-course corrections to improve that performance.

SBAs have increasingly been introduced as a component of public examinations with the intention of redressing some of the concerns raised about the potential of high-stakes assessment for unintended consequences (Kellaghan and Greaney, 2001). Ten papers (seven of medium and three of low rigour) are concerned with the implementation of school-based assessments (Bansilal, 2011; Beets and van Louw, 2011; Kapambwe, 2010; Kellaghan and Greaney, 2004; Lubisi and Murphy, 2002; Nsibande and Modiba, 2012; Ong, 2010; Pryor and Lubisi, 2002; Reyneke et al., 2010; Scherman et al., 2011). The preponderance of these concern the implementation of SBAs in South Africa (Bansilal, 2011; Beets and van Louw, 2011; Lubisi and Murphy, 2002; Pryor and Lubisi, 2002; Reyneke et al., 2010; Scherman et al., 2011); other countries include Swaziland (Nsibande and Modiba, 2012), Zambia (Kapambwe, 2010) and Malaysia (Ong, 2010). All but one (Kapambwe, 2010) report on the lack of outcomes or unintended consequences, ranging from increased teacher workload to teachers' alienation from the interactive pedagogies that SBA aims to cultivate (see Box 4 for an SBA case example in South Africa).

The medium-rigour study in Zambia (Kapambwe, 2010) suggested that SBA has potential in achieving its intended purposes and having desirable effects on teachers and students. The study did not report the impact on instructional practice but the author claims that SBA probably has a positive influence of continuous assessment on teachers' practices in the classroom based on the study finding of pre-post difference in student scores that were significantly higher for students in the pilot schools. However, this claim does not appear to be supported by student results alone without corresponding insight into teaching practices.

HIERARCHY OF IMPACT

Capacity development of educators is the intended aim of school-based assessment, an area that includes the 10 papers in our synthesis of assessment. The base of evidence for our discussion of hierarchies of impact draws on a mix of papers of medium and low quality. As we discuss in the preceding section, all but one of the papers (Kapambwe, 2010; medium rigour) report on undesirable outcomes, largely attributing these to the lack of resources and adequate teacher preparation, the press for results, and an overarching emphasis on bureaucratic 'answerability'. The papers that report undesirable consequences of SBA are detailed in their depiction of orders of impact for teachers, particularly Bansilal (2011), and we draw on these in our articulation of hierarchy of impact in Box 4.2.

Table 4.7: Hierarchy of impact: High-stakes examinations: Capacity development of

Box 4.**2:** Case example: School-based assessment in South Africa

All five studies of SBAs in South Africa report negative impact (Bansilal, 2011; Beets and van Louw, 2011; Lubisi and Murphy, 2002; Pryor and Lubisi, 2002; Reyneke et al., 2010; Scherman et al., 2011), Reyneke et al. (2010), citing the work of Van der Berg and Shepherd (2008, p.30) reports that the qualifications authority in South Africa, Umalusi, found that: 'inaccurate continuous assessments were sending the wrong signals to learners and parents … resulting in a large number of under-prepared students entering the matriculation examination' (p.279). Reyneke et al. comment, 'Only 62.5% of candidates prepared for the final external examination through a process of continuous SBA passed. In 2009 the pass rate dropped to a new low of 60.7%, which begs the question: why is SBA not serving its purpose of enhancing learning and preparing candidates for the high stakes external examinations?'

Some believed that there was 'Far too much focus on daily (and) weekly assessment if one implemented it, there would hardly (be) any time for teaching', that assessment for learning 'is designed in the way that it caters for classes with reasonable learners - not black schools' and that to get the learners involved in learning activities was nothing but 'playing around' (Reyneke et al., 2010, p.286).

Bansilal (2011) conducted an intensive single case study of a highly competent algebra teacher's attempt to implement SBA in a school in challenging circumstances in KwaZulu-Natal, South Africa. Despite mastery of the domain and of approaches to teaching that domain, she was stymied by the large size of her class, the lack of material, and in particular, her students' struggles with basic mathematical concepts. Frustrated, she abandoned the prescribed, constructivist-oriented approach in favour of direct instruction to ensure that students generated some sort of written product for external assessment, all the while aware that they would fail.

In addition, Bansilal (2011) highlighted the lack of support for teachers' efforts to implement and sustain changes in their approach to assessment. Lack of guidance and support was also emphasised in Reyneke et al. (2010) and Pryor and Lubisi (2002). Reyneke et al. noted that the board responsible for SBA, the Independent Examination Board in South Africa, ran workshops for examiners and moderators but did not offer support for the training of those who worked directly with teachers around the implementation of SBA in their classrooms. Pryor and Lubisi (2002) commented that available professional development focused on 'procedural or bureaucratic functions, such as how to fill in and calculate official mark sheets rather than helping teachers to understand the rationale behind CA [Continuous Assessment] and its formative potential' (p.674).

SBA as a bureaucratic requirement, rather than as a process integral to teaching and learning, results in an increase in demands on teacher time. Reyneke et al. (2010) reported: 'Because of extensive record keeping and monitoring of individual learners, CASS in the South African system leads to an increase in teacher workload' (p. 287). They summarised the consequence of this constellation of the impact of SBA-related activity in South Africa: 'the poor quality and standard of the tasks set by educators; the low validity of internally set assessment tasks; the unreliability of marking instruments and the discrepancies in allocation of marks; and the unbalanced weighting of the cognitive demand and difficulty of the tasks (Umalusi, 2009, p.10)' (p.278).

educators through SBA (undesirable)

Activity	First-order	Second-order	Third-order
Implementation of school-based assessment as a component of national examinations in South Africa, Swaziland, Malaysia (Bansilal, 2011; Beets and van Louw, 2011; Kapambwe, 2010; Kellaghan and Greaney, 2004; Lubisi and Murphy, 2002; Nsibande and Modiba, 2012; Ong, 2010; Pryor and Lubisi, 2002; Reyneke et al., 2010; Scherman et al., 2011)	Professional development was disorganised, or facilitated by an educator lacking adequate experience Teacher professional development focus on procedural/bureaucratic aspects	Teachers' lack of understanding of underlying principles Teacher 'parroting' assessment tasks from guidance/previous examinations Practical difficulties (e.g., lack of resources, class-size) insurmountable Emphasis on summative, not formative, aspects of assessment Decreased time for other teaching; increased workload due to record keeping	Teacher perception of SBA as a technical procedure (compliance) unrelated to professional judgement Teacher alienation from interactive pedagogies Failure or limited implementation

C-M-O CONFIGURATIONS: HIGH-STAKES EXAMINATION: CAPACITY DEVELOPMENT OF EDUCATORS

The hierarchy of outcomes above leads us to identify compliance with bureaucratic authority as a key mechanism probably producing failure or limited implementation. A second suggested configuration is the lack of follow-up/follow-through and the failure to develop collective capacity for interpreting examination results as a result of conditions under which support is missing or misguided.

Table 4.8: C-M-O high-stakes examinations: Capacity development of educators

Conditions	Mechanisms	Outcomes
Inadequate teacher preparation	Compliance with bureaucratic authority	Failure or limited implementation Teacher alienation from interactive pedagogies

Conditions	Mechanisms	Outcomes
Teacher professional development focus on procedural/bureaucratic aspects

Emphasis on summative, not formative, aspects of assessment

Internal and external pressure for results | | Teacher lack of competence in translating lesson objectives into assessment tasks

Teacher 'parroting' assessment tasks from guidance/previous examinations

Teachers' perception of SBA as a technical procedure (compliance) unrelated to professional judgement |
| Lack of provincial support

Professional development disorganised and/or facilitated by an educator lacking adequate experience

Context and framing of task predetermined by national design and not appropriate to the level of student understanding

Internal and external pressure for results | Lack of follow-up/follow-through | Failure or limited implementation

Teachers' alienation from interactive pedagogies

Interactive pedagogies and assessment for learning are perceived as inappropriate or irrelevant |

LOW-STAKES ASSESSMENT: SETTING EXPECTATIONS THROUGH THE ESTABLISHMENT OF CURRICULUM STANDARDS

In contrast to the visible chain of influence that studies associate with high-stakes examinations, the evidence synthesised suggests that there is no clear indication of low-stakes assessments influencing policy making, managerial decisions or education practice. As we mentioned at the beginning of this chapter, the focus on quality and the trend towards decentralisation have received broad bilateral and multilateral support, which has, according to authors of the papers included in our review, emphasised the importance of comparative international and national systems of assessment as a means of monitoring system progress towards educational achievement and quality (Kellaghan and Greaney, 2001). Internationally, this has meant that an increasing number of countries (especially in Africa) are undertaking national assessments and participating in regional and international comparative studies in education (Howie, 2012). Our synthesis of four medium-rigour papers and one low rigour paper (Ferrer, 2006; Gvirtz, 2002; Kellaghan and Greaney, 2001, 2004; Ravela, 2002) highlights two of the programme pathways in which papers suggested that change had occurred: (1) setting expectations through the establishment of curriculum standards and (2) capacity development of educators through guidance documents and support to school leaders and teachers (see Section 4.4.5).

Two medium-rigour papers argue that the implementation of national assessments in Latin America may lead to the adoption of clear curricular targets which serve as de facto curriculum standards (Ferrer, 2006;

Kellaghan and Greaney, 2001). National assessments in Brazil are claimed to have led to changes to curricula along with innovations in pedagogy, as well as improved policies related to schools in challenging circumstances (Kellaghan and Greaney, 2004). Ferrer (2006) mentions Colombia, Ecuador and Uruguay as illustrating the ways in which the design of national assessments at least raises the question of curriculum standards and at best advances specific answers to the question of standards. Ferrer comments:

> Many countries in the region have professional assessment staff dedicated to establishing or specifying clear curricular targets that allow them to design more focused tests, while they continue to work (and sometimes further the debate) on what concrete outcomes of learning are expected as a priority from the students. Given the lack of concerted efforts to develop content and academic performance standards, the assessment agencies' endeavours are a significant step in the right direction. (p. 21)

Ferrer provides an example from Uruguay of how the process of validating national assessment questions among teachers led to a broader debate about the established curriculum and its lack of explicit emphasis on the kinds of higher-order thinking skills that assessment questions sought to measure. In Colombia, Ferrer notes that both national assessments and public examinations are designed with criteria that take national curriculum standards into account. The reference matrix for the assessments provides 'an explicit conceptual framework and operational definitions for each of the assessed skills' in such a way that 'different levels of student performance can be reported and illustrated more clearly than would be possible using the national curricular guidelines' (Ferrer, 2006, p.22). Ferrer highlights the potential contribution to coherence across the system:

> Both technically and politically, the basic skills that have been proposed can be regarded as curricular standards since they offer a means of effective communication among all the actors in the sector as to what students in the system are expected to learn. (p.22)

In sub-Saharan Africa, Kellaghan and Greaney (2004) provide examples in which assessments effectively identified areas of curricular need. In Namibia, results from the National Learner Baseline Assessment indicated that the expectations around competence in English were too high, which led to recommendations to revise the curricula. In Eritrea, the national assessment was able to pinpoint specific targets of difficulty in the curriculum (e.g., place value, word problem in mathematics), highlight the underperformance of girls relative to boys and identify specific areas of focus for teacher training and pedagogy.

A variant of this is the use of an assessment to expand the emphasis of schools and teachers on the development of non-cognitive areas of learning. Kellaghan and Greaney (2001) cite Pravalpruk (1996) to characterise how 'measures of affective outcomes, practical skills, and social perception were included in a national assessment in an effort to dislodge teachers' preoccupation with cognitive development in specific content areas (a preoccupation that was reinforced by end-of-school examinations for university entrance)' which 'led teachers to place greater emphasis on these outcomes in their teaching and assessments' (p.79).

However, unintended consequences can also occur that result in standards becoming established unintentionally. Ferrer's (2006) review of assessment in Latin America points to a lack of co-ordination and coherence among different functional units in the education system as a crucial contextual feature that constrains the impact of high-stakes assessment at all levels of the system. For example, Ferrer (2006) comments that 'Argentina's education assessment system includes a lack of communication between the assessment unit and the offices responsible for curricular development and teacher training, and resistance to quantitative assessment on the part of some academics and technical specialists within the ministry' (p.58). Many Latin American countries, according to Ferrer (2006), have a history of dedicated assessment units that are at arms length from ministries of education and have relative autonomy along with high degrees of

technical expertise. The authors claim that their autonomy bears the risk that 'they might become disconnected from ministry information needs and turn into programmes that, while of a high technical quality, have little impact on policy decisions geared to improving educational quality' (p.18).

In addition, Gvirtz (2002) notes that the move to standardise a curricular framework while decentralising curriculum development in Argentina inadvertently led to national assessment driving curriculum definition. The federal-level promulgated 'Common Basic Contents' (CBCs) that defined the focal areas for provincial curriculum development and also served as the basis for the development of a national assessment. Textbook publishers seized upon delays in the design of provincial-level curricula, developing textbooks that elaborated the CBCs and focused on areas targeted in the national assessment. The result, according to Gvirtz (2002, p.465), is that the 'CBCs were adopted as the new school curriculum' even though they were intended to serve only as guidelines.

HIERARCHY OF IMPACT

The evidence base for establishing service-delivery orders of impact from the synthesis in the area of low-stakes assessment is weak due to the predominant focus in the papers on the system level. There is very little detailed discussion of concrete ways in which schools have responded to the expectations put forward by low-stakes assessments, although all authors claim that service-delivery outcomes exist.

Table 4.9: Hierarchy of impact: Low-stakes assessment: Setting expectations through the establishment of curriculum standards

Activity	First-order	Second-order	Third-order
National assessment implementation: inaction on results (Ferrer, 2006)	Lack of co-ordination between assessment unit and education system	Lack of acceptance of results	Undesirable impact on schools and educational improvement
National assessment implementation: action on results (Ferrer, 2006; Gvirtz, 2002; Kellaghan and Greaney, 2001, 2004)	Consensus around the form and process of assessment	Acceptance of the results as valid and reliable system indicators	Influence teacher training, curricular change, pedagogical innovation, standards

C-M-O CONFIGURATIONS: LOW-STAKES ASSESSMENT: SETTING EXPECTATIONS THROUGH THE ESTABLISHMENT OF CURRICULUM STANDARDS

Given the paucity of evidence on service-delivery outcomes, it is difficult to identify promising mechanisms. Below we propose 'trust in pedagogical authority' as a key mechanism in order to translate consensus into influence.

Table 4.10: C-M-O low stakes assessment: Setting expectations: Establishment of curriculum standards

Conditions	Mechanisms	Outcomes
Consensus around the form and process of assessment Low-stakes reporting	Trust in the pedagogical authority of assessment approaches	Influence teacher training, curricular change, pedagogical innovation, standards
Lack of co-ordination between assessment unit and education system Low-stakes reporting	[Additional evidence needed]	Undesirable impact on the education system

LOW-STAKES ASSESSMENT: CAPACITY DEVELOPMENT FOR EDUCATORS THROUGH GUIDANCE AND SUPPORT

Two medium rigour papers discuss the guidance and support for schools and teachers around the topics and processes of national assessments as a potential pathway through which national assessments can have impact (Ferrer, 2006; Ravela 2002). Ferrer (2006) describes in detail the type of guidance and support provided for each country covered in his review. Ravela (2002) offers suggestions of potential impact in Argentina. Neither paper reports on the impact that the guidance and support actually had on the development of the capacity of school leaders and teachers to interpret assessment results effectively.

Units responsible for national assessments in Latin America have disseminated results in ways that allow schools to compare their performance with similar institutions (Ferrer, 2006, p.73). Ravela (2002) comments on efforts in Argentina to disseminate 'methodological notebooks', published regularly since 1993:

> These notebooks, which to some extent have served as models for other countries in the region, seek to foster didactic and disciplinary reflection among teachers on the basis of concrete examples of items and results. Emphasis is placed on those activities with a higher level of achievement and those that proved most difficult. (Ravela, 2002, Section IV.1, para. 1)

Several countries have developed more sophisticated methods of dissemination, including the following, as reported by Ferrer (2006):

> In the Dominican Republic, the institutional reports given to the schools disaggregate results by course or section and by knowledge area. They note the percentages of students who move forward a grade and who are kept behind, and the correlation between final school score (internal) and the score in the standardized test; they also provide a comparison of results with similar schools, as well as with all schools in the country, region, and district. (p.73)

In Uruguay, data disaggregated by school are given confidentially to each establishment; they present student results and those of students in schools in similar socioeconomic circumstances. (pp. 34, 73)

Bolivia's System for Measuring and Evaluating the Quality of Education (SIMECAL), for both its census-based and sample-based tests, gives the results to schools in the form of an institutional report on their students' performance, one that includes the average institutional score, an operational description of achievement levels by area, and the percentage of students in the school at each level. This information is followed by a description of the strengths and weaknesses of the entire student population by core topics in each area of the curriculum. (p.73)

Aguascalientes, Mexico: The improvement projects the schools are asked to implement call for the use of the available statistics and analysis of the in-school and out-of-school factors that have been shown to affect performance. Emphasis is placed on the in-school factors, since it is here that schools can have a direct effect by devising new teaching and curricular strategies. (p.73)

Sample-based national assessments cannot provide information about all schools. In an effort to ensure that all schools benefit from guidance, countries such as Uruguay provide material that enables schools not included in the sample to gauge their own performance using a selection of test items and grading guidelines. Ferrer (2006) concludes, 'Schools that were not part of the national sample can thus secure a more objective measure of their students' performance level, gain access to new assessment methods, and obtain an opportunity for more systematic reflection on the curriculum and on teaching-learning processes' (p.34).

HIERARCHY OF IMPACT

The evidence of service delivery outcomes in this area is not well articulated. The three medium-rigour papers are country overviews that do not detail service delivery outcomes.

Table 4.11: Hierarchy of impact: Low-stakes assessment: Capacity development of educators through guidance and support

Activity	First-order	Second-order	Third-order
Detailed analyses of student responses (Ravela, 2002)	Identification of variation in proficiency in different content areas	Concrete illustrations of mastery for teachers	[No outcomes reported]

C-M-O CONFIGURATIONS: LOW-STAKES ASSESSMENT: CAPACITY DEVELOPMENT FOR EDUCATORS THROUGH GUIDANCE AND SUPPORT

In terms of low-stakes assessment, there was frequent mention of the need for capacity development of educators in relation to interpretation of detailed analyses of student reports; however, there were no specific initiatives or reported outcomes connected with this suggestion.

Table 4.12: C-M-O low stakes assessment: Capacity development for educators through guidance and support

Conditions	Mechanisms	Outcomes
Detailed analyses of student responses	[additional evidence needed]	[No service delivery outcomes reported]

4.5 CONCLUSION

The following is a brief summary of the evidence of outcomes, categorised into the three generic programme pathways we identified in the synthesis. For each outcome, we also list corresponding conditions and the proposed mechanisms that may produce the outcome.

HIGH-STAKES EXAMINATIONS

A. SETTING EXPECTATIONS: SCHOOL EMPHASIS ON EXAM PERFORMANCE

Outcomes (undesirable): Evidence drawn from all 13 papers included in this area (9 papers of medium rigour and 4 papers of low rigour) identifies unintended consequences of high-stakes examinations that affect schools, teachers and students negatively.

Conditions: No internal or external support for teaching, instructional leadership, interpreting results; internal and external pressure for results.

Proposed mechanism: Fear of bureaucratic authority.

Outcomes (desirable): Three papers of medium rigour (Kellaghan and Greaney, 2004; Castro and Tiezzi, 2003; Ferrer, 2006) describe examinations achieving intended outcomes to improve the quality of teaching and learning and align what is taught with what is assessed.

Conditions: Internal and external support for teaching, instructional leadership, interpreting results; appreciation for productive persistence; internal and external pressure for results.

Proposed mechanism: Trust in the pedagogical authority of assessment approaches.

B. PROVIDING FEEDBACK/CONSEQUENCES

1. FEEDBACK THROUGH EXAM RESULTS

Outcomes (desirable): Four medium-quality papers propose the possibility of teaching improvements as a consequence of customised guidance, but we found no evidence about impact of the dissemination of results on service delivery or student learning.

Conditions: Information provision through on-site teacher and school leader training and professional development; guidance around exam preparation

Proposed mechanisms: Follow-up/follow-through.

2. INCENTIVES AS CONSEQUENCE

Outcomes (undesirable): Two high-quality studies reporting limited impact and/or unintended consequences described school-level incentives.

Conditions: Group, not individual, incentive; internal and external pressure for results.

Proposed mechanisms: Compliance with bureaucratic authority

Outcomes (desirable): A high-quality study of individual teacher incentives in rural areas of India reports significant positive results with limited unintended consequences.

Conditions: School leadership pressure for results; teacher recognition of the value of the incentive.

Proposed mechanisms: Individual desire for reward.

Outcomes (desirable): A high-quality study of individual incentives for female students in rural areas of Kenya reports significant positive results with limited unintended consequences.

Conditions: School leadership pressure for results; teacher acceptance; parental recognition of value and pressure for results

Proposed mechanism: Parental involvement.

C. CAPACITY DEVELOPMENT OF EDUCATORS THROUGH SCHOOL-BASED PERFORMANCE ASSESSMENT

Outcomes (undesirable): Evidence suggests the possibility of the adverse consequences of school-based assessment on instructional practice.

Conditions: Inadequate teacher preparation; teachers' professional development focuses on procedural/bureaucratic aspects, and it is disorganised and/or facilitated by inexperienced educators; emphasis on summative, not formative, aspects of assessment; internal and external pressure for results.

Proposed mechanisms: Compliance with bureaucratic authority; lack of follow-up/follow through for interpreting examination results to improve teaching and learning.

LOW-STAKES ASSESSMENT

A. SETTING EXPECTATIONS: THE ESTABLISHMENT OF CURRICULUM STANDARDS

Outcomes (desirable and undesirable): Four medium-rigour papers report examples of positive changes as well as instances of unintended consequences on the establishment of curriculum standards.

Conditions (desirable): Consensus around the form and process of assessment.

Proposed mechanism (desirable): Trust in the pedagogical authority of assessment approaches.

Conditions (undesirable): Lack of co-ordination between the assessment unit and the education system.

Proposed mechanism (undesirable): [additional evidence needed].

B. CAPACITY DEVELOPMENT OF EDUCATORS THROUGH GUIDANCE AND SUPPORT FOR SCHOOLS AND TEACHERS

Outcomes (desirable): Two medium rigour papers indicates the potential impact of low-stakes assessment through the provision of information and guidance to teachers and school managers. [No service delivery outcomes reported.]

Conditions: Detailed analyses of student responses.

Proposed mechanisms: [additional evidence needed].

5. IN-DEPTH REVIEW: MONITORING

5.1 INTRODUCTION

This chapter explores the conditions under which monitoring activity leads to improvement in schools and to positive learning outcomes for schoolchildren in low- and middle-income countries (LMICS), especially the poorest and most marginalised. Twenty-two papers met the criteria we established for in-depth review related to monitoring. Data were extracted from these papers, and evidence assessed for relevance and rigour as described in Chapter 2. The synthesis of findings presented in this chapter was generated from papers discussing monitoring activity primarily in Sub-Saharan Africa (n=10). An additional five papers covered multiple regions, and seven were country case studies in Latin America (n=2); East Asia (n=4) and South Asia (n=1).

As we noted in the preceding chapter on assessment, monitoring may be used in the literature to refer to all types of accountability activity that have as their purpose the evaluation of the system of schooling, including inspection and assessment, as well as collecting, disseminating and using performance information. It is this last set of activities around performance information that constitute our use of the term monitoring; as we noted in the introductory chapter, it encompasses the infrastructure and methods used to track school-level information collected primarily through quantitative/empirical methods. Monitoring refers specifically to the system-level processes designed to collect, compare and report school-level information about the composition, organisation and function of schools. This necessarily encompasses the ways that results of standardised assessments may be used, which we discussed in the preceding chapter. In this chapter, we look at the collection and use of the wider set of school-level performance information that might be collected and used in a system of monitoring. In addition to the presentation of test scores, reported information may include comparisons of school performance with other schools; students' socio-economic characteristics; the results from surveys of student and parent satisfaction with various school features; school financing and/or audit findings; school-level inputs and expenditures (Bruns et al., 2011). An example from an initiative to publicly monitor individual schools through the publication of a 'school report card' used in the Brazilian state of Parana provides an illustration of the range of information that this might include (See Box 5.1).

We begin our synthesis with an elaboration of our definition of monitoring and also review the quality of the studies. We then present our findings around the most prominent programme pathways, identifying initiatives that share similar features in terms of the generic pathways, as identified in our initial rough theory, through which monitoring programmes are intended to yield particular outcomes. This allows us to identify reported outcomes and then elaborate hierarchies of outcomes in order to develop the chain of evidence to support our inferences around likely C-M-O configurations.

> **Box 5.1:** State of Parana, Brazil, school report card
>
> The report card followed a standard format for the report on each school. In addition to aggregate results of students' test-based performance, information included student flows (promotion, retention and drop-out rates), school characteristics (average class size and teachers' qualifications), results from parental surveys (satisfaction about facilities, security, teaching practices, quality of education and parental involvement), and parent opinions on the availability of information about school performance and activities. Most of the items also included comparative municipal and state averages to help those using the report to understand how a particular school was positioned in terms of neighbouring schools (Bruns et al., 2011, p. 60); (see also Crouch and Winkler, 2008, p. 26, Text Box 3).

The chapter has the following sections:

Section 5.2: Elaborating the definition of monitoring

Section 5.3: Quality of studies

Section 5.4: Findings about types of activities

Section 5.5: Conclusion

5.2 ELABORATING THE DEFINITION OF MONITORING

Our synthesis led us to an elaboration of our initial definition of monitoring in LMICs that we want to highlight. The shift is to include local collection and reporting of data that is driven by a national or provincial agenda around system-wide monitoring. This shift allows us to include information generated at the local level by local stakeholders, but within a national or provincial framework of information geared towards system-wide monitoring and comparability across schools. The crucial distinction here is that the monitoring has implications for the system more broadly and is not simply a means of ensuring local school accountability. Six papers in our review (ADEA, 2001; Barr et al., 2012; Brock, 2009; Bruns et al., 2011; Crouch and Winkler, 2008; De Grauwe, 2007) specifically identify the shift towards an increased role for local actors in system-wide monitoring – including teachers, school leaders, parents and community-members – with the broader trends of decentralisation and school-based management.

5.3 QUALITY OF THE STUDIES

We assessed the quality of the 22 papers included in the monitoring synthesis, using the method described in Chapter 2. The papers included five working papers for the ADEA, USAID and the World Bank; four literature reviews; eight small-scale country or provincial case studies; and five quantitative measurements of the impact of interventions related to monitoring activity. The majority of these papers include limited descriptions of the underlying methodology and present self-reports of small (potentially non-representative) samples of respondents. Further information about the quality of all the included studies is presented in Table 5.1 and in Appendix 4.1.

Table 5.1: Reviewers' judgements about rigour and relevance of each study included in the monitoring synthesis

Studies (first author and date)	Rigour			Relevance		
	High	Medium	Low	High	Medium	Low
ADEA (2001)			✓		✓	
Andrabi (2013)	✓				✓	
Attfield (2013)		✓		✓		
Barr (2012)	✓			✓		
Brock (2009)		✓			✓	
Bruns (2011)		✓		✓		
Caddell (2005)		✓			✓	
Chen (2011)		✓			✓	
Crouch (2008)			✓	✓		
De Grauwe (2007)			✓		✓	
Gvirtz (2004)		✓			✓	
Higgins (2005)		✓			✓	
Lassibille (2010)	✓			✓		
Murimba (2005)			✓		✓	
Powell (2006)		✓			✓	
Prew (2010)			✓		✓	
USAID (2006)		✓		✓		
USAID (2007)		✓		✓		
Winkler (2005)			✓		✓	
Winkler and Herstein (2005)			✓			
World Bank (2008)		✓			✓	
World Bank (2010)			✓		✓	

5.4 FINDINGS ABOUT TYPES OF ACTIVITIES

Our primary focus in this section is the reported and suggested outcomes for schools as a consequence of different types of monitoring activity. In the papers we synthesise, 13 papers (one of high rigour, seven medium and five) discuss various aspects of EMIS, including school development planning, from regional and country-level perspectives (ADEA, 2001; Attfield and Vu, 2013; Brock, 2009; Bruns et al., 2011; Caddell, 2005; Crouch and Winkler, 2008; Lassibille et al., 2010; Murimba, 2005; Powell, 2006; Prew and Quaigrain, 2010; USAID, 2006; Winkler and Herstein, 2005; World Bank, 2010). Most papers emphasise the impact of EMIS implementation on system efficiency, especially the allocation of resources; implications for school-level service delivery is not consistently noted. Impact on student learning outcomes was only reported in one high-rigour study and was not significant (Lassibille et al., 2010). School monitoring report cards are the central

topic of five (one low-, two medium-, and two high-rigour) papers (Andrabi et al., 2013; Barr et al., 2012; Bruns et al., 2011; USAID, 2006; Winkler and Herstein, 2005). Andrabi et al. (2013) and Barr et al. (2012) involve experimental research. We highlight three general areas of monitoring activity, two of which relate to the programme pathway of 'providing feedback': uses of information provided from **Educational Management and Information Systems (EMIS)** to local education leaders for school-level management decisions (Section 5.4.1) and **school monitoring report cards** that are typically produced centrally and provided to a wider group f school stakeholders, not only educators but also parents and the wider community (Section 5.4.2). The third programme pathway we highlight is 'setting expectations' through generating demand for EMIS information for **local school development planning** (Section 5.4.3).

PROVIDING FEEDBACK: EMIS FOR SCHOOL-LEVEL MANAGEMENT DECISIONS

Monitoring activity that aims to provide feedback for decision making that has an impact on schools emphasises the design, implementation and uses of EMIS. The intended programme pathway of EMIS information is to provide education decision makers and policy analysts with information that allows them to understand how educational inputs relate to educational outputs. EMIS aims to improve decision making and the targeting of resources to areas most in need through access to high-quality and timely data. The papers synthesised in this chapter include several descriptive accounts and overviews of the implementation of EMIS and its impact on service delivery in schools (1 high rigour - Lassibille et al., 2010; 4 medium rigour – Chen, 2011; Higgins and Rwanyange (2005); Gvirtz and Larripa (2004); World Bank (2008); 4 low rigour - ADEA, 2001; De Grauwe (2007); Murimba, 2005; World Bank, 2010). One experimental study (Lassibille et al., 2010), which we discuss first, does examine the impact of EMIS at the level of the school. The findings from this study suggest that information alone had little effect but the provision of information with training and support at the local level had a sustained effect on management and student attendance.

The high-rigour RCT in Madagascar included a set of tools that were distributed to 909 randomly selected schools, with 303 randomly selected control schools (Lassibille et al., 2010). From 2005 to 2007, the control (no intervention) and four different treatments ran in parallel. The complete set of interventions consisted of school leaders and officers at the sub-district and district levels receiving summary reports that corresponded with their management level, along with management toolkits and guides as well as training. One of the four treatment groups received all interventions at all levels and the three other groups received constrained variations, either by level (e.g., local only, no sub-district or district support) or type of intervention (e.g., provision of guides only, no training). Baseline and follow-up data included data on the implementation of interventions along with test scores from standardised tests in three subjects. The programme showed significant impacts on manager, teacher and student behaviours, particularly for the treatment group that received all interventions at all levels:

> 'In its most direct and intensive form, the interventions changed the behaviour of all actors toward better management. These changes translated immediately into increases in student attendance and sizable reductions in dropout rates'. (Lassibille et al., 2010, p. 20)

After two years, more than one in three (37%) treatment schools versus fewer than one in six (15% control schools were considered relatively well-managed (i.e., teachers and the director perform essential responsibilities). However, the impact on student learning was small and not statistically significant. The study authors conclude: 'changing service providers' behaviour takes time and effort, and a two-year time frame was probably too short to produce clear-cut impacts on student test scores' (Lassibille et al., 2010, p. 20).

The Madagascar study was one of a series of pilot initiatives developed by AGEPA (Amélioration de la Gestion dans les Pays Africains), a regional programme that eventually included Madagascar as well as Benin, Burkina Faso, Guinea, Mali, Mauritania, Mozambique, Niger, Senegal and Togo. We reviewed a summary report of the project that was a descriptive account of project activities included in an overview of school accountability policies in sub-Saharan Africa (World Bank, 2010). The World Bank report included a more recent and more thorough reporting of activities in sub-Saharan Africa than did ADEA (2001), an earlier regional overview. The AGEPA project provided what the report described as 'high-quality, country-tailored technical support' to help countries 'define and conduct analytical work to diagnose management and accountability gaps', and from this analysis to develop practical interventions at the local, school, and classroom levels to address gaps (World Bank, 2010, p. 7). Aside from Madagascar, there were no other rigorous evaluations of impact conducted. The World Bank summary report suggests anecdotally that 'Country demand for technical assistance in the area of education and school management has been continuously growing' which it attributed in part to the success of the project (p. 3).

The SACMEQ (Southern Africa Consortium for Monitoring Educational Quality) initiative also had a substantial component of improving country infrastructure for EMIS. Murimba (low rigour) (2005) does not mention any school-level impacts in a descriptive account of the initiative, but the authors do suggest that the project improved the systems and processes for collecting and disseminating performance information, as well as building technical and strategic capacity for EMIS (p. 2).

HIERARCHY OF IMPACT

De Grauwe (2007) provides a panoramic view of monitoring and the implementation of Education for All in 45 low-income countries based on analyses of national policy documents. Murimba (2005) and the World Bank reports on sub-Saharan Africa (2008, 2010) offer high-level overviews of large-scale EMIS initiatives in Africa; Higgins and Rwanyange (2005) focus on the introduction of monitoring in Uganda. Chen (2011) offers a similar account of EMIS for school-level decision making in Indonesia, while Gvirtz and Larripa (2004) provide a high-level view of education monitoring in Argentina with scant evidence about school-level outcomes of monitoring activity. Lassibille et al. (2010) is the one study that details service delivery outcomes at the school level. We have explored the dependencies among intermediate outcomes for the treatment group that received all interventions at all levels and showed significant impacts on manager, teacher and student behaviours. We have used actual findings of impact from this study (third order) to infer a hierarchy of intermediate outcomes, beginning with programme activities in the first column and proceeding through: first-order impact, which corresponds with direct outcomes from implementation activities; second-order impact in the next column related to the reported consequences of implementation; and third-order impact, which are the outcomes Lassibille et al. (2010) reports from the intervention. These are the intermediate outcomes that we expect to contribute to our ultimate outcomes of interest, improved service delivery, through improvements in classroom and school management, and increased system efficiency through improvements in district and sub-district supervision and monitoring. (No significant outcomes were reported for student learning, although the authors suggest that this two-year study may not have been long enough to reveal these effects.)

Table 5.2: Hierarchy of impact: Providing feedback: EMIS for school-level management decisions

Activity	First-order	Second-order	Third-order
School level EMIS information, along with comparative performance data, accompanied by role definition, operational guidance, specific tools and intensive training Creation of tri-level (district, sub-district, school) operational tools and standardised processes for each role (e.g., teachers, school directors, sub-district and district administrators) in an education system focusing on pedagogy, student learning and follow-up, management of instructional time, administration, school statistics, partnership with local community Guidance delivered to those in each role Customised training modules delivered for those in each role emphasising how and when each tool is used (Lassibille et al., 2010)	Teacher and school-leader workflow processes performed as implemented Sub-district- and district-level administrators perform corresponding workflow processes	School-level processes reinforced at sub-district and district levels through supervision and monitoring	Classroom management improves in monitoring and following up on student absenteeism, preparing lesson plans and tracking progress in student learning School management improves in carrying out supervisory and monitoring duties Sub-district and district management improves in carrying out supervisory and monitoring duties

C-M-O CONFIGURATIONS: PROVIDING FEEDBACK: EMIS FOR SCHOOL-LEVEL MANAGEMENT DECISIONS

The hierarchy of impact enables us to infer potential causal pathways from the interconnections among outcomes. The first- and second-order impacts reported above do not in themselves explain how third-order impacts occurred. However, we can infer particular sequences of activity involving those holding different roles that are important to achieving particular outcomes. These causal pathways allow us to establish precursors, or conditions, that are necessary for particular outcomes to arise and then infer possible mechanisms triggered by those conditions that yielded the outcomes reported or suggested.

The high-rigour study of Lassibille et al., 2010, suggests that the mechanism that we call Follow-up/Follow-through explains why schools that received not only information from EMIS but also received guidance and training at the district, sub-district and school levels saw the greatest improvement in school and instructional management processes. The guidance and training across levels, according to the study authors, improved classroom management through the improved oversight of school and district leadership. Knowing how to follow up and when to follow through depended on several conditions, such as the clarity of key workflow processes across levels and support in implementing key processes across levels.

Table 5.3: C-M-O providing feedback: EMIS for school-level management decisions

Conditions	Mechanisms	Outcomes
Clarity of key workflow processes within and across classroom, school, sub-district and district levels Support in implementing workflow processes within and across levels Consistent and clear feedback about implementation of workflow processes and means to improve (Lassibille et al., 2010)	Follow-up/follow-through	Classroom management improves in monitoring and following up on student absenteeism, preparing lesson plans, and tracking progress in student learning School management improves in carrying out supervisory and monitoring duties Sub-district and district management improves in carrying out supervisory and monitoring duties.

PROVIDING FEEDBACK: SCHOOL REPORT CARDS

We now turn to another category of monitoring activity that is intended to influence service delivery at the school level through providing feedback of another kind. Increasing parental and, occasionally, student roles through school report cards is an aspect of five papers in this review (two high, two medium and one low on rigour) (Andrabi et al., 2013; Barr, 2012; Bruns, et al., 2011; USAID 2006; Winkler, 2005). Three of the papers

(two high, one low rigour) concentrate on particular national or provincial implementations. Two of the country-specific papers are RCTs, one of which is an experiment in Pakistan with the dissemination of information-rich school report cards to parents (Andrabi et al., 2013) and the other of which is a 'participatory scorecard' initiative in Uganda (Barr et al., 2012). The third case-specific paper is a descriptive account of a school report card initiative in Parana State, Brazil (Winkler, 2005). The remaining two papers are of medium rigour and discuss the use of school report cards in multiple countries. These are a USAID (2006) overview of school report cards and a World Bank report (Bruns, et al., 2011) that includes school report cards as part of a non-systematic review of accountability-focused reforms in 11 developing countries.

The two medium- and low-rigour papers (Bruns, et al., 2011; USAID 2006; Winkler, 2005) suggest an increased likelihood of beneficial effects on service delivery and learning outcomes from combining the dissemination of school performance information with participatory processes that involve parents in collecting and analysing school-level information. This claim is supported by one RCT study, judged as high rigour, that finds a participatory intervention to have greater impact than one that did not include participation (Barr et al., 2012). However, both that study in Uganda and another RCT in Pakistan (Andrabi et al., 2013) found that enriched provision of information alone yielded a desirable impact on school management without participatory processes. In the latter study, the anticipation of parental action by school leaders following the public release of school report cards appears to have contributed to improvements in school leaders' oversight of instruction that led to students' increased test scores.

The aim of school report cards is to use performance information that may be generated by EMIS or other sources to amplify local actors' political voice or promote parental agency through choice (Bruns et al., 2011). We concentrate in our review on system-wide initiatives around the uses of performance information for local school accountability; we do not include efforts that concentrate exclusively on community accountability independent of wider processes (for a recent systematic review of community accountability, see Westhorp, et al., 2014).

The high-rigour study of Andrabi and others (2013) examined the impact of including school performance and fee information and comparative data about local schools' performance and fees along with children's scores on a standardised test. Grade 3 pupils in 112 villages, comprising 800 public primary schools, were given a standardised achievement test. Report cards were then distributed to a random selection of half of the villages. Researchers conducted school and household surveys along with follow-up testing of children to determine impact. The study found that: the additional provision of information positively affected children's test score gains; was accompanied by the lowering of private school fees; and influenced parents' beliefs about school quality. Test score gains for children in treatment villages was 0.11 standard deviations higher than those in control villages; fees in private schools were 20 percent lower. Household surveys confirmed that the provision of information changed parental views of school quality, such that they closely aligned with school test scores, which the authors conclude is 'consistent with information increasing the precision of the quality signal for parents' (Andrabi et al., 2013, p.4). However, survey results suggest that changes came about primarily through the organisational responses of schools and shifts in allocation of resources towards hiring better-qualified teachers, increasing the use of textbooks and increasing the length of the school day (p. 5). School report cards did not change household investments in education (e.g., parental expenditures, time spent on education). This led researchers to suggest that household investment 'may have been directed toward greater pressure on the school rather than investments at home' (Andrabi, et al., 2013, p. 5).

Barr et al. (2012) is an RCT of two variations of a school monitoring scorecard. One hundred primary schools from districts in each of Uganda's four regions participated in a control group (40) or one of two treatment

groups (30 each). In the standard treatment group, members of the school management committee (SMC), which comprised parents and community members, received training in the use of a *standardised scorecard* that measured indicators valued by the Ministry of Education. The alternative treatment included similar training in the principles of monitoring but gave SMC members freedom to design unique scorecards for their schools, or *participatory scorecards*. In treatment schools, the SMC collected termly data on all indicators. Schools in the control group did not receive any intervention. The participatory design showed statistically and economically significant effects in reducing pupil (8.9%) and teacher (13.2%) absenteeism and a desirable effect on pupil test scores, equivalent to increasing a pupil's standing from 50th to 58th percentile (Barr et al., 2012, pp. 16-17). The study authors concluded that 'the participatory design component of community-monitoring interventions may be important to their success. Delegation of this process appears to have fostered a stronger sense of ownership among school stakeholders.' The intervention did not show significant impact on student enrolment, progression or retention.

The medium-rigour World Bank publication, *Making Schools Work*, by Bruns et al. (2011), highlights school report cards in one section (pp. 42-46). None of the high-quality studies reviewed in this report are of school report cards; the authors draw conclusions about school report card initiatives based on case studies and reports of small pilot studies in Brazil (the Parana state initiative reported in Winkler, 2005), and sub-Saharan Africa, which is reviewed in the preceding section on school development planning.

A low-rigour, descriptive overview of efforts in the State of Parana, Brazil, by Winkler (2005) suggests desirable effects, with parents engaging teachers in discussions about school improvement and an increase in parental voice in the policy deliberations of district and regional school councils (also summarised in Bruns, et al., 2011, p. 45). Winkler comments that, 'by giving school-level data high visibility, school and parents' councils became a small army of quality controllers, reporting discrepancies in state and national databases' (Winkler, 2005, p. 3). Bruns, et al. (2011) summarises the state of evidence in this area in the following way:

> 'This largely qualitative and anecdotal evidence suggests that information-for-accountability reforms might have positive impacts: greater collaborations and better communications between parents and teachers, improved parental participation in school matters, better and more frequent data reporting mechanisms, better resource flows, and some suggestion of improved education outcomes'. (p. 49)

We did not find any suggestion of specifically undesirable effects of school report cards; although as Bruns et al. (2011) note, the largely anecdotal and highly contextual nature of the evidence that currently exists precludes any strong claims about desirable or undesirable effects.

HIERARCHY OF IMPACT

Our exploration of hierarchies of impact focuses on the two high-rigour studies of Andrabi et al. (2013) and Barr et al. (2012) as these provide the most detailed information relevant to identifying potential causal pathways. Table 5.4 provides a sequence of intermediate outcomes reported in each study to begin to sketch dependencies among first-, second- and third-order impact. Most notably, Barr et al. (2012) suggest that 'increased sense of ownership' of school issues not only led to improvements in children's test scores but also reductions in pupil and teacher absenteeism.

Table 5.4 Hierarchy of impact: Providing feedback: School 'report cards'

Activity	First-order	Second-order	Third-order
Provision of school-level reports of centrally collected and analysed EMIS information, including comparisons of school performance, for school personnel and local stakeholders (Andrabi et al., 2013)	Parental determination of school quality based on comparative performance and cost Parental pressure on school to improve performance and lower cost	School leadership reallocates resources (e.g., hires better qualified teachers, more use of textbooks, increase school day) School fees lowered	Children's test scores improve
Local stakeholders on SMC are trained in information literacy and design of a display of EMIS information about their school (Barr et al., 2012)	Local stakeholders develop 'participatory scorecard' for their school	Local stakeholders develop capacity to use information effectively to analyse school performance	Increased sense of ownership for school quality among local stakeholders and school personnel Reductions in pupil and teacher absenteeism Children's test scores improve

C-M-O CONFIGURATIONS: PROVIDING FEEDBACK: SCHOOL REPORT CARDS

The two high-rigour studies concerning school report cards (Andrabi et al., 2013; Barr et al., 2012) suggest that the horizontal mechanisms of parental engagement explain improvements in school quality and children's learning. Andrabi et al. (2013) proposes parental ability to exert pressure to improve school performance as the mechanism that led school personnel to reallocate school resources towards instructional improvement, lower school fees and make changes that improved children's test performance. Important conditions that we infer triggered parental pressure were initial capacity to analyse information effectively to evaluate the strengths and weaknesses of the school and relate these to areas for improvement. Barr et al. (2012) attribute parents' ability to exert pressure to the involvement of local stakeholders in monitoring activities, making decisions around what information to collect and becoming involved in the processes of collecting, analysing and displaying that information. We infer from this a C-M-O configuration that yields stakeholder sense of ownership, Parental Participation in school monitoring. Parental Participation, according to Barr et al., may also explain reductions in pupil and teacher absenteeism in schools carrying out the participatory scorecard approach.

Table 5.5 C-M-O providing feedback through school report cards

Conditions	Mechanisms	Outcomes
Decisions of local stakeholder group (e.g., SMC) of consequence to school personnel	Parental ability to exert pressure to improve school performance (Parental pressure)	School leadership reallocates resources (e.g., hires better

Conditions	Mechanisms	Outcomes
Local stakeholders develop the capacity to use information effectively to understand school performance School personnel capacity to work with local stakeholders (Andrabi et al., 2013; Barr et al., 2012)		qualified teachers, more use of textbooks, increase school day) School fees lowered Children's test scores improve
Local stakeholders engaged in decisions around what information to collect and the process of collecting information Local stakeholders develop the capacity to use information effectively to understand school performance School personnel capacity to work with local stakeholders (Barr et al., 2012)	Parental participation in monitoring activity	Shared sense of ownership of the school among local stakeholders and school personnel Children's test scores improve Reductions in pupil and teacher absenteeism

The generic pathway of providing feedback is the anticipated means through which EMIS intends to shape the behaviours of school personnel and thus improve service delivery. We propose three specific mechanisms relating to providing feedback, Follow-up/Follow-through in relation to the use of EMIS for management decisions affecting the school level and Parental Pressure and Parental Participation in relation to school report cards. Follow-up/Follow-through may be considered a vertical mechanism in the sense that it relies on coherence across hierarchical levels. Parental Pressure and Parental Participation are horizontal mechanisms in that these are triggered by the engagement of local stakeholders in scrutiny of school organising processes and priorities. All three mechanisms hinge on the development of the capacity of those involved to interpret information and use it to evaluate school performance and shape decision making around how to improve performance.

SETTING EXPECTATIONS: USES OF EMIS WITH LOCAL SCHOOL DEVELOPMENT PLANNING

We now turn to the generic pathway of setting expectations as a way that EMIS intends to shape service delivery at the school level. We focus on locally-based school development planning (SDP), highlighted in nine papers (six medium- and three low-rigour papers), as a means of improving the quality of data inputs and developing demand and associated capacity to use performance information outputs effectively (Attfield and Vu, 2013; Brock, 2009; Bruns et al., 2011; Caddell, 2005; Crouch and Winkler, 2008; Powell, 2006; Prew and Quaigrain, 2010; USAID, 2006; Winkler and Herstein, 2005). Local SDP processes are presented as vehicles for setting expectations through generating demand for EMIS information, developing capacity to use that

information effectively and increasing the quality of data inputs provided to EMIS (Powell, 2006, p. 19). All the papers use descriptive accounts of existing interventions to support their claims; none is an experimental or comparative study. Suggestions of desirable impact of one effort in Nigeria appear in four papers, two of medium rigour (Bruns et al., 2011; USAID, 2006) and two of low rigour (Crouch and Winkler, 2008; Winkler and Herstein, 2005). Two medium-rigour papers describe the positive impact of SDP initiatives in Vietnam (Attfield and Vu, 2013) and in China (Brock, 2009). Two papers (one low and one medium rigour) suggest mixed impact from an initiative in Ghana (Prew and Quaigrain, 2010; Powell, 2006), and one paper of medium rigour suggests undesirable impact in a national initiative in Nepal to promote local SDP (Caddell, 2005).

Local SDP coupled with national or provincial EMIS is viewed in six papers as engaging local school leadership more directly in the processes not only of generating plans but also of determining which indicators merit measurement and being involved in the presentation and use of results. Five papers report a promising initiative implemented in Kano, Nigeria (Bruns et al., 2011; Powell, 2006; Winkler and Herstein, 2005; Crouch and Winkler, 2008). Crouch and Winkler (2008) describe:

> 'A new collection tool and a data management system were created after soliciting input from stakeholders state-wide, including central planners and managers; local education officers; PTA and teachers' union members; and members of the legislature, the governor's office, testing authorities, and the Ministry of Finance. In addition to capturing information relevant to all stakeholders, the new information system also generates multidimensional reports targeting different issues and different users, as designed by the end-users themselves'. (p. 27)

One paper of medium-rigour provides an overview of an initiative in China (Brock, 2009). The initiative coupled the strengthening of EMIS at the system level with local initiatives using SDPs to improve school management and promote community engagement with schools. The initiative, the Gansu Basic Education Project, also had an inspection component discussed in Chapter 6. Impacts attributed to the project include increases in net enrolment from 79% in 1999 to 91% in 2005, with the largest increases in primary schools in remote areas and among minority girls (Brock, 2009, p. 456). The authors suggest that desirable changes in student enrolment were attributable to improvement in local school management that stemmed from the SDP process.

> 'The effects of SDP have been very positive. It has introduced a level of real operational planning to schools and has shown County Education Bureau (CEB) staff how delegation of responsibility to schools and headteachers, while running the risk of abuse, does in most cases result in increased ownership of local issues – thereby reducing the demands on the CEB'. (Brock, 2009, p. 457)

The author notes that as of 2010, the SDP was being advanced in 10 other provinces in China (Brock, 2009, p. 457).

A descriptive account of the implementation of minimum school standards in Vietnam suggests that the effective use of EMIS in SDP hinged on lowering the standards that were used to evaluate school-level performance, especially for schools in the most challenging circumstances. The Primary Education for Disadvantaged Children project developed a set of minimum standards for school quality that were then used as the basis for an annual survey of all schools. Attfield and Vu (2013) comment that the existing national standards had promoted the inequitable allocation of resources to high-performing schools; the recalibrated standards, the authors claim:

> 'Enabled accurate, quantitative benchmarking of [minimum requirements] with a sufficiently rich data set that has been exploited in a diverse range of ways to demonstrate change, track equity of

investments, plan on both the macro and micro level and test hypotheses on the variables and factors that hide within the 'black box' of a classroom and somehow determine learning'. (p. 83)

Service-delivery outcomes were not explicitly detailed by Attfield and Vu (2013) but the reports of system-level changes in response to the shift in standards as well as the mention of impact at the school and classroom level suggests that recalibrating expectations enabled information to be used more effectively for planning at the school level.

Two papers report of an initiative to promote SDP, coupled with development of a system-wide EMIS in Ghana and suggest mixed results at the level of the school but more positive suggested outcomes at the district level (Powell, 2006; Prew and Quaigrain, 2010).

The suggested positive or mixed results of the locally-focused interventions in Vietnam, China, Nigeria and Ghana run counter to a narrative account of an initiative in Nepal. Caddell (2005) suggests that efforts to promote greater involvement of local school staff and community members alongside the development of national EMIS capability resulted in 'participatory/micro-planning exercises' that were more concerned with the 'extraction of data' than eliciting the input of school personnel and members of the local community:

> 'In practice, however, the [District Education Plan] process remained largely focused on the extraction of data to meet national planning, monitoring and evaluation objectives as opposed to districts or schools being able to set their own agendas. Rather than offering opportunities for local voices to be heard, these attempts sought to transfer central-level concerns and responsibilities, including the pressure to meet EFA-related objectives, to the district and sub- district level'. (p. 462)

The author does not appear to take into consideration that translating pressure to meet EFA-related objectives to the local level may have been a primary policy objective. Nonetheless, the paper highlights the ways in which an emphasis on technical/bureaucratic processes may lead to displacing participatory aims.

HIERARCHY OF IMPACT

The medium- and low-rigour papers that comprise this category of monitoring activity provide descriptive detail about first-, second- and third-order impacts, but none is a direct focus of research as in the areas covered earlier. Reports of SDP processes in Kanu, Nigeria, and Gansu, China, describe improvements in school management that led to more effective community engagement, which resulted in Gansu Province in improved primary school enrolment of minority girls. These projects emphasised eliciting a demand for EMIS information, whereas the minimum school standards project in Vietnam approached increasing demand for data from another direction, by recalibrating school standards to the basic needs of schools. We also trace the hierarchy of undesirable impact suggested by Caddell (2005) in his description of the implementation of 'participatory microprocesses' in schools in Nepal, which Caddell claims, led to an increase in the quality of data but no change in stakeholders 'sense of shared ownership' in school decision making.

Table 5.6 Setting expectations: Uses of EMIS with local school development planning

Activity	First-order	Second-order	Third-order
Desirable impacts			
Local School Development Planning (SDP) processes in Kano, Nigeria (Bruns et al., 2011; Crouch and Winkler, 2008; Powell, 2006; Winkler and Herstein, 2005)	School personnel are engaged in determining EMIS indicators and presentation of results Provision of customised 'multidimensional reports' that take into consideration the specific needs and uses of different stakeholders, including school personnel	EMIS information is used effectively to develop local operational plans Increased demand for EMIS information Increased quality of local information provided to EMIS	Improvements in school management and community engagement
Local school development planning (SDP) processes in Gansu Province, China (Brock, 2009)	Improvement in local school operational planning and management processes	Increased school-community engagement	Improvements in primary school enrolment among minority girls Increased ownership of local education issues by school and community

Activity	First-order	Second-order	Third-order
Implementation of revised, less demanding minimum school standards in Vietnam (Attfield and Vu, 2013)	Provision of accurate, quantitative benchmarking of minimum standards	[no outcomes reported]	EMIS information is used more effectively for planning and equitable allocation of resources at school and classroom levels
Undesirable impacts			
'Participatory/micro-planning exercises' involve school personnel and community members in the elaboration of District Education Plans in Vietnam (Caddell, 2005)	Schools conduct local planning exercises with community members	District plans include data from planning exercises to meet national planning, monitoring and evaluation objectives District plans do not reflect input from local level other than data	Improvement in the quality of information provided to EMIS Decreased sense of ownership of issues by local school and community

C-M-O CONFIGURATIONS – SETTING EXPECTATIONS: USES OF EMIS WITH LOCAL SCHOOL DEVELOPMENT PLANNING

We propose three configurations of conditions, mechanisms and outcomes that relate to the implementation of SDPs to set expectations around service delivery and learning and teaching. Two of these mechanisms yield desirable outcomes, 'Learning from Failure' and 'Reality Testing'. The third results in an undesirable impact, Lack of Follow-up/Follow-through.

Brock (2009) emphasises that the Gansu Basic Education Project (GBEP) had created an environment for experimentation that encouraged school personnel to take increased long-term responsibility for their school and, at the same time, did not hold them responsible for short-term failure. These conditions encouraged Learning from Failure within schools as well as in external relations of school to district and school to community. Brock (2009) suggests that experimentation enabled the elaboration of operational SDPs which facilitated school-community connections, most notably towards improving enrolment among minority girls.

Attfield and Vu's (2013) account of the promulgation of new minimum school standards in Vietnam suggests a mechanism that we label Reality testing, with the intention of highlighting how the focus on meeting basic needs led to a concrete and accurate portrayal of systemic disparities in schooling. The public recognition of the actual state of the education system and local schools led to more effective uses of EMIS information for planning and equitable allocation of resources at school- and classroom-levels. Reality Testing hinged on a sophisticated EMIS infrastructure that was able to collect and disseminate accurate and timely information.

Finally, setting expectations appears to be impeded by the lack of follow-up/follow-through, a mechanism that we proposed in the preceding section on providing feedback as triggered by the coherence of processes across hierarchical levels. In this instance, that of the elaboration of detailed District Educational Plans in Nepal (Caddell, 2005) through 'participatory microprocesses' at the school level, the disconnection between district and school, Caddell claims, led to local alienation rather than increased participation in taking responsibility for local school issues.

Table 5.7 C-M-O setting expectations: Uses of EMIS for local SDP

Conditions	Mechanisms	Outcomes
Supplemental funding for implementation of a new national curriculum Delegation of some planning and resource allocation responsibility to school level Training of school leadership in education management, coupled with teacher training and development of new curricular materials	Learning from failure	Improvements in primary school enrolment among minority girls Increased ownership of local education issues by school and community

Conditions	Mechanisms	Outcomes
Environment for experimentation (Brock, 2009)		
Revised minimum school standards focused on basic inputs for learning EMIS infrastructure for accurate and timely monitoring and reporting of school and district-level progress against standards (Attfield and Vu, 2013)	Reality testing	EMIS information is used more effectively for planning and equitable allocation of resources at school- and classroom levels Increase in reliability of information provided by schools
Lack of capacity at the district level to support participatory processes at the school level 'Participatory processes' are aimed at meeting the district demand for information, not local participation (Caddell, 2005)	Lack of follow-up/follow-through	Decreased ownership of local education issues by school and community

5.5 CONCLUSION

We briefly summarise the evidence of outcomes, categorised by the two generic programme pathways we identified in the synthesis. We list the related conditions and corresponding proposed mechanisms that may explain the outcomes.

PROVIDING FEEDBACK

USES OF EMIS FOR MANAGEMENT DECISIONS THAT AFFECT THE SCHOOL LEVEL

Outcomes: Evidence was drawn from one high- four medium- and four low-rigour papers. Findings from the high-rigour study suggest that information alone had little effect but the provision of information with training and support at the local level had a sustained effect on management and student attendance.

Conditions: Clarity of key workflow processes within and across classroom, school, sub-district and district levels; support in implementing workflow processes within and across levels; consistent and clear feedback about the implementation of workflow processes and means to improve.

Proposed mechanism: Follow-up/Follow-through.

SCHOOL REPORT CARDS

Outcomes (including participatory processes): Four papers (one high, two medium and one low on rigour) (Barr et al., 2012; Bruns, et al., 2011; USAID 2006; Winkler, 2005) suggest an increased likelihood of beneficial effects on service delivery and learning outcomes from combining the dissemination of school performance information with participatory processes that involve parents in collecting and analysing school-level information. The high-rigour study reports improvement in pupils' test scores and reductions in pupil and teacher absenteeism, and suggests that this resulted from a 'shared sense of ownership' around the school among local stakeholders and school personnel.

Conditions: Local stakeholders were engaged in decisions around what information to collect and the process of collecting information; local stakeholders developed the capacity to use information effectively to understand school performance; school personnel developed the capacity to work with local stakeholders.

Proposed mechanism: Parental participation in monitoring activity.

Outcomes (not including participatory processes): Two high-rigour studies (Andrabi et al., 2013; Barr et al., 2012) found that enriched provision of information alone yielded greater parental awareness of school quality, with a consequent desirable impact on school management without participatory processes.

Conditions: Decisions of local stakeholder group (e.g., the SMC) were of consequence to school personnel; local stakeholders developed the capacity to use information effectively to understand school performance; school personnel developed the capacity to work with local stakeholders.

Proposed mechanism: Parental ability to exert pressure to improve school performance (Parental pressure).

SETTING EXPECTATIONS: USES OF EMIS WITH LOCAL SCHOOL DEVELOPMENT PLANNING

Outcomes (desirable): Suggestions of positive impact of one effort in Nigeria appear in four papers, two of medium rigour (Bruns et al., 2011; USAID, 2006) and two of low rigour (Crouch and Winkler, 2008; Winkler, 2005). Two medium-rigour papers describe the positive impact of SDP initiatives in Vietnam (Attfield and Vu, 2013) and in China (Brock, 2009). Two papers (one low and one medium rigour) suggest a mixed impact from an initiative in Ghana (Prew and Quaigrain, 2010; Powell, 2006)

Conditions (China, Brock, 2009): Supplemental funding for implementation of the new national curriculum; delegation of some planning and resource allocation responsibility to school level; training of school leadership in educational management, coupled with teacher training and development of new curricular materials; environment for experimentation.

Proposed mechanism: Learning from failure.

Conditions (Vietnam, Attfield and Vu, 2013): Revised minimum school standards focused on basic inputs for learning; EMIS infrastructure for accurate and timely monitoring and reporting of school and district-level progress against standards.

Proposed mechanism: Reality testing.

Outcomes (undesirable): One paper of medium rigour suggests undesirable impact of a national initiative in Nepal to promote the local SDP: decreased ownership of local education issues by the school and community (Caddell, 2005).

Conditions: Lack of capacity at the district level to support participatory processes at the school level; 'participatory processes' were aimed at meeting district demand for information, not local participation.

Proposed mechanism: Lack of follow-up/follow-through.

6. IN-DEPTH REVIEW: INSPECTION

6.1 INTRODUCTION

This systematic review explores the conditions under which school inspections lead to improvement in schools and to positive learning outcomes for schoolchildren in low- and middle-income countries, especially the poorest and most marginalised children. The review focuses on developing countries in Asia, Africa and Latin-America.

The chapter has the following sections:

Section 6.2: Defining inspection

Section 6.3: Quality of studies

Section 6.4: Findings about types of activities

Section 6.5: Alternative strategy for external school inspections

Section 6.6: Conclusion and discussion

6.2 DEFINING INSPECTION

School inspections are understood as external evaluations of schools, undertaken by officials outside the school with a mandate from a national or 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). As De Grauwe explains (2001, 2007), 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 (2001, 2007) describes, the supervisors' role is often 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 even have managerial tasks and are, for example, responsible for the deployment of teachers, or deciding on the promotion of teachers and head teachers.

In this review, we recognise that a developmental brief held by the same role holder may give rise to different mechanisms and yield distinctly different outcomes. We are, however, particularly interested in one particular pathway to impact – the school-level evaluative dimensions of the role. This means that we will only look at inspection/supervision that has at its core an element of judgement, using a framework that allows for some level of comparison between schools, where the person responsible for making the judgement is external to the school (not present in the school on a day-to-day basis) or is responsible for more than one school. The judgement would typically also have consequences for schools/school staff, which may be punitive or in the form of additional support for schools/head teachers. We recognise however that those consequences are often not put in place in developing countries, due to limited resources (e.g. no funding for additional monitoring), but the authority undertaking inspections/supervision should have a formal role/position to potentially enact such consequences. A judgement can include an aggregate score for the school (e.g. as failing or well performing), but may also include an overview of strengths and weaknesses. The judgement is

communicated to the school, and typically also (but not necessarily) published in an inspection report, and made available to the school and the wider community.

The sources we reviewed discuss the implementation of inspections in low and middle income countries (LMICs), and particularly highlight the problems these countries face in ensuring high quality inspections of schools. These problems are explained below as they are relevant conditions to all the mechanisms of possible outcomes of inspections discussed in subsequent sections.

LACK OF RESOURCES

The first set of problems has to do with the lack of financial and material resources, such as cars and fuel to visit (remote) schools and the lack of computers and stationery to prepare those visits and follow up with written inspection reports. Six studies in Botswana, Namibia, Tanzania, Zimbabwe, Timor Leste and South Africa present findings from primary research to explain how the lack of resources results in infrequent and limited visits to schools, particularly those in remote areas (high-rigour study by De Grauwe, 2001; medium rigour study by Uwazi, 2009; low rigour studies by Herselman and Hay, 2002; MacPherson, 2011; Mazibuko, 2007; Wanzare, 2002). These schools may go without an inspection visit for many years. Four other studies include a discussion of similar problems in Peru and Africa but do not present actual research data (high-rigour study by Alcazar et al., 2006; medium-rigour study by Moswela, 2010; low rigour studies by De Grauwe, 2007, 2008).

Many countries have seen an increase in the number of students and schools over the last years in an effort to meet the millennium goals of equal and full access to schools for children. The number of supervisors has often not kept pace with the number of schools and teachers, leading to a high supervisor/school ratio. This is particularly a problem when supervisors have to cover long distances to visit schools in remote areas and are also tasked with many other (administrative, managerial or school improvement) duties.

One of the studies in Tanzania (medium-rigour study by Uwazi, 2009) also describes how the distribution of schools across inspection districts is not based on a school's need to be inspected; rather, inspection schedules focus on administrative coverage. As a result, schools are inspected infrequently or not at all and there is little practice of more targeted visits to schools in need to potentially increase the impact of inspections. The only study that reports sufficient resources is a medium-rigour study by Chen (2011) in Indonesia.

Lack of stationery and computers also limit the publication and dissemination of inspection reports to schools and other stakeholders and also limit the collection and analysis of relevant school documents and data (e.g. school development plans, school self-evaluations, census data) in the preparation of upcoming inspection visits.

WORKLOAD

A related issue is the ambiguity around the main functions and workload of school inspectors. Four studies (one high rigour, one medium rigour, two low rigour) in Botswana, Namibia, South Africa and Kenya present results from primary studies to explain how inspectors, particularly in African countries, are often tasked with a number of roles around the control and support of schools and lack a clear mandate (De Grauwe, 2001; Mazibuko, 2007; Moswela, 2010; Wanzare, 2002). The number of activities they are expected to undertake, given the number of schools within their remit, adds to their work load and also limits the time they can

actually spend on school inspection visits. De Grauwe (low rigour, 2007, 2008) emphasises these findings in two non-empirical papers in which he talks about school inspections, particularly in Africa:

> 'There is a profound conflict between the mandate of the service and its resources. The mandate is very demanding: to exercise control over and offer support to all schools and teachers, while informing schools of ministry policies and bringing school realities to the attention of decision-makers. The expansion in the numbers of schools and teachers has not been accompanied by an equal expansion in the numbers of supervisors, the evident result being that each supervisor has so many schools under his or her charge that they simply cannot visit all schools more than once or twice a year, if at all'. (De Grauwe, 2008, p.3)

EDUCATION CONTEXT

Not surprisingly, some of the same problems we found in the implementation of school inspections can also be found in the broader context of schools, and they are also considered to be major reasons for the low performance of schools and the limited capacity to improve in response to inspections. The available studies suggest that contextual issues, such as an overall lack of resources (in trained teachers, textbooks, support of schools) need to be addressed before schools have the capacity to improve and school inspections can have an impact.

The first set of problems in the context of schools is a lack of resources. Five studies in Peru, Ghana, Pakistan, Botswana and South Africa of varied quality (two high rigour, one medium rigour, two low rigour) indicate a lack of teachers, textbooks and low salary of teachers, requiring them to take on a second job, as conditions that hamper improvement of schools in response to inspections (Alcazar et al., 2006; Darvas and Balwanz, 2014; Jaffer, 2010; Mazibuko, 2007; Moswela, 2010). Similar findings are discussed for Africa and Timor Leste by De Grauwe (low rigour, 2008) and Macpherson (low rigour, 2011).

The rigorous study by Alcazar et al. (2006) for example explains how, in Peru, a community's remoteness and poverty level are strong predictors of high teacher absence in primary education, how there are few incentives (or consequences) to avoid teacher absenteeism as teachers' pay is, for example, not related to their performance. Jaffer (low rigour, 2010) for example states:

> 'Many government teachers hold other jobs to supplement their income, for example teaching in a private school or managing their own tutoring centre. This creates many issues, including teachers paying more attention to their other work'. (p.387)

De Grauwe (low-rigour study, 2008) also discusses the lack of availability of support services for teachers; those services have, according to this study, been ignored in almost every country (Botswana, Namibia, Tanzania, Zimbabwe) for a long time since resources have become more scarce and have caused a deterioration in the quality of basic education. Schools, however, require support to act on inspection findings and to prepare for inspection visits.

The size of the country and the accessibility of rural community schools are key issues here as they have an impact on the accessibility of schools for inspection visits as well as for follow-up support and resources. Alcazar et al. (high rigour, 2006), for example, explain how a community's remoteness and poverty level are strong predictors of teacher absenteeism in Peru, while Darvas and Balwanz (high rigour, 2014) talk about an 'access challenge' and inequitable distribution of resources to schools.

Other relevant issues in the context of schools are the *lack of knowledge in schools* to improve school quality and to understand and prepare for evaluations (two low-rigour papers on South Africa), the overall *lack of consequences* in the system to improve performance and teach to a high standard (one study of high rigour in Peru, one of low rigour in Uganda).

The cultural and political context is referred to in seven papers, on Ghana, South Africa, Pakistan and India. Six of these seven papers present findings from primary studies, but only the Ghana study, from Darvas and Balwanz (2014), is of high rigour. All the papers, however, present similar issues in explaining how the political and cultural context has an impact on the effectiveness of school inspections, such as when teacher unions resist school inspections and provide buffers for teachers and head teachers to ignore inspection recommendations (such as in Africa and India; see papers of low rigour from De Grauwe, 2008 and in Pakistan, Jaffer, 2010), or when the state invests its efforts and support in groups who are important to the survival of the state, favouring schools with school staff that have strong political affiliations. A high-rigour study in Ghana also suggests that inspections in those schools are biased and only lead to favourable reports. The cultural and political context is also relevant to the overall pressure on politicians to improve education and may impact on the distribution of resources to education in general and to specific regions or groups of schools (see Darvas and Balwanz, 2014). Chen (2011) additionally explains, in a medium-rigour study, how the cultural context has an impact on the functioning of such local decision making and accountability structures. According to Chen, accountability of schools to parents in Indonesia is, for example, not likely to work as 'community harmony is highly valued and a majority of parents are reserved and do not openly complain or express dissatisfaction' (p.14).

6.3 QUALITY OF THE STUDIES

In this review we analysed 22 papers, using the method previously described in Chapter 2. The papers included one conference paper, one dissertation, four case studies and eleven scientific papers. Three non-empirical papers provide narrative descriptions of school inspections across a number of countries, particularly in Africa, outlining the problems that these inspection systems face in inspecting schools. The majority of empirical papers are from small-scale case studies, often including limited descriptions of underlying methodologies and presenting self-reports of small (potentially non-representative) samples of respondents. Only three papers report quantitative results from surveys and secondary data. Table 6.1 provides an overview of our assessment of rigour of these 22 sources, indicating that only three studies were counted as rigorous (and two of them did not have inspections as the main topic of the study but only discussed them as a sideline), while eight papers were of medium rigour and 11 were of low rigour).

Table 6.1: Reviewers' judgements about the rigour and relevance of each study included in the inspection synthesis

Studies (first author and date)	Rigour			Relevance		
	High	Medium	Low	High	Medium	Low
Alcazar (2006)	✓				✓	
Barrett (2011)			✓		✓	
Brock (2009)		✓			✓	
Chen (2011)		✓			✓	

Studies (first author and date)	Rigour			Relevance		
	High	Medium	Low	High	Medium	Low
Churches (2013)			✓	✓		
Crouch (2008)			✓	✓		
Darvas (2014)	✓				✓	
De Grauwe (2001)	✓			✓		
De Grauwe (2007)			✓	✓		
De Grauwe (2008)			✓	✓		
Harber (2006)		✓			✓	
Herselman (2002)			✓	✓	✓	
Jaffer (2010)			✓	✓		
Kingdon (2012)		✓		✓		
Macpherson (2011)			✓		✓	
Mazibuko (2007)			✓	✓		
Moswela (2010)		✓		✓		
Opoku-Asare (2006)		✓		✓		
Santiago (2012)		✓			✓	
Uwazi (2009)		✓		✓		
Wanzare (2002)			✓	✓		
World Bank (2010)			✓		✓	

6.4 FINDINGS ABOUT TYPES OF ACTIVITIES

The available papers highlight a lack of impact of school inspections. The evidence tables highlight that only 11 papers (one of high rigour, three of medium rigour, seven of low rigour) address the potential effectiveness of inspections, of which only three (medium-rigour) sources actually present study findings on the (lack of) effects of inspections, while other papers are primarily presentations of the authors' personal viewpoints.

Two of the 11 papers point to unintended consequences where teachers in Ghana put on an act during inspection classroom observations (medium-rigour, Opoku-Asare, 2006), and principals use inspections to threaten their teachers in South Africa (low rigour, Mazibuko, 2007). According to Opoku-Asare (2006), school inspections are often pre-announced and thus enable the teachers concerned to prepare adequately for the observation lessons by arming themselves with all the teaching materials they can possibly lay hands on, and sometimes rehearse the lessons they intend to teach for the exercise (p. 112).

The papers that indicate a lack of impact include small-scale studies of low rigour in Uganda (Crouch and Winkler, 2008), South Africa (Herselman and Hay, 2002; Mazibuko, 2007), and Timor Leste (Macpherson, 2011). Only one study of medium rigour from the Tanzania National Audit Office (Uwazi, 2009) refers to a lack of improvement in national student achievement data to support the claim of limited impact of school

inspections. Mazibuko (2007) in a case study of low rigour, however, suggests that school staff learn from evaluations even though they haven't implemented any recommendations. Alcazar et al. 2006, high rigour), De Grauwe (low rigour, 2007, 2008) and Jaffer (low rigour, 2010) additionally refer to anecdotes and discussions about a lack of impact of school inspections in Peru, Botswana, Namibia, Tanzania, Zimbabwe and Pakistan.

Only two papers (one of low rigour, one of medium rigour) suggest that inspections have an impact. Macpherson (2011), a case study, describes how school inspections in Timor Leste have the potential to contain the scale of corruption in the misuse of school grants when policing transparency in their collection and disbursement, while not engaging in the processes of collection and disbursement themselves. As the study only looked at how schools were investigated and how inspectors investigate allegations of misuse, no claims can be made about school inspections actually leading to a decrease in corruption.

Brock (2009) draws on a number of case studies in Gansu when explaining how increased power to school inspectors to report on the quality of schools, and to propose changes and support in/of schools, lead to an improvement in school development planning: schools set out specific goals for their development in close cooperation with the local community, in which they take into account the needs of poorest children and developed learning materials to address these needs. These school development goals could subsequently be measured by inspectors.

Below, we use the programme pathways defined in our initial rough theory to organise our presentation of outcomes, related conditions and proposed mechanisms.

SCHOOL INSPECTIONS: PROVIDING FEEDBACK INFORMATION

In our initial rough theory, we suggested that inspections could have an impact on service delivery and improved learning outcomes through the feedback from school inspectors on strengths and weaknesses in school quality. Feedback refers to the priorities for improvement that are set and communicated to schools (adapted to the local context), and targeting weak schools for visits and feedback. Fourteen papers reflect on inspection feedback and communication in relation to school inspections. Of these, only eight papers (one of high rigour, five of medium rigour, two of low rigour) present actual study findings, while six papers only make claims about the functioning of inspection feedback.

All of the papers describe a lack of impact from inspection which is, in the authors' views, caused by specific attributes of inspection feedback, and by a number of conditions. These attributes and conditions fail to 'fire' any improvement mechanism from inspection feedback, such as when school staff accept inspection feedback and use it to improve the school's weaknesses to enhance student outcomes. Each of these conditions will be described below and how they have failed to lead to improved outcomes.

CONTENT OF THE FEEDBACK

Several authors discuss the importance of feedback and communication of inspection findings in school improvement and claim that the content of the feedback is an important cause of limited improvement from inspection feedback. Three papers (two of medium rigour, one of low rigour) in Indonesia and Ghana present findings from primary research which indicate that inspections particularly focus on bureaucratic and administrative issues, checking figures and compliance to regulations which are not considered to be relevant

for school improvement and are often outside the school's span of control (Chen, 2011; Darvas and Balwanz, 2014; Opoku-Asare, 2006). Inspection recommendations do not focus on vital problems in schools and are often repeated in a routine manner, year after year. As a result, schools fail to accept and/or implement inspection recommendations. Uwazi (2009, medium rigour) presents study findings which suggest that such a focus on administrative and bureaucratic issues takes time away from schools that they could use to focus on actual improvement of student outcomes. Similar claims have been made by Santiago et al. (2012, medium rigour) in a country review of Mexico.

As Uwazi (medium rigour, 2009) and Opoku-Asare (medium rigour, 2006) explain, the inspection recommendations are often generic and unrealistic and require additional resources that the school administration is not able to acquire. These papers highlighted how the Tanzania and Ghana Inspectorates of Education often provided advice to schools which should be aimed at the Ministry of Education, such as hiring more teachers, acquiring more textbooks, or constructing/renovating school buildings. Uwazi (2009) suggests that inspections can only be effective if they address issues of poor-performing students, how to address dropout rates, and how to improve learning and instruction and/or training gaps in schools. Similar issues are discussed in papers from De Grauwe (low rigour, 2007), Jaffer (low rigour, 2010), Santiago et al. (medium rigour, 2012) and Wanzare (low rigour, 2002) referring to Africa, Pakistan, Mexico and Kenya.

The conditions that are expected to have caused such ineffective feedback are explained in papers from Churches and McBride (2013, low rigour), Darvas and Balwanz (2014, low rigour), De Grauwe (2001, highly rigorous), De Grauwe (2007, low rigour), De Grauwe (2008, low rigour), Harber (2006, medium rigour; Jaffer (2010, low rigour) and Wanzare (2002, low rigour). These authors point to the work overload of inspectors (both in numbers of schools and in number of indicators to inspect) which lead them to focus on a simple checking and control of administrative protocols. Performance management systems hold inspectors to account for the number of schools visited (instead of impact and quality of feedback), which would lead them to focus on checking facts and figures. Other conditions of ineffective feedback are a lack of professionalism of school inspectors and lack of training in evaluation of school quality. School inspections are often also prioritised for schools that are suspected of irregular use of resources and misconduct of teachers, while inspectors also seem to feel that control of administrative procedures gives them power over schools and authority in their evaluation, as it would signal a clear mandate from central government. Such status, credibility and authority is often lacking (see the section below).

A number of authors also suggest that inspection feedback and standards need to fit within the local context to effectively motivate school improvement. The local context is seen by many authors as important in ensuring that school inspections address local priorities and issues. None of these papers have, however, actually investigated a relationship between adapting inspection standards to local contexts and the improvement of schools, but they suggest that inspection purposes and priorities need to be adapted to the history and culture that underpin the local context of a country, or of different regions within a country to advance school improvement, particularly in heterogeneous countries (low rigour, De Grauwe, 2008). Inspectors need to have an open mind in order to recognise excellence and understand the existing restraints on pedagogy that exist in a specific context (such as class size and resourcing). Such understanding is needed to provide relevant solutions and feedback that support the improvement of schools and to ensure acceptance and use of the feedback by relevant stakeholders.

COMMUNICATION AND TONE OF VOICE

Feedback is often also not considered to be effective because of the hostile and intimidating tone of voice of school inspectors. This is mentioned as a problem in Africa by De Grauwe (low rigour, 2007, 2008) and Wanzare (low rigour, 2002). They haven't studied the actual communication of school inspectors in schools but suggest from experience and from referencing other studies that their attitudes can be condescending (De Grauwe, 2007, p.711) and that:

> 'Inspection of schools in Kenya has at times been marked by impromptu, irregular visits by some inspectors with the object of 'catching' the teachers doing the wrong. Some school inspectors have been criticized for being harsh to teachers and for harassing teachers even in front of their pupils'. (Wanzare, 2002, online version)

The environment in which instructional supervision takes place in schools (in Botswana) is rather hostile and too intimidating for teachers to make any meaningful impression on the improvement of teaching standards (medium rigour, Moswela, 2010).

As the findings from Moswela's study suggest, a condescending tone in communicating with schools and presenting feedback leads to a poor relationship between teachers and inspectors and makes no impression on teachers, resulting in a lack of impact on improvement of schools.

LACK OF CREDIBILITY OF INSPECTORS AND INSPECTION FEEDBACK

The lack of perceived expertise, status and credibility of school inspectors by school staff is also expected to limit the implementation of inspection feedback. Eight papers (one high rigour, four medium rigour, three low rigour) in Nigeria, Botswana, Namibia, Tanzania, Zimbabwe, Eastern and Southern Africa, Pakistan, Timor Leste, Mexico and Kenya present findings from primary research to explain that school inspectors are often not trained in the evaluation of schools and have limited expertise in doing such evaluations which would limit their credibility and the credibility of inspection findings in the eyes of school staff (De Grauwe, 2001; Harber, 2006; Jaffer, 2010; Macpherson, 2011; Moswela, 2010; Santiago et al., 2012; Uwazi, 2009; Wanzare, 2002). They have no expertise in how to objectively evaluate schools/teachers, how to provide accurate and consistent feedback on strengths and weaknesses or how to engage schools in a professional dialogue about school improvement. As Santiago et al. (2012) comment in an OECD review from Mexico:

> 'However, it was reported that in general there is much variation in the quality of advice and support supervisors may be able to offer schools. The capacity of supervisors in general to engage in school evaluations in ways which may promote school improvement as well as resulting in accurate evaluation of the quality of a school's work is limited under present conditions' (p.155).

These papers also highlight how the overall lack of systems and structures around human resource management and development to support the hiring and training of a high-quality inspection core hampers the credibility of school inspectors. Studies in Botswana, Namibia, Tanzania, Zimbabwe (low rigour, De Grauwe, 2007; medium rigour, Moswela, 2010), Pakistan (low rigour, Jaffer, 2010) and Mexico (medium rigour, Santiago et al., 2012), for example, they explain that inspectors are often recruited on an ad hoc basis from a pool of teachers and principals with long service, where personal connections and political affiliations are used to transfer people into inspection posts.

Similar issues are discussed in papers from Churches and McBride (low rigour, 2013), Darvas and Balwanz (low rigour, 2014), De Grauwe (low rigour, 2008) and Harber (medium rigour, 2006), referring to Nigeria, Ghana, and Eastern and Southern Africa. Studies in Pakistan (low rigour, Jaffer, 2010), Botswana and Tanzania (high rigour, De Grauwe, 2001) also suggest that the lack of credibility is caused by recruitment issues (favouring individuals with high political influence), and the pay scale of school inspectors, which is on a lower grade than that of the head teachers they are inspecting, causing head teachers to believe that inspectors are not of a high status and that their feedback can be disregarded:

> 'The inspector's position was equivalent to the teaching grade of a high school teacher, so these teachers and the inspectors were at the same grade and salary scale. Hence, individuals could not be held accountable for sub-standard performance or rewarded for good performance. No pre-service training was provided to prepare the inspectors for the specific roles and responsibilities of the post. These lacunae further weakened the inspectors' position and authority, and also impacted adversely on the efficiency of the inspection system'. (low rigour, Jaffer, 2010, p.378)

There also seems to be little incentive in place to improve the overall quality of inspections as the number of visits to schools is the main performance indicator used in the evaluation and monitoring of school inspectors, and not the quality of their work. A study of medium rigour in Tanzania by the National Audit Office (Uwazi, 2009) also highlights that there is no monitoring of inspection systems to learn about what works and doesn't work and to improve the functioning and impact of inspections. According to this study, such monitoring needs to analyse the extent to which schools have implemented recommendations, stakeholders' perceptions of the recommendations and the impact of implemented recommendations:

> 'It shows that the school inspectorate programme is not functioning properly and therefore fails to safeguard quality of instruction and its improvement by: failing to prioritise the issues of poor performance of students in the inspection cycle, not effectively communicating and following up on implementation of recommendations, failing to monitor the effectiveness of school inspections' (Uwazi, 2009, p.1).

An important condition for feedback to lead to improvement is also the capacity and knowledge in schools to address and implement improvements and to effectively engage in whole-school evaluation and school inspections. As Mazibuko (2007), in a non-rigorous study of school inspections in South Africa (KwaZulu-Natal), explains:

> 'Principals and educators interviewed in this study maintained that financial constraints and other crucial issues make it difficult for their schools to address areas or issues identified by the supervisors as areas that need attention for the development of these schools. Lack of resources, overcrowding, lack of support from DfE (resources, information). Principals and educators claim that financial constraints prevent them from addressing the recommendations of the supervisors'. (p.229)

and

> 'Clear understanding of whole-school evaluation and its implications can lead to proper implementation. But the study reveals that participants, particularly principals and educators, had only a general understanding of whole-school evaluation, as they had not undergone training. This implies that principals and educators do not really understand the pros and cons of whole-school evaluation. Because of this shortcoming, schools cannot conduct whole-school evaluation effectively. The principal of school D contended that most educators have a negative attitude towards

wholeschool evaluation. This can be attributed to the fact that most educators have not been trained on whole-school evaluation and may not fully appreciate the significance of whole-school evaluation in schools. Lack of understanding and knowledge impede most educators from fully participating with the supervisors during the external evaluation'. (p.201)

SCHOOL INSPECTIONS: CONSEQUENCES FROM INSPECTION FEEDBACK

The lack of consequences for failing schools and the lack of follow-up on inspection visits is also claimed to be an important condition for the overall lack of impact of inspections in LMICs in 10 sources. Seven studies (one of high rigour, three of medium rigour, three of low rigour) report results from primary research which shows that inspectorates of education in Indonesia, Uganda, Namibia, Tanzania, Ghana and Kenya cannot sanction failing schools and have no interventions in place to motivate school improvement. Even if schools can be sanctioned by law, there are no means or mechanisms in place to actually implement such consequences:

> 'Key aspects of quality assurance and consequences are missing from these [inspection] visits [in Indonesia]' (medium rigour, Chen, 2011, p.13).

> 'The lack of rewards and sanctions for good or bad performance leaves the system weak' (medium rigour, Chen, 2011, p.24).

A study of medium rigour in Ghana (Opoku-Asare, 2006) also shows that a lack of consequences actually results from too close relationships between schools and inspectors and negative inspection reports not being published. Jaffer (2010, low rigour) also presents study findings which indicate that the lack of credibility of inspectors (due to a lack of training, political appointments into post, and pay scale similar to teachers) renders them powerless in holding schools accountable for low performance. Similar findings on the lack of consequences from inspections have been discussed in a study of low rigour by Churches and McBride (2013) on Nigeria.

Consequences of school inspections, such as rewards for high-performing schools, sanctions for failing schools and follow-up support for and monitoring of school improvement, are expected to motivate improvement through their enforcement of schools' compliance to inspection standards, their incentives for effective behaviours, and the fact that they give prominence to inspection feedback and credibility to school inspectors, and force schools to act on inspection feedback. Findings from a study by Brock (2009, medium rigour) suggest that inspectors' power to report, propose changes and support may enhance school development planning, particularly when inspectors measure school goals. De Grauwe (2001, high rigour) also explains how school heads in Nigeria and Tanzania face similar issues, as they often have no power to ensure that their staff implement improvements from inspections, while there is also limited management capacity at the district and school levels to support the implementation of inspection recommendations, leading to an overall lack of impact of inspections.

SCHOOL INSPECTIONS: SETTING EXPECTATIONS

Six studies (one of low rigour, two of medium rigour, three of low rigour) present findings from Gansu (China), Uganda, Pakistan, South Africa, Mexico and Peru that indicate how the development of standards, guidelines and frameworks to inspect schools can be an important driver for improvement, as they inform schools where to focus on in their improvement plans, support school self-evaluation and ensure consistency of inspection assessments and feedback to schools. Currently, many developing countries do not have such guidelines to

evaluate the quality of schools which leads to inconstancy in the evaluation of schools and also limits schools in their preparation for visits and in the incorporation of inspection standards and criteria in their school development planning. As Brock (2009) explains that frameworks and guidelines are made available to both schools and inspectors. Availability to schools supported school development planning as a vehicle for change in which schools and local communities are brought together to create a unified approach to the school's development, and in which the relationship between the county education bureau and the school changes from top down to bottom up development.

It seems that openness of frameworks and inspection schedules allows schools to prepare for visits and creates buy-in to the inspection process, which would promote improvement. As De Grauwe (2001, p.17, high rigour; 2008, low rigour, p.14) suggests that openness and transparency are increasingly encouraged, implying that schools will be informed beforehand of visits. As a result, where these reforms are actually being implemented, teachers are beginning to consider inspectors as sources of help rather than of criticism, and start applying the same frameworks and norms to the evaluation and improvement of their work throughout the country (Grauwe et al, 2001, 2008).

Brock (2009, medium rigour) also found that making frameworks and guidelines available to schools and inspectors created a more bottom-up and unified approach to school development as it brought together schools and local communities in setting priorities for improvement. These processes to school development are, according to Churches and McBride (2013, low rigour) and Moswela (2010, medium rigour), enhanced when stakeholders, such as principals, proprietors, employers, higher education providers and teachers are involved in the design of frameworks and buy-in is created for the evaluation of teachers and schools. Such buy-in is an important condition for the impact of inspections, as teachers and teacher unions have strong power positions to resist inspections and often do so (De Grauwe, 2008, low rigour). Openness and transparency of frameworks, buy-in to these frameworks, and bottom-up processes of school development seem to result in standardisation of quality across a country and may have, according to De Grauwe (2008) a desirable impact on improvement of schools, particularly in homogenous countries with few disparities.

Santiago et al. (medium rigour, 2012) suggest that such tools and guidelines can support schools in engaging in self-evaluations and enhance a common language of quality in a country. Guidelines and frameworks also support school inspectors in their evaluation of schools and enable them to have a professional dialogue with school staff about potential improvements. Such a dialogue is considered to support the school's acceptance and use of inspection feedback. An important condition is, however, according to Santiago et al. (2012) to prevent schools from being overloaded with different types of guidelines and materials, as this will confuse them when deciding which approach to focus on in improving the quality of their school. Wanzare (low rigour, 2002) also discusses how schools can use inspection handbooks in their evaluations if these handbooks are not too detailed, bureaucratic or rigid. Churches and McBride (2013) suggest that buy-in and use of these handbooks and standards by stakeholders (such as schools) is enhanced when they are involved in their design.

SCHOOL INSPECTIONS: CAPACITY BUILDING OF EDUCATORS

Capacity building of educators particularly refers to the ability of schools to evaluate and improve their own performance, and the capacity to implement improvements. Studies discuss how linking external inspections and internal school self-evaluation may motivate self-evaluation of schools, and suggest how this may lead to more sustainable improvement. None of the authors have however studied the relationship between strengthening internal evaluations in inspection systems and actual school improvement.

De Grauwe (high rigour 2001; low rigour, 2007, 2008) and Herselman and Hay (2002, low rigour) expect that an increase in school internal evaluation will strengthen the participation and commitment of teachers in school change and create a culture of quality in which teachers reflect on their own practices, which is expected to lead to more sustainable improvement. External support for internal evaluations and resulting improvement, such as from resource centres, may strengthen linkages between schools and prevent their isolation. It is also expected to strengthen school management and culture and the school's capacity to improve.

According to De Grauwe (low rigour, 2008, p.15), internal evaluations of schools in response to external inspections can however only be effective if school inspectors take these evaluations serious when they inspect the schools, if there is overlap in internal and external frameworks and criteria for making a judgement about school practices, and if the agenda for these self-evaluations fit the improvement priorities of schools and countries (instead of being driven by donor organisations). Schools also need support in the implementation of rigorous self-evaluations. Wanzare (low rigour, 2002) discusses how schools can use inspection handbooks in their evaluations if these handbooks are not too detailed, bureaucratic or rigid, while Moswela (2010) suggests that teachers who have an active part in inspections throughout the year are better able to improve their teaching.

SCHOOL INSPECTIONS: CAPACITY DEVELOPMENT OF STAKEHOLDERS

A final condition for effective inspections is the alignment of actions of stakeholders in the education system and their capacity to implement and support school improvement. These conditions refer to both the building of capacity of stakeholders in our initial rough theory and the setting of expectations.

De Grauwe (high rigour, 2001), Mazibuko (low rigour, 2007) and Opoku-Asare (medium rigour, 2006) describe how limited co-ordination between the inspectorate of education and other national stakeholders, such as teacher training or resource centres in the dissemination and use of inspection findings, potentially limits the impact of school inspections. Limited co-ordination between the inspectorate and other stakeholders in the education system particularly leads to a lack of follow-up on school inspection visits and findings, and limited support to schools in the implementation of inspection feedback. Alignment and co-ordination between inspectorates of education and other education service providers or stakeholders in the education system are expected to enhance the impact of school inspections, as they ensure that school improvement efforts across the system focus on the same standards (preventing confusion of schools) and that there are consequences and follow-up from inspections.

Six papers discuss the relation between alignment and follow-up on inspection assessments and school improvement, but only one study (low rigour, Jaffer, 2010, in Pakistan) actually report findings that support such a relationship:

> 'The problem, as the respondents indicated, was that others rarely followed up on the supervisor's recommendations. As one respondent put it, 'the higher authorities just write 'seen' on the supervisor's recommendations without taking any action. There is no decision on the actions that we have suggested for school improvement. They ignore our note. And so we know that nothing will come out of these reports and efforts'. (p.386)

Other authors discuss similar issues, such as De Grauwe (2001, p.143), who explains:

'Co-ordination is difficult, especially between the supervision service and other services which work towards pedagogical improvement, such as teacher training, teacher resource centres, curriculum development and examinations. The follow-up to school visits suffers from this lack of coordination. Recommendations made in inspection reports and addressed to the administrative and/or pedagogical authorities, remain words in the wind, which frustrates the school staff as well as the supervisors'.

Co-ordination and alignment are needed to disseminate knowledge from inspection visits and make sure that relevant actors follow-up on inspection recommendations. Alignment is also strongly related to buy-in to inspection standards and recommendations from relevant actors (such as teacher unions), who will then support and pressure school staff to act on inspection findings (see the previous section).

Follow-up on inspection recommendations through better alignment of the actions of stakeholders in the education system is needed to prevent an overall sense of inertia and demotivation, as is evidenced in a statement from De Grauwe (low rigour, 2007):

'Several strategies have as an objective to make external supervision more effective, by systematising the follow-up or formulating a more coherent and therefore less demanding job description. Supervision visits seldom lead to a well-organized follow-up, by the supervisors themselves, by the administration or by the schools. ….This lack of follow-up, the result of the powerlessness of supervisors and of the complexity of decision-making in a bureaucracy such as the educational administration, is frustrating to teachers and discredits the supervision system. It is also a core reason for supervision's feeble impact on quality'. (p.711)

Examples of how such alignment can be improved are given in study of high rigour by De Grauwe (2001), who describes:

'Relationships between supervision and the other pedagogical services are close and institutionalized in Botswana, supervisors being members of committees and panels in charge of curriculum development, teacher training and examinations. In Zimbabwe, their involvement is less institutionalized but still quite intense: supervisors serve as resource persons in training and participate in writing test items, marking examinations and preparing evaluation reports. In Namibia, however, the situation tends to the opposite, with no formal contacts and very few informal ones between supervisors and other pedagogical support staff. Supervisors are, for instance, not represented on the examination board, neither will they be on the Regional Education Forum…In Tanzania, in principle, supervisors do sit on curriculum panels and help with examinations, but in practice their involvement in pedagogical improvement is limited because of the practical constraints on their work' (p.44).

Alignment of and co-ordination between different agencies and offices are, however, difficult, according to De Grauwe (2008, low rigour), as it goes against the sense of independence that many of these agencies have, and their differences in opinion about, for example, adequate teaching methods. Mazibuko (2007, low rigour) also found that in South Africa (KwaZulu-Natal) there was a lack of clarity in the roles of support offices, districts and inspectors which hampered the support of schools in using inspection recommendations to improve.

6.5 ALTERNATIVE STRATEGY FOR EXTERNAL SCHOOL INSPECTIONS

A range of studies suggest alternative strategies to evaluate and monitor schools which are thought to be more effective, given the limited resources outlined in the introduction and the lack of impact. The suggested alternative strategy focuses on enhancing the level of monitoring and support of schools through the involvement of the local community and of parents. One study of high rigour from De Grauwe (2001) in Botswana, Namibia, Tanzania and Zimbabwe, presents actual findings from research on how local accountability was strengthened in these countries. Other studies (high rigour: Alcazar et al., 2006; medium rigour: Moswela, 2010, Santiago et al., 2012; low rigour: De Grauwe, 2007, 2008; Herselman and Hay, 2002; Wanzare, 2002) also discuss potential benefits of local accountability in Peru, Africa (South Africa, Kenya, Botswana) and Mexico. These studies assume that enhanced local accountability will benefit school improvement through enhanced monitoring, but none of the authors have actually studied the relationship between strong local accountability and monitoring and school improvement.

Local accountability is expected to address the limited resources for school inspection visits, and for follow-up on visits, particularly in remote areas. Local communities, district offices, resource centres etc. are seen as relevant actors to take over some of the evaluation tasks of the national inspectorate of education and, as a result, to increase the level of monitoring of schools.

This increased monitoring through the local community, with locally organised support for school improvement, is expected to improve the performance of schools. Alcazar et al. (2006), in a study of high rigour (although not on inspections), for example, state that in Peru's active oversight and involvement of the local community may improve performance through better monitoring, and De Grauwe (2007), in a paper of low rigour, states that 'several strategies have as an objective to make external supervision more effective, by bringing supervision closer to the school.

Chen (medium rigour, 2011) and Crouch and Winkler (low rigour, 2008), however, contradict the need for these local forms of evaluation and accountability and emphasise an expansion of the inspection system to include all schools and improve standards. These authors suggest that decentralised evaluation models cannot compensate for limited inspections due to, for example, cultural context, where parents do not openly complain or express dissatisfaction, such as in Indonesia, where community harmony is highly valued.

6.6 CONCLUSION AND DISCUSSION

In brief, our synthesis of the literature we have reviewed on school inspection in LMICs indicates the following:

- Inspections generally fail to have an impact in LMICs.
- Only three paper suggest potentially desirable effects when school staff learn from inspections, when school inspections contain the scale of corruption, or when school inspections improve school development planning, although these studies present no primary results to substantiate these claims.
- A small number of studies also suggest unintended consequences, such as teachers in Ghana putting on an act during inspection classroom observations, and principals using inspections to threaten their teachers in South Africa.

- The 'school-site supervision model' is suggested as an alternative model to school evaluation when a country has limited resources to implement school inspections. Such a model includes more localised forms of monitoring and evaluation, which would particularly work in heterogeneous countries where different regions have different priorities and cultures that need to be taken into account, and in countries that have strong local communities where there is a culture which supports such accountability.
- Alternatively, a number of studies suggest that the evaluation of schools, in the absence of external inspections, should be enhanced through the strengthening of school self-evaluation, where principals are tasked with the 'inspection' of their staff and/or school.

7. CONCLUSIONS AND IMPLICATIONS

We begin this chapter with a discussion of the changes we made to the review process. We then present a summary of the findings from the individual synthesis chapters, using this as a jumping-off point for the elaboration of our initial rough theory based on the configurations of context-mechanisms-outcomes that we discussed in relation to school inspection, assessment and monitoring. The elaboration of theory provides the basis for a composite discussion of findings and reflections on the initial rough theory as a means of explaining the connections among context and outcomes, with attention drawn to five hypothesised mechanisms. The final section offers an overview of the implications of the review for policy, practice and research.

7.1 SUMMARY OF THE FINDINGS

Firstly, we present a summary of key findings from each of the synthesis chapters, concentrating our focus on those areas of programme activity for which there exists some evidence of desirable or undesirable impact. We then use the corresponding C-M-O configurations to inform a discussion of implications for our initial rough theory and the hypothesised mechanisms that we have used throughout to present our syntheses: setting expectations, providing feedback/consequences, capacity development of educators, capacity development of stakeholders and institutionalisation of norms.

The question our review aimed to address was the following:

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

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

Our review focuses primarily on impact at the level of service delivery, emphasising the implications at the school level of accountability interventions, with a secondary focus on system efficiency and learning outcomes. Our rationale for this focus is that improvement of school quality has been the overarching concern of most LMIC national governments, as well as multilateral agencies, such as the World Bank and UNESCO, which has led to the widespread promotion of monitoring, inspection and assessment as policy levers for overseeing and improving schools (Bruns et al., 2011). We used a realist synthesis approach to understand the connections between particular conditions and school and system outcomes in LMICs, and the mechanisms of change that motivate school improvement from these three accountability approaches. 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 that might lead to equitable learning and, ultimately, overall system efficiency for the poorest and most marginalised children in LMICs.

Our initial rough theory included a set of generic hypotheses about how systemic accountability is intended to influence service delivery, as well as learning outcomes and system efficiency, based on the work of Bouckaert

and Halligan (2008), Ehren et al. (2013) and Hatch (2013). We identified five generic programme pathways. These are the hypothesised mechanisms which programme activities aim to trigger in order to lead to desired outcomes:

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

The following summarises our findings, first as a review of the key outcomes for each accountability element, and then a discussion of conditions and finally a recapitulation of the proposed mechanisms that we characterised based on our syntheses of the papers.

OUTCOMES

Papers relating to all three accountability elements show limited evidence of improved service delivery, improved learning outcomes or system efficiency. The assessment papers indicate unintended consequences, such as manipulating results and teaching to the test. The lack of evidence is partly due to the limited number of studies that specifically address these connections, as well as the lack of rigorous studies in the field. Across the three accountability elements, we identified programme activities that were designed to trigger outcomes through four of the five programme pathways: setting expectations, providing feedback/consequences, capacity building for educators, and capacity building for stakeholders. We summarise key outcomes for each accountability element and then offer comments looking across the three elements for each programme pathway. Table 7.1 depicts key outcomes per accountability element for each programme pathway.

ASSESSMENT

We examined two major types of programme activity: high-stakes examinations and low-stakes assessment.

Setting expectations: A range of evidence (five papers of medium rigour and three of low rigour) point to the undesirable impact of high-stakes examinations in setting expectations that yield adverse outcomes for school management, teaching and student learning.

Providing feedback/consequences: Four high-quality studies looked at a variety of different interventions related to incentives. Two report limited or undesirable impacts from group incentive initiatives at the level of teacher group and school. Two report desirable impacts, one of individual teacher incentives and the other of incentives for girl pupils.

Capacity development for educators: School-based assessments (SBAs) are examined in ten papers, seven of medium rigour and three low. All but one medium-rigour paper report a lack of outcomes, or unintended outcomes, including increased teacher workload, as well as teachers' alienation from the type of pedagogy that SBAs are designed to promote.

MONITORING

We synthesised 22 papers about monitoring, 4 of high, 12 of medium and 7 of low rigour. The papers examined three programme areas: the uses of EMIS for school-level management decisions; school report cards; and uses of EMIS with local school development planning (SDP). The first two areas are intended to bring about change by *providing feedback* to school personnel and other stakeholders; the last, SDP, aims to *set expectations* about the quality of schooling and school performance.

Setting expectations: Some evidence (six medium-rigour and three low-rigour papers) exists of positive impact of SDP initiatives in Nigeria, China and Vietnam; with mixed impact suggested in Ghana.

Providing feedback: A solid base of evidence, drawing on one high-, one medium- and one low-rigour paper, suggests that information alone had little effect but the provision of information with training and support at the local level had a sustained effect on management and student attendance. The evidence around school report cards is stronger, with three high-, two medium- and one low-rigour studies. All papers report an increased likelihood of beneficial effects on service delivery and learning outcomes from report cards. The evidence is mixed as to whether participatory approaches involving community members contribute significantly to effecting change at the school level.

INSPECTION

Our synthesis of inspection studies comprised 22 papers; three were of high rigour (although two of these did not have inspection as the main topic of study), eight were of medium rigour and eleven were of low rigour. Overall, there is limited evidence on the impact of school inspections in developing countries and little evidence of underlying mechanisms of change.

Setting expectations: Six studies (1 high, two medium and three low rigour) present findings from Gansu (China), Uganda, Pakistan, South Africa, Mexico and Peru that indicate how development of standards, guidelines and frameworks can be an important driver for school-level improvement. One high- and one low-rigour study suggest that openness of frameworks and inspection schedules allow schools to prepare for visits and create buy-in to the inspection process which can promote improvement. One study of medium rigour suggests that tools and guidelines can support schools in engaging in self-evaluations.

Providing feedback/consequences: Eight studies (one of high rigour, five of medium rigour, two of low rigour) present findings about the lack of impact from inspection caused by specific attributes of inspection feedback and inspectors providing the feedback, such as lack of credibility of inspectors (e.g. due to low pay scale), disrespectful tone of voice, and recommendations on administrative procedures and conditions out of the school's control. Six studies (one of high rigour, three of medium rigour, two of low rigour) report results from primary research showing that inspectorates of education in Indonesia, Uganda, Namibia, Tanzania, Ghana and Kenya cannot sanction failing schools and have no interventions in place to motivate school improvement. One highly rigorous study (De Grauwe, 2001) explains lack of impact by pointing to the limited power of school leaders to enact change in Nigeria and Tanzania, along with lack of district support for improvements. Brock (2009, medium rigour) suggests that inspectors' power to report, propose changes and support may enhance SDP, particularly when inspectors measure school goals.

Capacity development of educators: One high-, one medium- and four low-rigour studies suggest that an increase in school internal evaluation will (when used in school inspections) strengthen participation in and commitment of teachers to school change and sustainable improvement.

Capacity development of stakeholders: One high-rigour paper, one medium and one low describe how limited co-ordination between the inspectorate of education and other national stakeholders, such as teacher training or resource centres, in the dissemination and use of inspection findings, potentially limits the impact of school inspections.

Table 7.1: Outcome summaries for each accountability element by programme pathway

Programme pathway	Assessment	Monitoring	Inspection
Setting expectations	High-stakes examinations: limited evidence highlights unintended, undesirable consequences (6 papers of medium rigour and 2 papers of low rigour)	EMIS and School Development Planning (SDP): Limited evidence of SDP initiatives in Nigeria, China and Vietnam; mixed impact in Ghana (5 medium-rigour and 3 low-rigour papers)	Development of standards, guidelines and frameworks can be important drivers for school-level improvement (Gansu, China; Uganda; Pakistan; South Africa; Mexico; and Peru) (1 low- and 5 medium-rigour papers) Openness of frameworks and inspection schedules allows schools to prepare for visits and creates buy-in to the inspection process which can promote improvement (1 high- and 1 low-rigour paper) Tools and guidelines can support schools in engaging in self-evaluations (1 medium-rigour paper)
Providing feedback / consequences	Incentives as consequence: Limited impact and/or unintended consequences as a result of school-level incentives (2 high-rigour studies) Incentives as consequence: Significant positive results from individual teacher incentives in rural areas of India with limited	EMIS for school-level management: Information alone had little effect but the provision of information with training and support at the local level had sustained effect on management and student attendance (1 high-, 1 medium-, 1 low-rigour paper)	Feedback: Lack of impact from inspection caused by specific attributes of inspection feedback (1 high-, 5 medium-, 2 lo- rigour papers) Consequences: Inspectorates of education in Indonesia, Uganda, Namibia, Tanzania, Ghana and Kenya cannot sanction failing schools and have no interventions in place

Programme pathway	Assessment	Monitoring	Inspection
	unintended consequences (1 high-rigour study) Incentives as consequence: Significant positive results with limited unintended consequences in rural areas of Kenya from individual incentives for girls (1 high-rigour study)	School report cards: Increased likelihood of beneficial effects on service delivery and learning outcomes from combining dissemination of school performance information with participatory processes (1 high-, 2 medium-, 1 low-rigour paper) School 'report cards': Positive impact on school management without participatory processes; provision of comparative information alone yielded greater parental awareness of relative school quality (2 high-rigour studies)	to motivate school improvement (1 high-, 3 medium-, 2 low-rigour papers) Feedback/consequences: Lack of implementation of inspection recommendations due to limited power and management capacity of school leader to ensure implementation; limited management capacity at district level to monitor (1 high-rigour study) Feedback/Consequences: Inspectors' power to report, propose, support changes may enhance school development planning, when inspectors measure school goals (1 medium-rigour study)
Capacity development of educators	School-based assessment: Adverse consequences of school-based assessment on instructional practice and teachers (6 medium- and 3 low-rigour papers)		School internal evaluation will (when used in school inspections) strengthen participation and commitment of teachers in school change and sustainable improvement (1 high-, 3 low-rigour papers)
Capacity development of stakeholders			Impact of school inspections may be limited by lack of co-ordination between inspectorate of education and other national stakeholders (1 high-, 1 medium-, 1 low-rigour paper)

CONDITIONS

Our summary characterises conditions that lead to undesirable or unintended outcomes and those that lead to desirable school-level outcomes.

Setting expectations: The conditions that lead to undesirable outcomes are not surprising: they are those that promote goal displacement for school personnel, encouraging short-term or instrumental means to achieve the immediate goals of boosting examination results, preserving the impression of quality by manipulating information or putting on a show for inspectors. Pressure to perform that is not effectively mediated by appropriate and relevant internal and external support distorts the ultimate goal of high-quality schooling. Conditions that lead to setting undesirable expectations include:

- no internal support for teaching.
- no instructional leadership.
- no internal or external support for interpreting results.
- no external support for teaching or instructional leadership.

On the other hand, for all three accountability elements, evidence exists of ways in which educators have worked effectively with accountability pressures towards improvement responding to positive expectations. Setting expectations is enhanced when standards for evaluation and improvement are strongly aligned across levels of the education system and reflected in the required activities of stakeholders, encompassing such conditions as:

- internal (school-level) support for teaching and instructional leadership.
- external support for high-quality teaching and instructional leadership.
- internal and external support for interpreting results.

Conditions that limited the efficacy of feedback/consequences were those that promoted compliance rather than a developmental mindset towards improvement. Some characteristics of such conditions are:

- provision of feedback without local training and support;
- feedback lacks relevance to school priorities;
- incentive (as consequence) is generalised to group rather than individual;

Conditions that promoted the efficacy of feedback/consequences are those in which the feedback goes beyond one-sided sharing of information (e.g. of aggregated assessment results or school report cards); rather it is communicated in a meaningful manner to address local priorities, feeding into local school development processes, addressing issues that are within the control of the school to improve, and focusing on school conditions that are conducive to improvement of learning outcomes. This often requires support from national policy makers, or other stakeholders on the national level who are responsible for the implementation of inspections, assessment and monitoring (adding an element of capacity-building of national educators). The high-rigour study of Lassibille et al. (2010), in Madagascar provides some specific examples of conditions that promoted effective feedback:

- clarity of key workflow processes within and across classroom, school, sub-district and district levels.
- support in implementing workflow processes within and across levels.
- consistent and clear feedback about implementation of workflow processes and means to improve.

Repeatedly in our synthesis of conditions yielding undesirable outcomes, we encountered the lack of capacity of educators and stakeholders to engage with the envisioned processes. Simply put, ineffective interventions did not provide conditions for the development of educators' capacity to engage, whether in interpreting exam results or in making effective use of EMIS information or in conducting school self-evaluations as part of inspection. Conditions conducive to capacity development of educators include support of schools in the implementation of feedback from inspections, assessment and monitoring, and in their implementation of internal quality assurance systems to implement improvements on a more continuous basis. Papers across all three accountability elements argued for participatory approaches as providing conditions for the buy-in of schools to inspection, the quality of monitoring information, and bridging the gap between assessment results and action for improvement. Such participatory approaches were claimed to be beneficial for developing the expertise of school staff in the area of school evaluation and, it was argued, would shift perceptions and beliefs of educators towards more effective standards of high quality teaching and learning. Papers asserted that participatory approaches may also address some of the problems around lack of centralised resources for accountability (e.g. lack of transportation to deploy inspectors to school inspections in remote villages or those very difficult to access) and creating structures for more frequent evaluation of schools.

PROPOSED MECHANISMS

From our overview of conditions, we can now look more closely at each pathway to infer potential mechanisms at work in producing particular outcomes. Here we summarise the mechanisms we inferred from our synthesis of outcomes and conditions across the papers. The robustness of the inference is indicated by colour coding, with green consisting of multiple sources of evidence around conditions and outcomes, with at least one study of high rigour and a majority of papers of low rigour; orange consisting of one or more high-rigour paper or multiple source with a preponderance of medium-rigour papers; and red consisting of only one paper of medium rigour or more than one paper of low-rigour. Undesirable outcomes are indicated with grey text.

Table 7.2 Proposed mechanisms

Assessment	
High-stakes examinations	
Setting expectations through…	
emphasis on exam results (not intended)	Fear of bureaucratic authority
emphasis on exam results (intended)	Trust in the pedagogical authority of assessment approaches
Providing feedback/consequences through…	
information provision tailored to stakeholder groups	[additional evidence needed]

guidance and training	Follow-up/Follow-through
teacher and school group incentives (not intended)	Compliance with bureaucratic authority
individual teacher incentives (intended)	Individual desire for reward
Individual student incentives (intended)	Parental ability to exert pressure to improve child's performance
Capacity development of educators through school-based performance assessment…	
emphasis on procedural aspects (not intended)	Compliance with bureaucratic authority
lack of external support	Lack of follow-up/follow-through
Low-stakes assessment	
Setting expectations through…	
the establishment of curriculum standards (intended)	Trust in the pedagogical authority of assessment approaches
the establishment of curriculum standards (not intended)	[additional evidence needed]
Capacity development through guidance and support	[additional evidence needed]
Monitoring	
Setting expectations through uses of EMIS with local school development planning accompanied by…	
guidance and support, with some local autonomy	Learning from failure
revised minimum school standards	Reality testing
Providing feedback through…	
uses of EMIS for school-level management decisions	Follow-up/follow-through
school report cards, non-participatory design	Parental ability to exert pressure to improve school performance
school report cards, participatory design	Parental participation in monitoring activity

Inspection	
Setting expectations through…	
involvement in design of inspection frameworks	
support of school self-evaluations	Buy-in to inspection process
openness of frameworks and inspection schedules	
Providing feedback/consequences when…	
no rewards/sanctions from inspection results and disrespectful tone of voice	Lack of motivation to improve
feedback lacks relevance to school priorities and is beyond the school's control (not intended)	Compliance with bureaucratic authority
Capacity development of educators through self-evaluation	
inspectors' use of school self-evaluation in inspection	Buy-in to self-evaluation process and inspection feedback
provision of external support, guidelines, handbooks	
Capacity development of stakeholders through…	
alignment of standards and activities in support of school improvement	Follow-up/Follow-through

7.2 IMPLICATIONS FOR THE INITIAL ROUGH THEORY

Our findings around outcomes and conditions, together with the inferences we have made in advancing proposed mechanisms, cast our initial rough theory in a new light. Our hypothesised programme pathways of *setting expectations* and *providing feedback/consequences* featured in all three accountability elements. This was not surprising, given that these two programme pathways might be considered the Janus face of performance accountability, with one face anticipating feedback in the form of performance standards and the other delivering results. Capacity development, both of educators and of the broader universe of stakeholders, however, attained a different position and priority than initially envisaged. Our syntheses highlight repeatedly the central role of capacity development across all levels and within all activities. Moreover, the lack of development of capacity was central to reports of undesirable or unintended consequences in assessment and

inspection. The lack of explicit mention of capacity development in our account of monitoring may have more to do with the paucity of high-quality studies emphasising service delivery in this arena than with the state of monitoring, per se. Many of the papers synthesised in monitoring emphasised the system-level capacities necessary for the implementation of EMIS, or the community level in terms of reception of school report cards, but few studies explicitly addressed the implications of school-level capacity or the lack thereof. Finally, we did not find evidence of institutionalisation of norms as a programme pathway through which accountability interventions aimed to influence resources and reasoning at the school level. The institutionalisation of norms aims to highlight the underlying processes that enable effective communication and ongoing development; however, this is not necessarily a salutary process in all settings. Unintended consequences may arise around short-term goals and fulfilling bureaucratic requirements. We have also seen that the kinds of norms that the system might want to promote may be in conflict with one another – engagement of local stakeholders in participatory processes and upholding agreed-upon national education standards of quality, for example. On the other hand, our proposed mechanisms highlight processes that suggest dependence on shared values, attitudes and beliefs around accountability activity, such as the horizontal responsibility and mutual interdependence highlighted for effective monitoring activity. These underscore the institutionalisation of norms, even if that programme pathway is not one that is explicitly advanced.

CONNECTIVITY OF MECHANISMS

Additionally, our findings also indicate that our five mechanisms are interrelated and cannot be separated when explaining how each of the three accountability elements leads to improvement. Schools' acceptance and use of performance feedback (from inspections, assessment and monitoring systems) is, for example, indicated as an important mechanism of change across the papers, but many indicate the high level of support schools need to effectively use feedback and implement improvements, as well as adequate communication and distribution of feedback (e.g. the tone of voice of school inspectors, and proper communication of aggregated assessment results). Schools (and national policy makers), for example, need support in the interpretation of assessment data and in the implementation of data collection activities (e.g. in monitoring systems) in order to ensure the accuracy of the feedback and the identification of actual weaknesses and lack of resources that need to be addressed. Such support also sets expectations in schools around standards of good education and institutionalises external accountability standards. Support for the use of feedback therefore strongly links to our description of capacity building of educators and stakeholders, setting of expectations and institutionalisation of norms.

Similarly developing accountability systems with schools and local stakeholders (e.g. developing inspection standards, school report cards) enhances their capacity, but also sets expectations around evaluation and improvement and institutionalises external accountability norms.

In Figure 7.1, 'Capacity development' of both educators and stakeholders shifts position. In our initial rough theory, it was triggered by the provision of feedback/consequences from accountability-related activities. Our revisions, based on inferences from our syntheses of evidence, place capacity development as a property of the system that is part of a chain of configurations, in that it serves as a mechanism for triggering the setting of expectations or intended responses from the provision of feedback/consequences. Presumably, it also plays a critical role in the institutionalisation of system norms, although our review does not provide evidence of this, except in the inverse. That is, the lack of development of capacity for all accountability elements was shown to lead to the institutionalisation of undesirable norms, most vividly in terms of high-stakes examinations and inspections.

Figure 7.1: Revised theory
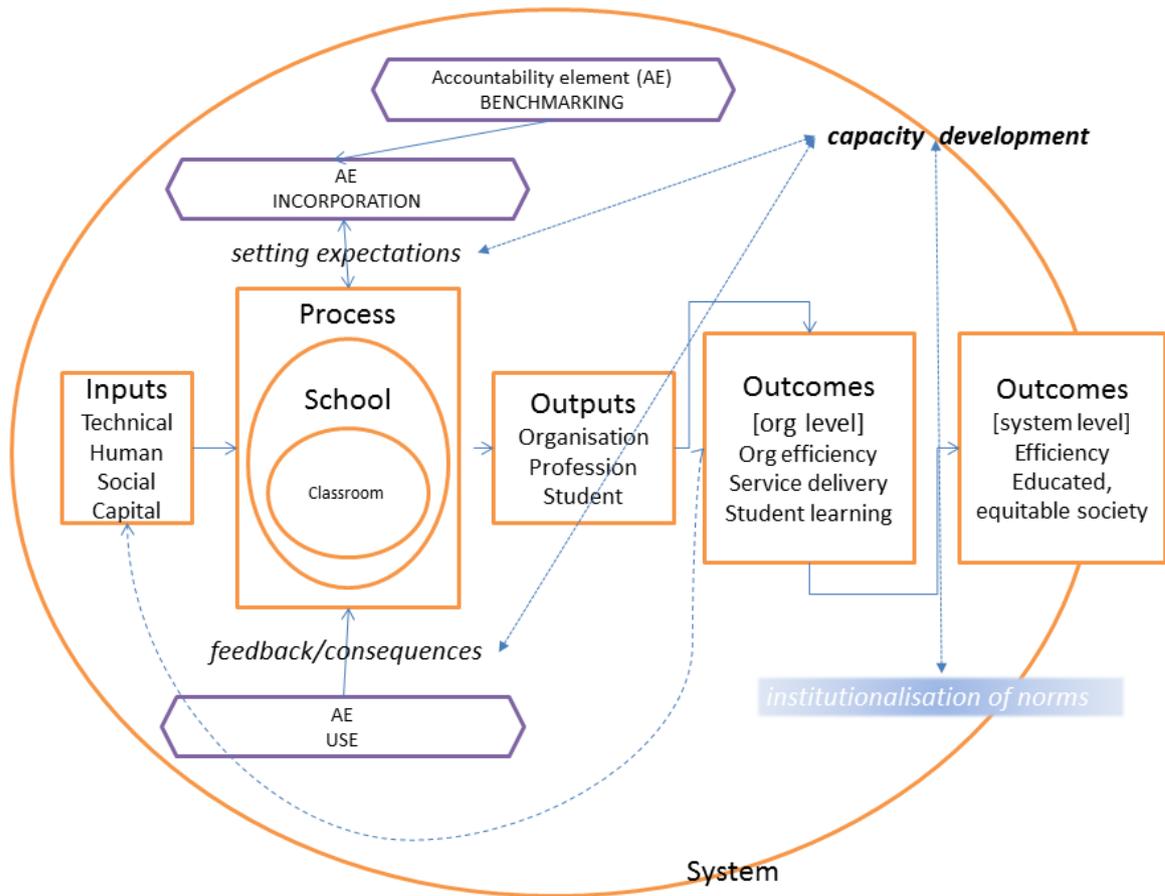
Our proposed mechanisms that have the most secure foundations of evidence are those concerning the production of undesirable outcomes from high-stakes examinations and inspections. These proposed mechanisms suggest a sequence or chain of producing responses by school personnel that may yield unintended outcomes. Our syntheses of evidence in high-stakes examinations and inspections led us to the inference that internal and external pressure for results triggered 'fear of bureaucratic authority', which when unmediated by 'follow-up/follow-through', i.e. credible sources of support and guidance emphasising instructional application and relevance at the school level and other levels, triggers 'lack of motivation to improve' (or in more extreme instances, manipulation and corruption) and encourages responses that seek only to satisfy administrative demands or 'compliance with bureaucratic authority', rather than seek to improve the quality of instructional management, teaching and student learning. The virtuous cycle, which is less well supported by evidence in this review, might be that 'trust in pedagogical authority' when combined with appropriate and relevant follow-up and follow-through yields capacity to improve, which produces improvements and intended outcomes. The ideal causal pathway of capacity development would see system norms shift from the 'answerability' that seeks primarily to satisfy bureaucratic mandates to the 'responsibility' that might animate continuous improvements in schooling (Gregory, 2003). As we note below, the paucity of evidence supporting the virtuous cycle yielding intended outcomes raises questions around whether and in what specific ways the approaches to accountability that we have reviewed here cultivate responsibility at the school level.

Figure 7.2: Virtuous and undesirable cycles
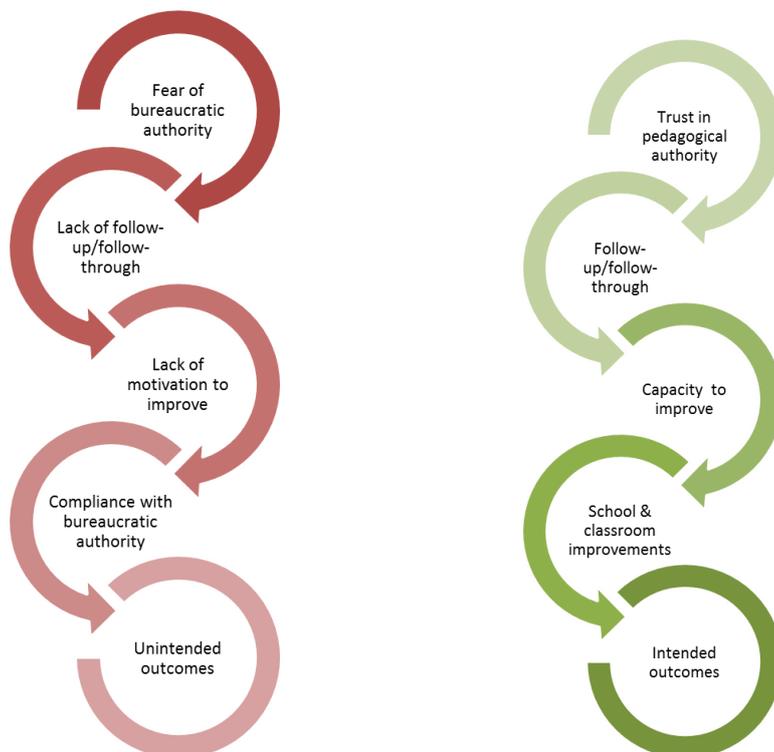
7.3 IMPLICATIONS

Interpretation and application of the results of this review require further work by different users of research. Initial implications include:

POLICY

At a policy level, awareness of the mechanisms we have elaborated may be helpful in assessing the impact of existing initiatives as well as designing new initiatives:

- Assessment, monitoring and inspections may lead to improved outcomes under appropriate conditions in LMICs through the mechanisms of setting expectations and providing feedback. However, capacity development of educators and stakeholders may be an essential component that, if overlooked, can lead to undesirable or unintended consequences. Also, explicit attention may need to be directed towards processes through which the institutionalisation of norms occurs through the system and, locally, in schools.
- Our review indicates that these five mechanisms are interrelated and may need to be considered when explaining how each of the three accountability elements leads to improvement. For example, assessment, monitoring and inspection systems set expectations when developed in close cooperation with key stakeholders (e.g. participation of schools and local community in developing inspection standards, or designing and using school report cards), which equally builds capacity for evaluation and improvement.
- Capacity development within and across levels of the system of education can be an important way of achieving coherence of expectations within schools, and capacity development may need to take place in conjunction with implementation of accountability approaches for intended outcomes to occur. On the other hand, it is worth highlighting that it may be important to consider what kinds of capacities are being developed. Are the capacities geared towards improving teaching and student learning, or are they intended to ensure that school personnel are responsive to compliance with the bureaucracy of schooling?
- There is some evidence in support of participatory approaches, and some evidence that finds that they are of little additional value. It is likely that their use is highly contingent on context.
- There is evidence about the benefits of involving educators in the design, data collection and reporting of accountability activities in all three elements, proposing that such involvement, which varies in degree of participation, promotes the quality of data inputs, active engagement with the results, and use of the results for improvement.
- Interaction among hypothesised mechanisms suggests that there might be mutual dependencies among various approaches to accountability. The interdependence of accountability initiatives is not well documented in the literature, but may be an important consideration for further development of policy.
- Our knowledge of how low-stakes assessment operates and influences school-level decision making is not extensive. This may be because it has been introduced more recently than other accountability activities.

PRACTICE

The concentration of our review on school-level processes and outcomes means that implications for practice at the local level are highlighted. This means emphasis on school management and instructional practices, as well as participatory engagement of local stakeholders. The most salient implications concern ways of resolving what we identified as the tension between 'answerability' and 'responsibility' (Gregory, 2003). It may not be enough for accountability activity to require answerability by placing demands on the local level, especially when those at the local level do not have capacity or resources to fulfil those demands. Higher-level demands may only be fulfilled when they are designed in conjunction with close attention to the needs and capacity of those at the local level. A key insight of this review is the ways in which development of capacity may need to occur within and across levels in order for accountability activity to yield desirable school, system and student outcomes. An enduring question relates to the question of, 'Capacity for what?' Does the capacity demanded by the system aim to enforce compliance or promote the quality of teaching and learning in schools? Must it necessarily be one or the other, or can it be both?

RESEARCH

There is a need for more robust research around what it takes to improve teaching and learning in schools. Our review highlights a paucity of high-quality studies in most areas, particularly in low-stakes assessments, and this may be one of the areas that offers promise of revealing the dynamics of change in schools and in classrooms. Moreover, the connectivity of mechanisms of change suggests that one-dimensional research approaches of looking at the causes and effects of inspections, assessment and monitoring translate with difficulty when trying to explain the impact of accountability in LMICs. Traditional methods of analysing large, longitudinal datasets to link schools' status on accountability measures to improved student achievement results (see for example Allen and Burgess, 2012; Hussain, 2012) are one kind of evidence for research in developing countries but may not provide a complete picture. New research methodologies and approaches need to be developed which specifically look at interlocking mechanisms and conditions of change, examining cyclical cause and effect relations to explain and understand the impact of accountability in different settings. New approaches to conceptions of rigorous research as deeply embedded in continuous improvement of practice within and across levels of the education system may have particular salience in this regard (e.g., Bryk et al., 2015).

Our elaboration of the interdependent mechanisms of accountability activity may offer insight into the most salient relationships to explore in research. Because of the limitations of the literature that we identified, we were unable to conduct comparative analyses across geographic regions or even within regions. Such work could make an important contribution to understanding how systematic variation in historical, social, organisational and cultural contexts shapes responses to system-wide accountability initiatives at the local level if it were structured to examine those contexts in depth. In a related way, intensive studies of the interdependence of accountability initiatives within a single education system would provide valuable insight into how accountability elements operate in concert to produce various outcomes.

8. REFERENCES

8.1 STUDIES INCLUDED IN IN-DEPTH REVIEW (N=214) AND SYNTHESIS* (N=68)

Abadzi H (2009) Instructional time loss in developing countries: Concepts, measurement, and implications. *World Bank Research Observer* 24: 267-283.

Acedo C, Gorostiaga JM, Senen-Gonzalez S (2007) Decentralization and structural change in secondary education in Argentina: The case of the Province of Buenos Aires. *Prospects: Quarterly Review of Comparative Education* 37: 127-145.

Adams D (2012) Introduction: World-class basic education. *Prospects* 42 (3): 247–267.

ADB (2009) *Third Primary Education Development Project: Bangladesh*. Manila: Asian Development Bank.

ADB (2012) *Bangladesh: Third Primary Education Development*. Manila: Asian Development Bank.

*ADEA (2001) *What works and what's new in education: Africa speaks!* Paris: Association for the Development of Education in Africa.

ADEA (2011) *Definitions and properties of African Union indicators for monitoring the plan of action for education in the second decade*. Paris: ADEA.

ADEA (2012) *Monitoring the African Union Second Decade of Education action plan: Outlook report released in Abuja*. Paris: ADEA.

Adegbesan S (2010) Establishing quality assurance in Nigerian education system: Implication for educational managers. *Educational Research and Reviews* 5 (7): 380-384.

Aderemi T (2013) An appraisal of the national economic empowerment and development strategy I (NEEDS I) in Nigeria: The educational perspective. *Journal of Economics and Behavioral Studies* 5 (10): 718-729.

Adeyemi T (2006) The effective management of primary schools in Ekiti State, Nigeria: An analytical assessment. *International Journal of Emotional Psychology and Sport Ethics* 8: 26-33.

AED (2002) *A transnational view of basic education: Issues of access, quality, and community participation in West and Central Africa*. SARA Project. Washington, DC: AED.

Ahmed M (2011) *The sector-wide approach in Bangladesh primary education: A critical view*. CREATE Pathways to Access Research Monograph No. 57. Falmer: University of Sussex. http://www.create-rpc.org/pdf_documents/PTA57.pdf (accessed 3 November 2016).

Akhter M, Chaudhuri J (2013) Real time monitoring for the most vulnerable: Pre-primary education in Bangladesh. *IDS Bulletin* 44: 97-112.

Akporehe D (2011) An assessment of accountability among teachers in secondary schools in Delta State. *African Journal of Social Sciences* 1: 115-125.

Akyeampong K, Pryor J, Ampiah J (2006) *A vision of successful schooling: Ghanaian teachers' understandings of learning, teaching and assessment*. *Comparative Education* 42 (2): 155-176.

*Alcazar L, Rogers FH, Chaudhury N, Hammer J, Kremer M, Muralidharan K (2006) Why are teachers absent? Probing service delivery in Peruvian primary schools. *International Journal of Educational Research* 45: 117-136.

Alejandro JG (2009) *How much are Latin American children learning? Highlights from the Second Regional Student Achievement Test (SERCE)*. Washington, DC: PREAL.

Alvarez J, Moreno VG, Patrinos HA (2007) *Institutional effects as determinants of learning outcomes: Exploring state variations in Mexico*. Policy Research Working Paper Series 4286. Washington, DC: World Bank.

*Andrabi T, Das J, Khwaja AI (2013) *Report cards: The impact of providing school and child test scores on educational markets*. Washington, DC: World Bank.

Arunatilake N, Jayawardena P (2008) *Will formula-based funding and decentralized management improve school level resources in Sri Lanka?* Cahiers de recherche PMMA. PEP-PMMA. http://portal.pep-net.org/documents/download/id/13538 (accessed 3 November 2016).

*Attfield I, Vu BT (2013) A rising tide of primary school standards: The role of data systems in improving equitable access for all to quality education in Vietnam. *International Journal of Educational Development* 33 (1): 74–87.

*Bansilal S (2011) Assessment reform in South Africa: Opening up or closing spaces for teachers? *Educational Studies in Mathematics* 78: 91-107.

*Barr A, Bategeka L, Guloba L, Kasirye I, Mugisha F, Serneels P, Zeitlin A (2012) *Management and motivation in Ugandan primary schools: An impact evaluation report*. PEO. http://ageconsearch.umn.edu/bitstream/164412/2/1357671419-WP_2012-14_PIERI-11857.public.pdf (accessed 3 November 2016).

*Barrera-Osorio F, Raju D (2010) *Short-run learning dynamics under a test-based accountability system: Evidence from Pakistan*. Policy Research Working Paper 5465. Washington, DC: World Bank.

*Barrett A (2011) A Millennium Learning Goal for education post-2015: A question of outcomes or processes. *Comparative Education* 47: 119-133.

Beaumont JA (2007) Investigating Internet use in Jamaican primary classrooms. *Dissertation Abstracts International Section A: Humanities and Social Sciences* 68: 1891.

*Beets P, van Louw T (2011) Social justice implications of South African school assessment practices. *Africa Education Review* 8: 302-317.

Best M, Knight P, Lietz P, Lockwood C, Nugroho D, Tobin, ML (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. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Bisaso R, Kereteletswe O, Selwood I, Visscher A (2008) The use information technology for educational management in Uganda and Botswana. *International Journal of Educational Development* 28: 656-668.

*Braun H, Kanjee A, Bettinger E, Kremer M (2006) *Improving education through assessment, innovation, and evaluation.* Cambridge, MA, American Academy of Arts and Sciences.

Bray M, Adam K, Mark B, Khadeeja A (2001) The dialectic of the international and the national: secondary school examinations in Maldives. *International Journal of Educational Development* 21 (3): 231–244.

*Brock A (2009) Moving mountains stone by stone: reforming rural education in China. *International Journal of Educational Development* 29: 454-462.

Broekman A (2013) The rationale and effects of accountability policies on the work and motivation of teachers: Evidence from Indonesia. In Verger A, Altinyelken H, de Koning M (eds) *Global managerial education reforms and teachers: Emerging policies, controversies and issues in developing contexts*. Amsterdam: American Academy of Arts and Sciences.

Brookings Institution (2012) *A global compact on learning: Taking action on education in developing countries*. Policy Guide. Washington, DC: Brookings Institution.

*Brown G, Hui S, Flora W, Kennedy K (2011) Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Education Research* 50: 307-320.

*Bruns B, Filmer D, Patrinos H (2011) *Making schools work: New evidence on accountability reforms*. Washington DC: World Bank.

*Caddell M (2005) Listening to local voices? International targets and decentralised education planning in Nepal. *International Journal of Educational Development* 25: 456-469.

Cassidy T (2006) *Education Management Information Systems (EMIS) in Latin America and the Caribbean: Lessons and challenges*. Washington, DC: IDB.

*Castro M, Tiezzi S (2003) *The reform of secondary education and the implementation of ENEM in Brazil*. Oxford: Oxford University Press.

*Chen D (2011) *School-based management, school decision-making and education outcomes in Indonesian primary schools*. Washington, DC: World Bank.

*Chisholm L, Wildeman R (2013) The politics of testing in South Africa. *Journal of Curriculum Studies* 45: 89-100.

*Churches R, McBride C (2013) *Making external school review effective*. Reading: CfBT Education Trust.

Cohen J (2001) *The regulatory structure supporting basic education in Indonesia: Analysis covering 1989 to present*. Washington, DC: Academy for Educational Development.

Cohen J, Bloom D, Malin M (2007) *Educating all children: A global agenda*. Cambridge, MA, MIT Press.

Courtney J (2008) Do monitoring and evaluation tools, designed to measure the improvement in the quality of primary education, constrain or enhance educational development? *International Journal of Educational Development* 28: 546-559.

Crouch L, Vinjevold P (2007) South Africa: Access before quality, and what to do now? Revista de currículum y formación del profesorado* 10 (1): 1-16.

Crouch L, Winkler D (2008) *Governance, management and financing of Education for All: Basic frameworks and case studies*. Paris: UNESCO.

Cueto S (2005) *Empirical information and the development of educational policies in Latin America: Regional dialogue on education of the Inter-American Development Bank*. Washington DC: Inter-American Development Bank.

*Darvas P, Balwanz D (2014) *Basic education beyond the Millennium Development Goals in Ghana: How equity in service delivery affects educational and learning outcomes*. Washington DC: World Bank.

Darvas P, Krauss A (2010) *Education in Ghana: Improving equity, efficiency and accountability of education service delivery*. Washington, DC: World Bank.

Davies L, Harber C, Dzimadzi C (2003) Educational decentralisation in Malawi: A study of process. *Compare* 33: 139-154.

De Clercq F (2008) Teacher quality, appraisal and development: The flaws in the IQMS. *Perspectives in Education* 26 (1).

*De Grauwe A (2001) *School supervision in four African countries: Vol. I: Challenges and reforms*. Paris: International Institute for Educational Planning/UNESCO.

De Grauwe A (2005) Improving the quality of education through school-based management: Learning from international experiences. *International Review of Education* 51: 269-287.

*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*. 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.

*De Grauwe A, du Pouget F, Asunción V (2007) *Review of national education planning documents in 45 countries*. Background paper prepared for the Education for All Global Monitoring Report 2007. http://unesdoc.unesco.org/images/0014/001474/147464e.pdf (accessed 3 November 2016).

Dembélé M, Schwille J, Martial D, John S (2006) Can the global trend toward accountability be reconciled with ideals of teacher empowerment? Theory and practice in Guinea. *International Journal of Educational Research* 45 (4–5): 302–314.

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).

Di Gropello E (2003) *Monitoring educational performance in the Caribbean*. Washington DC: World Bank.

Di Gropello E (2004) *Education decentralization and accountability relationships in Latin America*. Washington, DC: World Bank.

Duflo E, Hanna R (2005) *Monitoring works: Getting teachers to come to school*. Cambridge, MA: National Bureau of Economic Research.

Faleye B, Afolabi E (2005) The predictive validity of Osun State Junior Secondary Certificate Examination. *Electronic Journal of Research in Educational Psychology* 3: 131-144.

*Ferrer G (2006) *Educational assessment systems in Latin America: Current practices and future challenges*. Washington, DC: Partnership for Educational Revitalization in the Americas (PREAL).

Firpo S (2012) *Electoral impacts of uncovering public school quality: Evidence from Brazilian municipalities*. Bonn: Institute for the Study of Labor.

Fischman G, Gvirtz S (2001) An overview of educational policies in the countries of Latin America during the 1990s. *Journal of Education Policy* 16 (6): 499-506.

Franco Vivanco E (2013) When change matters: Identifying score gains school determinants in Mexico: An intra-cohort value-added approach. *Economía Mexicana*, *Nueva Epoca* 23 (4): 361-405.

Gershberg A, Gonzalez P (2012) Understanding and improving accountability in education: A conceptual framework and guideposts from three decentralization reform experiences in Latin America. *World Development* 40: 1024–1041.

Gershberg A, Meade B, Andersson S (2009) Providing better education services to the poor: accountability and context in the case of Guatemalan decentralization. *International Journal of Educational Development* 29 (3): 187–200.

Giri R (2005) *The adaptation of language testing models to national testing of school graduates in Nepal: processes, problems and emerging issues*. PhD thesis, Victoria University of Technology.

Glewwe P (2014) *Education policy in developing countries*, Chicago, IL: University of Chicago Press.

*Glewwe P, Ilias N, Kremer M (2010) *Teacher incentives*. Cambridge, MA: NBER. http://www.nber.org/papers/w9671.pdf (accessed 3 November 2016).

Goldman C, Kumar K, Liu Y (2008) *Education and the Asian surge: A comparison of the education systems in India and China*. Occasional Paper. Santa Monica, CA: Center for Asia Pacific Policy.

Gorostiaga J, Ferreira A (2012) Discourses and policies on educational quality in Argentina: 1990-2010. *Research in Comparative and International Education* 7 (3): 364-375.

Govinda R, Bandyopadhyay M (2008) *Access to elementary education in India: Country analytical review*. Falmer: University of Sussex. http://www.create-rpc.org/pdf_documents/India_CAR.pdf (accessed 3 November 2016).

Gunnar E, Filmer D (2007) Autonomy, participation and learning: Findings from Argentine schools, and implications for decentralization. *Education Economics* 15 (1): 103-127.

*Gvirtz S (2002) Curricular reforms in Latin America with special emphasis on the Argentine case. *Comparative Education* 38: 453-469.

*Gvirtz S, Larripa S (2004) National evaluation system in Argentina: Problematic present and uncertain future. *Assessment in Education: Principles, Policy & Practice* 11: 349-364.

Hanushek E, Link S, Woessmann L (2013) Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics* 104: 212-232.

Harber C (2001) *State of transition: Post-apartheid educational reform in South Africa*. Monographs in International Education. Didcot: Symposium Books.

*Harber C (2006) Democracy, development and education: Working with the Gambian inspectorate. *International Journal of Educational Development* 26: 618-630.

Herriot A, Crossley M, Juma M, Waudo J, Mwirotsi M, Kamau A (2002) The development and operation of headteacher support groups in Kenya: A mechanism to create pockets of excellence, improve the provision of quality education and target positive changes in the community. *International Journal of Educational Development* 22: 509-526.

*Herselman M, Hay D (2002) Quality assurance in the Foundation Phase in the Eastern Cape province: A case study. *South African Journal of Education* 22: 239-245.

*Higgins L, Rwanyange R (2005) Ownership in the education reform process in Uganda. *Compare: A Journal of Comparative Education* 35: 7-26.

Holsinger D, Jacob W (2008) *Inequality in education: Comparative and international perspectives*. CERC Studies in Comparative Education, vol. 24. New York: Springer; Hong Kong: Comparative Education Research Centre.

*Howie S (2012) High-stakes testing in South Africa: Friend or foe? *Assessment in Education: Principles, Policy & Practice* 19: 81-98.

Humayun S (2013) Regulating low fee private schools in Islamabad: A study on policy and practice. In Srivastava P (ed.) *Low-fee private schooling: Aggravating equity or mitigating disadvantage?* Didcot: Symposium Books.

Hungi N, Thuku FW (2010) Variations in reading achievement across 14 southern African school systems: Which factors matter? *International Review of Education* 56 (1): 63–101.

Hunt F (2013) *Review of national policies on learning and teaching*. Paris: UNESCO.

Iyengar R, Surianarain S (2010) A comparative analysis of education policy and practice: The case of institutions in Mumbai and Delhi. *Perspectives on Urban Education* 8: 19-28.

*Jaffer K (2010) School inspection and supervision in Pakistan: approaches and issues. *Prospects* 40: 375-392.

Jansen J (2001) On the politics of performance in South African education: Autonomy, accountability and assessment. *Prospects* 31 (4): 553-564.

Kamens D, Benavot A (2011) National, regional and international learning assessments: Trends among developing countries: 1960-2009. *Globalisation, Societies and Education* 9: 285-300.

*Kapambwe W (2010) The implementation of school based continuous assessment (CA) in Zambia. *Educational Research and Reviews* 5: 99-107.

Karim S, Santizo R, Claudia A, Cabrero M (2004) *Transparency in education. Study 1: Report card in Bangladesh; Study 2: Quality Schools Programme in Mexico*. Paris: International Institute for Educational Planning.

*Kellaghan T, Greaney V (2001) *Using assessment to improve the quality of education*. Paris: International Institute for Educational Planning.

*Kellaghan T, Greaney V (2004) *Assessing student learning in Africa*. Washington DC: World Bank.

Khalil S, Salihu A (2011) Delivery of efficient and effective primary education and healthcare services by local governments of Oyo State in Nigeria. *International Journal of Research in Commerce and Management* 2 (10): 8-12.

Khattri N, Ling C, Jha S (2012) The effects of school-based management in the Philippines: An initial assessment using administrative data. *Journal of Development Effectiveness* 4: 277-295.

*Kingdon G, Muzammil M (2012) The school governance environment in Uttar Pradesh, India: Implications for teacher accountability and effort. *Journal of Development Studies* 49: 251-269.

Klenowski V (2006) Learning oriented assessment in the Asia Pacific region. *Assessment in Education: Principles, Policy & Practice* 13: 131-134.

*Kremer M, Miguel E, Thornton R (2004) *Incentives to learn*. Cambridge, MA: NBER.

Kristiansen S, Pratikno, Stein K (2006) Decentralising education in Indonesia. *International Journal of Educational Development* 26 (5): 513–531.

*Lassibille G, Tan J-P, Tesse C, Nguyen T (2010) Managing for results in primary education in Madagascar: Evaluating the impact of selected workflow interventions. *World Bank Economic Review* 24: 303-329.

Leclercq F (2003) Education policy reforms and the quality of the school system: A field study of primary schools in Madhya Pradesh, India. *Economic and Political Weekly* 38 (19): 1855-1869.

Leopold R, Pema L, Syed Rashed Al-Zayed J (2007) *Education for All in Bangladesh: Where does Bangladesh stand in terms of achieving the EFA Goals by 2015?* Washington, DC: World Bank.

Liang LL, Yuan H (2008) Examining the alignment of Chinese national physics curriculum guidelines and 12th-grade exit examinations: A case study. *International Journal of Science Education* 30: 1823-1835.

Little A (2008) *EFA politics, policies and progress*. CREATE Pathways to Access, Research Monograph No. 13. Falmer: CREATE, University of Sussex.

Little A (2010) *Access to elementary education in India: Politics, policies and progress*. CREATE Pathways to Access, Research Monograph No. 44. Falmer: CREATE, University of Sussex.

*Lubisi R, Murphy R (2002) Assessment in South African schools. *Assessment in Education: Principles, Policy & Practice* 9: 255-268.

*Luxia Q (2005) Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing* 22: 142-173.

Machado F (2013) *Decentralization and accountability: The curse of local underdevelopment*. Washington, DC: IDB.

*Macpherson R (2011) Educational administration in Timor Leste: Language policy and capacity building challenges in a post-conflict context. *International Journal of Educational Management* 25: 186-203.

Maile S (2002) Accountability: An essential aspect of school governance. *South African Journal of Education* 22 (4): 326-331.

Malik MR (2007) Improving decision-making systems for decentralized primary education delivery in Pakistan. *Dissertation Abstracts International Section A: Humanities and Social Sciences* 68: 2650.

Marcelo CN, Gabriel B (2008) *The Brazilian Education Quality Index (Ideb): Measurement and incentives upgrades*. Rio de Janeiro: Fundação Getulio Vargas, Escola De Pós-Graduação em Economia.

Marriott N, Goyder H (2009) *Manual for monitoring and evaluating education partnerships*. Paris: IIEP.

Marshall J, Chinna U, Nessay PO, Hok U, Savoeun V, Tinon S, Veasna M (2009) Student achievement and education policy in a period of rapid expansion: assessment data evidence from Cambodia. *International Review of Education* 55: 393-413.

Marshall J, Chinna U, Hok UN, Tinon S, Veasna M, Nissay Put (2012) Student achievement and education system performance in a developing country. *Educational Assessment, Evaluation and Accountability* 24: 113-134.

Martin C, Guzman E, Chris M, Elsa G (2005) Performance evaluation or standardized testing of aptitudes? Innovations at the margins Mexico's of [sic] school system. *International Journal of Educational Development* 25 (2): 145–155.

*Mazibuko SP (2007) The managerial role of the principal in whole-school evaluation in the context of disadvantaged schools in Kwazulu-Natal. Doctor of Education dissertation, University of South Africa.

Mbelle A (2008) *The impact of reforms on the quality of primary education in Tanzania: An assessment of primary education reforms in Tanzania*. Dar es Salaam: Research on Poverty Alleviation (REPOA).

Meade B, Gershberg AI (2008) Making education reform work for the poor: Accountability and decentralization in Latin America. *Journal of Education Policy* 23: 299-322.

Mobegi F, Ondigi A, Oburu P (2010) Secondary school headteachers' quality assurance strategies and challenges in Gucha District, Kenya. *Educational Research and Reviews* 5 (7): 408-414.

*Moswela B (2010) Instructional supervision in Botswana secondary schools: An investigation. *Educational Management Administration & Leadership* 38: 71-87.

Mujeri MK (2010) The rights-based approach to education in Bangladesh. *Bangladesh Development studies* 33: 139-203.

*Mukhopadhyay R, Sriprakash A (2011) Global frameworks, local contingencies: Policy translations and education development in India. *Compare* 41 (3): 311-326.

*Muralidharan K, Sundararaman V (2011) Teacher performance pay: Experimental evidence from India. *Journal of Political Economy* 199: 39-77.

*Murimba S (2005) The impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). *Prospects: Quarterly Review of Comparative Education* 35: 91-108.

Murtin F (2013) *Improving education quality in South Africa*. Paris: OECD Publishing.

Naidoo JP (2005) *Educational decentralization and school governance in South Africa: From policy to practice*. Paris: International Institute for Educational Planning (IIEP).

Ncube N (2007) Managing the quality of education in Zimbabwe: The internal efficiency of rural day secondary schools. *Dissertation Abstracts International Section A: Humanities and Social Sciences*.

Nenty H, Adedoyin O, Odili J, Major T (2007) Primary teacher's perceptions of classroom assessment practices as means of providing quality primary/basic education by Botswana and Nigeria. *Educational Research and Reviews* 2: 74-81.

Ngware M, Oketch M, Ezeh A (2011) Quality of primary education inputs in urban schools: Evidence from Nairobi. *Education and Urban Society* 43: 91-116.

*Nsibande R, Modiba M (2012) 'I Just Do as Expected': Teachers' implementation of continuous assessment and challenges to curriculum literacy. *Research Papers in Education* 27: 629-645.

Nusche D, Santiago P, Shewbridge C, Radinger T (2013*) OECD reviews of evaluation and assessment in education synergies for better learning: An international perspective on evaluation and assessment*. Paris: OECD.

Nwaka N (2010) Secondary school administration in Anambra State today: Challenges and the way forward. *African Research Review* 4: 1-14.

Nzomo J, Kariuki M, Guantai L (2001) *The quality of education: Some policy suggestions based on a survey of schools: Kenya*. SACMEQ Policy Research Report. Paris: IIEP.

OECD (2010) *Improving schools: Strategies for action in Mexico*. Paris: Organisation for Economic Co-operation and Development.

Okkolin M-A, Lehtomaki E, Bhalalusesa E (2010) The successful education sector development in Tanzania: Comment on gender balance and inclusive education. *Gender and Education* 22: 63-71.

*Ong S (2010) Assessment profile of Malaysia: High-stakes external examinations dominate. *Assessment in Education: Principles, Policy & Practice* 17 (1): 91-103.

*Opoku-Asare N (2006) Quality control for effective basic education in Ghana. *Journal of Science and Technology* 26 (3): 106-113.

Ozturgut O (2011) Standardized testing in the case of China and the lessons to be learned for the U.S. *Journal of International Education Research* 7: 1-6.

Pandey P, Goyal S, Sundararaman V (2009) Community participation in public schools: Impact of information campaigns in three Indian states. *Education Economics* 17: 355-375.

Parveen S (2008) An evaluative study of primary education in the light of policies and plans in Pakistan (1947-2006). *Journal of College Teaching & Learning* 5: 17-25.

Patrinos H, Larocque N (2007) *Enhancing accountability in schools: What can choice and contracting contribute*. Washington, DC: World Bank.

Patrinos H, Velez E, Wang C (2013) *Framework for the reform of education systems and planning for quality*. Washington DC: World Bank.

Peng WJ, Thomas S, Yang X, Li J (2006) Developing school evaluation methods to improve the quality of schooling in China: A pilot 'value added' study. *Assessment in Education: Principles, Policy & Practice* 13 (2): 135-154.

Piper B, Korda M (2011) *EGRA Plus: Liberia: Program evaluation report*. Research Triangle Park, NC: RTI International.

*Postlethwaite T (2004) *Monitoring educational achievement*. Paris: IIEP.

*Powell M (2006) *Rethinking education management information systems: Lessons from and options for less developed countries*. InfoDev. http://www.infodev.org/infodev-files/resource/InfodevDocuments_504.pdf (accessed 3 November 2016).

*Prew M, Quaigrain K (2010) Using school performance data to drive school and education district office accountability and improvement: The case of Ghana. *Educational Management Administration & Leadership* 38: 728-744.

*Pryor J, Lubisi C (2002) Reconceptualising educational assessment in South Africa: Testing times for teachers. *International Journal of Educational Development* 22: 673-686.

*Ravela P (2002) *How do Latin America's national systems of educational assessment present their results?* Washington, DC: PREAL.

*Ravela P, Wolfe R, Valverde G, Esquivel JM (2001) *How to proceed with learning assessment in Latin America?* Washington, DC: PREAL.

Ravela P, Arregui P, Valverde G, Wolfe R, Ferrer G, Rizo, FM, Aylwin M, Wolff W (2008) *The educational assessments that Latin America needs*. Washington, DC: PREAL.

Reinikka R, Svensson J (2011) The power of information in public services: Evidence from education in Uganda. *Journal of Public Economics* 95: 956–966.

*Reyneke M, Meyer L, Nel C (2010) School-based assessment: The leash needed to keep the poetic 'unruly pack of hounds' effectively in the hunt for learning outcomes. *South African Journal of Education* 30: 277-292.

Rincon H, Sala M, Sala F, Tarnapol Whitacre P, Lee T (eds) (2005) *Basic education in El Salvador: Consolidating the foundations for quality and equal opportunities*. Washington, DC: Academy for Educational Development.

Rogers F, Vegas E (2009) *No more cutting class? Reducing teacher absence and providing incentives for performance*. Policy Research Working Paper Series: 4847. Washington, DC: World Bank.

Saiwa C (2008) *Exploring the challenges of monitoring systems of teaching and learning in primary schools in Malawi in the context of a decentralised education system*. Zomba: University of Malawi.

*Santiago P, McGregor I, Nusche D, Ravela P, Toledo D (2012) *OECD reviews of evaluation and assessment in education: Mexico 2012*. Paris: Organisation for Economic Co-operation and Development.

Sara R, Rakesh R (2013) *Are our children learning? Literacy and numeracy across East Africa*. Nairobi: UWEZO.

Sasaoka Y, Nishimura M, Yuichi S, Mikiko N (2010) Does universal primary education policy weaken decentralisation? Participation and accountability frameworks in East Africa. *Compare* 40 (1): 79-95.

Sauvageot C, Dias Da Graca P (2007) *Using indicators in planning education for rural people: A practical guide*. Paris: International Institute for Educational Planning, UNESCO.

*Scherman V, Howie S, Bosker R (2011) Constructing benchmarks for monitoring purposes: Evidence from South Africa. *Educational Research and Evaluation* 17: 511-525.

Shalem Y (2003) Do we have a theory of change? Calling change models to account. *Perspectives in Education* 21: 29-49.

Smith MD (2012) School self-evaluation: A Jamaican perspective. *Journal of Third World Studies* 29: 137-153.

Smith WJ, Ngoma-Maema WY (2003) Education for All in South Africa: Developing a national system for quality assurance. *Comparative Education* 39 (3): 345-365.

Somerset A (2011) Strengthening educational quality in developing countries: The role of national examinations and international assessment systems. *Compare: A Journal of Comparative and International Education* 41: 141-144.

Tapo M (2004) *National standards/local implementation: case studies of differing perceptions of national education standards in Papua New Guinea*. PhD thesis, Queensland University of Technology.

*Taylor N (2009) Standards-based accountability in South Africa. *School Effectiveness and School Improvement* 20 (3): 341-356.

Tooley J, Dixon P (2006) 'De facto' privatisation of education and the poor: Implications of a study from Sub-Saharan Africa and India. *Compare: A Journal of Comparative Education* 36: 443-462.

Tshwane University of Technology (2012) *Assessment and education quality in sub-Saharan Africa: Prospects and pitfalls*. Pretoria: Tshwane University of Technology. http://www.oxes.org.uk/wordpress/wp-content/uploads/2013/03/Assessment-and-education-quality-in-sub-Saharan-Africa-presentation-with-timings.pdf (accessed 3 November 2016).

Tuin M, Verger A (2013) Evaluating teachers in Peru: Policy shortfalls and political implications. In Verger A, Altinyelken H, de Koning M (eds) *Global managerial education reforms and teachers: Emerging policies, controversies and issues in developing contexts*. Brussels: Education International Research Institute.

Tung K-C (2006) *EFA mid-decade assessment: Meeting report: Annual EFA Coordinators Meeting/EFA Mid-Decade Assessment Planning Meeting* (7th, Bangkok, Thailand, 24-29 October 2005). Bangkok: UNESCO. http://www.unescobkk.org/fileadmin/user_upload/efa/documents/Final_Meeting_Report.PDF (accessed 3 November 2016).

Umalusi (2009). Report on the quality assurance of the National Senior Certificate Assessment and Examination. Pretoria: Umalusi.

UNESCO (2006) *Situation analysis of teacher education in Pakistan: Towards a strategic framework for teacher education and professional development*. Paris: UNESCO and Washington, DC: USAID.

UNESCO (2007) *Education for All by 2015: Will we make it?* Paris: UNESCO and Oxford: Oxford University Press.

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

*USAID (2006) *School report cards: Some experience*. Washington DC: USAID.

*USAID (2007) *More effective decentralized education: Management and governance*. Washington, DC: USAID. http://pdf.usaid.gov/pdf_docs/Pdacm855.pdf (accessed 3 November 2016).

*Uwazi (2009) *When school inspection doesn't deliver: Highlights from the CAG audit of the Secondary Schools Inspection Programme in Tanzania*. Dar es Salaam: Uwazi InfoShop at Twaweza.

Van Der Berg S, Louw M (2006) Unravelling the mystery: Understanding South African schooling outcomes in regional context. Paper presented at: *Conference of the Centre for the Study of African Economies*, Oxford University, 21 March. http://www.jet.org.za/events/conferences/School%20quality%20research%20seminar%202/VanderBerg-Louw%20SA%20schooling%20outcomes.pdf (accessed 3 November 2016).

Van der Berg S, Shepherd D (2008) Signalling performance: an analysis of continuous assessment and matriculation examination marks in South African schools. Pretoria: Umalusi.

Van Der Gaag J, Adams A (2010) *Where Is the learning? Measuring schooling efforts in developing countries*. Policy Brief 2010-04. Washington DC: Brookings Institution.

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

Verger A, Novelli M, Altinyelken H (2012) *Global education policy and international development: New agendas, issues, and policies*. London: Bloomsbury.

Verger A, Altinyelken H, de Koning M (2013) *Global managerial education reforms and teachers: Emerging policies, controversies and issues in developing context*. Brussels: Education International.

Verspoor AM (2008) The challenge of learning: Improving the quality of basic education in sub-Saharan Africa. In Johnson D (ed.) *The changing landscape of education in Africa: Quality, equality and democracy*. Didcot: Symposium Books.

Vincent G, Thomas K (2012) *Implementing a national assessment of educational achievement.* Washington, DC: The World Bank.

Wagner DA (2011) *Smaller, quicker, cheaper: Improving learning assessments for developing countries*. Paris: IIEP.

Wagner DA, Lockheed M, Mullis I, Martin MO, Kanjee A, Gove A, Dowd AJ (2012) The debate on learning assessments in developing countries. *Compare* 42 (3): 509-545.

Wako T (2003) *Education Management Information Systems (EMIS): A guide for young managers*. Harare: NESIS/UNESCO.

*Wanzare Z (2002) Rethinking school inspection in the third world: The case of Kenya. *Educational Management, Administration & Leadership* 30 (2): 213-229.

Wanzare Z (2012) Instructional supervision in public secondary schools in Kenya. *Educational Management Administration & Leadership* 40 (2): 188-216.

Welmond M, Murphy D, Sjetnan C (2012) *Cameroon: Governance and management in the education sector*. Washington, DC: World Bank.

Winkler D (2004) *Strengthening accountability in public education*. EQUIP2 Policy Brief. Washington, D.C.: Academy for Educational Development.

*Winkler D (2005) *Increasing accountability in education in Parana State, Brazil*. EQUIP2. Washington DC: USAID.

*Winkler D, Herstein J (2005) *Information use and decentralized education: Challenges to the education management information system (EMIS) in the context of decentralisation*. Washington, DC: Academy for Educational Development.

World Bank (2003) *World development report 2004: Making services work for poor people*. Washington, DC: World Bank.

World Bank (2007) *Toward high-quality education in Peru: Standards, accountability, and capacity building: A World Bank country study*. Washington, DC: World Bank.

*World Bank (2008) *Governance, management, and accountability in secondary education in Sub-Saharan Africa*. Washington, DC: World Bank.

*World Bank (2010) *Improving education management in African countries*. Washington, DC: World Bank.

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

World Bank (2012) *Learning outcomes in Thailand: What can we learn from international assessments?* Washington, DC: World Bank.

Wu KB (2001) *Peruvian education at a crossroads: Challenges and opportunities for the 21st century: A World Bank country study*. Washington, DC: World Bank.

Xu Z-Y (2009) On the policy structure and orientation of governing the school choice problem in China's basic education. *US-China Education Review* 6 (6): 1-8.

Yao G, Aka RG (2006) The educational community running the school system: The example of school management committees (COGES). *Prospects: Quarterly Review of Comparative Education* 36: 185-204.

Yuan G (2008) An analysis of national educational assessment policy in the People's Republic of China and the United States. *Dissertation Abstracts International Section A: Humanities and Social Sciences* 68: 4944.

Zeng T, Deng Y, Yang R, Zuo X, Chu Z, Li X (2007) Balanced development of compulsory education: Cornerstone of education equity. *Frontiers of Education in China* 2: 469-493.

8.2 OTHER REFERENCES USED IN THE TEXT OF THE TECHNICAL REPORT

ADEA (2002) Reaching Out, Reaching All. Sustaining Effective Policy and Practice for Education in Africa. Fighting HIV/AIDS. Proceedings of the ADEA, Biennial Meeting, Arusha, Tanzania, October 7-11, 2001. Paris: ADEA.

Allen R, Burgess S (2012) *How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England*. Bristol: CMPO.

Anderson JA (2005) *Accountability in education*. Paris: UNESCO.

Argyris C (1977) Double loop learning in organizations. *Harvard Business Review* 55 (5): 115-125.

Barakat S, Hardman F, Rohwerder B, Rzeszut K (2012) *Low-cost private schools in Afghanistan and Pakistan: what evidence to support sustainable scale up*. Protocol. http://r4d.dfid.gov.uk/pdf/outputs/systematicreviews/Low_cost_schools_2012Barakat_protocol.pdf

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.

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 (2008) *Managing performance: International comparisons*. London: Routledge.

Bryk AS, Gomez LM, Grunow A, LeMahieu PG (2015) *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

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

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?* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3208 [accessed 3 November 2016).

Carrón G, De Grauwe A (2007) *Current issues in supervision: A literature review*. Paris: Unesco, IIEP.

Cheng YC (1993) *Conceptualization and measurement of school effectiveness: An organizational perspective*. Atlanta, GA: AERA.

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.

Department of Education (2010) Evidence-based report: National Senior Certificate (NSC) intervention improvement strategies. Pretoria: Department of Education.

DfID (undated) *Learning achievement: Engaging with evidence: A working evidence paper*. London: DfID

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.

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.

Eisemon TO (1990) Examination Policies to Strengthen Primary Schooling in African Countries. International Journal of Educational Development, 10:69

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 2015).

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

Gregory R (2003) Accountability in modern government. In Peters BG, Pierre J (eds) *Handbook of public administration*. London: Sage, pages 557–568.

Greenhalgh T, Kristjansson E., Robinson V. (2007) Realist *review to understand the efficacy of school feeding programmes.* BMJ, **335**(7625): p. 858-861.

Greenhalgh, J. (2014). Realist synthesis. In P. K. Edwards, J. O'Mahoney, & S. Vincent (Eds.), Studying organizations using critical realism: a practical guide (pp. 264-281). Oxford, United Kingdom: Oxford University Press.

Glewwe P, Ilias N, Kremer M (2003). "Teacher Incentives", National Bureau of Economic Research Working Paper #9671.

Guerrero G, Leon J, Zapata M, Sugimaru C, Cueto S (2012) *What works to improve teacher attendance in developing countries? A systematic review*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=MdlBcLkonuc%3d&tabid=3377 (accessed 17 November 2016).

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.

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.

Hussain, I. (2012). Subjective performance evaluation in the public sector: Evidence from school inspections. (Discussion paper No. CEEDP0135). London: Centre for the Economics of Education, London School of Economics.

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 17 November 2016).

Kellaghan T, Madaus GF, Raczek A (1996) The Use of External Examinations to Improve Student Motivation. Washington, D.C.: American Educational Research Association.

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.

Kingdon K, Aslam M, Rawal S, Das S (2013) *Are contract teachers and para teachers a cost-effective intervention to address teacher shortage and improve learning outcomes?* London: EPPI- Centre, Social Science Research Unit, Institute of Education, University of London http://r4d.dfid.gov.uk/pdf/outputs/systematicreviews/Parateachers2013Kingdon.pdf (accessed 17 November 2016).

Klerks M (2013) The effect of school inspections: A systematic review. Paper presented at: *ORD*, Wageningen, The Netherlands, 20-22 June 2012. http://janbri.nl/wp-content/uploads/2014/12/ORD-paper-2012-Review-Effect-School-Inspections-MKLERKS.pdf (accessed 3 November 2016).

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.

London NA (1997) *A National Strategy for Systems-Wide Curriculum Improvement in Trinidad and Tobago* In Chapman DW, Mahlck LO, and Smulders AEM., eds., From Planning to Action: Government Initiatives for Improving School-Level Practice. Paris: International Institute for Educational Planning; Oxford: Pergamon.

Luginbuhl R, Webbink D, De Wolf IF (2009) Do school inspections improve primary school performance? *Educational Evaluation and Policy Analysis* 31 (3): 221-237.

Lynch U, McGrellis S, Dutschke M, Anderson M, Arnsberger P, Macdonald G (2013) *What is the evidence that the establishment or use of community accountability mechanisms and processes improves inclusive service delivery by governments, donors and NGOs to communities?* EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=8khQy7mGMxw%3d&tabid=3425 (accessed 17 November 2016).

Mehrens WA, Kaminski J (1989) Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice* 8 (1): 14-22.

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://doc.utwente.nl/93480/1/review.pdf (accessed 3November 2016).

Orr D, Westbrook J, Pryor J, Durrani N, Sebba J, Adu-Yeboah C (2013) *What are the impacts and cost-effectiveness of strategies to improve performance of untrained and under-trained teachers in the classroom in developing countries?* London: EPPI-Centre, Social Science Research Centre, Institute of Education, University of London. http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=tmSH1jOjwZA%3d&tabid=3382 (accessed 17 November 2016).

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. https://www.researchgate.net/profile/Gill_Harvey/publication/228855827_Realist_synthesis_an_introduction/links/0fcfd507f0b7cbb2ce000000.pdf (accessed 3 November 2016).

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.

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).

Pravalpruk, K. 1996. "National assessment in Thailand." In: Murphy, P. et al. (Eds.), National assessments. Testing the system. Washington DC: World Bank

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.

Rollnick M, Manyatsi S, Lubben F, and Bradley J (1998) A Model for Studying Gaps in Education: A Swaziland Case Study in the Learningof Science.Ó International Journal of Educational Development 18:453

Rosenkvist MA (2010) *Using student test results for accountability and improvement: A literature review*. OECD Education Working Papers, No. 54. Paris: OECD. http://dx.doi.org/10.1787/5km4htwzbv30-en (accessed 7 June 2014).

Rosenthal L (2004) Do school inspections improve school quality? *Economics of Education Review* 23 (2): 143-151.

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 (1999) *School effectiveness in developed and developing countries: A review of the research evidence*. Washington, DC: World Bank.

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.

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).

Yu G (2007) *Research evidence of school effectiveness in Sub-Saharan Africa*. EdQual Working Paper No. 7. Bristol: University of Bristol. http://www.edqual.org/publications/workingpaper/edqualwp7.pdf (accessed 17 November 2016).

APPENDICES

APPENDIX 1.1: AUTHORSHIP OF THIS REPORT

DETAILS OF ADVISORY GROUP MEMBERSHIP

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

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

None

APPENDIX 2.1: RAMESES PUBLICATION STANDARDS: REALIST SYNTHESIS

TITLE			
1		In the title, identify the document as a realist synthesis or review	Subtitle
ABSTRACT			
2		While acknowledging publication requirements and house style, abstracts should ideally contain brief details of: the study's background, review question or objectives; search strategy; methods of selection, appraisal, analysis and synthesis of sources; main results; and implications for practice.	Executive summary
INTRODUCTION			
3	Rationale for review	Explain why the review is needed and what it is likely to contribute to existing understanding of the topic area.	Chapter 1
4	Objectives and focus of review	State the objective(s) of the review and/or the review question(s). Define and provide a rationale for the focus of the review.	Chapter 1
METHODS			
5	Changes in the review process	Any changes made to the review process that was initially planned should be briefly described and justified.	Chapter 2
6	Rationale for using realist synthesis	Explain why realist synthesis was considered the most appropriate method to use.	Chapter 2
7	Scoping the literature	Describe and justify the initial process of exploratory scoping of the literature.	Chapter 2
8	Searching processes	While considering specific requirements of the journal or other publication outlet, state and provide a rationale for how the iterative searching was done. Provide details on all the sources accessed for information in the review. Where searching in electronic databases has taken place, the details should include, for example, name of database, search terms, dates of coverage and date last searched. If individuals familiar with the relevant literature and/or topic area were contacted, indicate how they were identified and selected.	Chapter 2
9	Selection and appraisal of documents	Explain how judgements were made about including and excluding data from documents, and justify these.	Chapter 2
10	Data extraction	Describe and explain which data or information were extracted from the included documents and justify this selection.	Chapter 2
11	Analysis and synthesis processes	Describe the analysis and synthesis processes in detail. This section should include information on the constructs analyzed and describe the analytic process.	Chapter 2
RESULTS			
12	Document flow diagram	Provide details on the number of documents assessed for eligibility and included in the review with reasons for exclusion at each stage as well as an indication of their source of origin (for example, from searching databases, reference lists and so on). You may	Chapter 3

		consider using the example templates (which are likely to need modification to suit the data) that are provided.	
13	Document characteristics	Provide information on the characteristics of the documents included in the review.	Chapter 3
14	Main findings	Present the key findings with a specific focus on theory building and testing.	Chapter 4,5,6
DISCUSSION			
15	Summary of findings	Summarize the main findings, taking into account the review's objective(s), research question(s), focus and intended audience(s).	Chapter 7
16	Strengths, limitations and future research directions	Discuss both the strengths of the review and its limitations. These should include (but need not be restricted to) (a) consideration of all the steps in the review process and (b) comment on the overall strength of evidence supporting the explanatory insights which emerged. The limitations identified may point to areas where further work is needed.	Chapter 7
17	Comparison with existing literature	Where applicable, compare and contrast the review's findings with the existing literature (for example, other reviews) on the same topic.	Chapter 7
18	Conclusion and recommendations	List the main implications of the findings and place these in the context of other relevant literature. If appropriate, offer recommendations for policy and practice.	Chapter 7
19	Funding	Provide details of funding source (if any) for the review, the role played by the funder (if any) and any conflicts of interests of the reviewers.	Acknowledgement, Chapter 1 and Chapter 2

INCLUSION AND EXCLUSION CRITERIA AT THE MAPPING STAGE

Inclusion criteria	Exclusion criteria
Types of intervention: Studies or reports that investigate or explore accountability (monitoring, assessment and/or, inspection) of education system	Types of intervention: a) Studies or papers DO NOT focus on accountability elements OR b) Studies or papers DO NOT clearly state or make reference to a sub-national, national, regional and/or international level of an assessment, inspection or monitoring programme
Geographical location: Conducted in low- or middle-middle-income countries according to World Bank classification[5]	Geographical location: NOT conducted in low- or lower-middle-income countries according to World Bank classification
Setting: Targeting primary, secondary and/or compulsory education	Setting: NOT designed for primary, secondary and/or compulsory education
Types of studies: All types of study designs, policy and theoretical/conceptual framework documents	Types of studies: No restriction
Language: Published in English	Language: NOT published English
Date: Published in or after 1990	Date: Published before 1990

---

[5] http://data.worldbank.org/about/country-classifications (accessed 15 February 2014

APPENDIX 2.2: SEARCH STRATEGY FOR ELECTRONIC DATABASES

AEI (PROQUEST) 24 APRIL 2014

Set one (Indexed terms 1 AND countries AND settings)

((primary PRE/1 school*) OR (elementary PRE/1 school*) OR (high PRE/1 school*) OR (secondary PRE/1 School*) OR (Secondary PRE/1 Teach*) OR (secondary PRE/1 education) OR (primary PRE/1 education) OR (compulsory PRE/1 education) OR (elementary PRE/1 education) OR (schools) OR (school PRE/1 girl*) OR (school PRE/1 boys) OR (school) OR (schools)) AND ((MJSUB.EXPLODE('Educational quality') OR MJSUB.EXPLODE('Performance indicators') OR MJSUB.EXPLODE('Institutional evaluation') OR MJSUB.EXPLODE('Measurement objectives') OR MJSUB.EXPLODE('Report cards') OR MJSUB.EXPLODE('Management information systems') OR MJSUB.EXPLODE('Performance factors') OR MJSUB.EXPLODE('Recordkeeping') OR SU.EXACT('Government school relationship') OR MJSUB.EXPLODE('Personnel management') OR MJSUB.EXPLODE('Information utilisation') OR MJSUB.EXPLODE('Monitoring (Assessment)') OR MJSUB.EXPLODE('Competency based assessment') OR MJSUB.EXPLODE('Performance tests') OR MJSUB.EXPLODE('Criterion referenced tests') OR MJSUB.EXPLODE('Budgeting') OR MJSUB.EXPLODE('Performance based assessment') OR MJSUB.EXPLODE('Administrator role') OR MJSUB.EXPLODE('Educational indicators') OR SU.EXACT('Alternative assessment') OR SU.EXACT('Educational administration') OR MJSUB.EXPLODE('Records management') OR MJSUB.EXPLODE('Management systems') OR MJSUB.EXPLODE('Audits (Verification)') OR MJSUB.EXPLODE('Educational assessment') OR MJSUB.EXPLODE('Benchmarking') OR SU.EXACT('Bureaucracy') OR MJSUB.EXPLODE('Supervision') OR SU.EXACT('Records (Forms)') OR SU.EXACT('Formative evaluation') OR SU.EXACT('Access to information') OR MJSUB.EXPLODE('Administrative organisation') OR MJSUB.EXPLODE('Supervisors')) AND ((Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR kirghiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda

OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America') OR (developing PRE/1 nation*) OR (developing PRE/1 country*) OR (developing PRE/1 world) OR (developing PRE/1 economy*) OR (less* PRE/1 developed PRE/1 countries) OR (less* PRE/1 developed PRE/1 nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 economy*) OR (underdeveloped PRE/1 country*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 economy*) OR (low* PRE/2 middle PRE/2 country*) OR (LMIC) OR (laics) OR (LLMIC) OR (Llaics) OR (third PRE/1 world) OR (underserved PRE/1 country*) OR (underserved PRE/1 nation*) OR (deprived PRE/1 country*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 country*) OR (poor* PRE/1 nation*)))

Set 2: Free texts

(Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon

Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America') OR (developing PRE/1 nation*) OR (developing PRE/1 countr*) OR (developing PRE/1 world) OR (developing PRE/1 econom*) OR (less* PRE/1 developed PRE/1 countries) OR (less* PRE/1 developed PRE/1 nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 econom*) OR (underdeveloped PRE/1 countr*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 econom*) OR (low* PRE/2 middle PRE/2 countr*) OR (LMIC) OR (LMICs) OR (LLMIC) OR (LLMICs) OR (third PRE/1 world) OR (underserved PRE/1 countr*) OR (underserved PRE/1 nation*) OR (deprived PRE/1 countr*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 countr*) OR (poor* PRE/1 nation*)

AND

Accountability OR benchmarking OR (monitoring PRE/1 activit) OR (monitorin PRE/1 system*) OR (progress PRE/1 monitoring) OR (monitoring PRE/1 mechanism*) OR (monitoring PRE/1 process*) OR (monitoring PRE/1 procedure*) OR (targeted PRE/1 monitoring) OR (inspection*) OR (inspector*) OR (supervis*) OR (EMIS) OR (education PRE/1 management PRE/1 information PRE/1 system) OR (performance PRE/1 review*) OR (financial PRE/1 management) OR (audit*) OR (budget*) OR (education* PRE/1 finance) OR (Total PRE/1 quality PRE/1 management) OR (quality PRE/1 assurance) OR (quality PRE/1 control) OR (information PRE/1 management) OR (database PRE/1 management) OR (information PRE/1 system*) OR (decision PRE/1 support PRE/1 system*) OR (standardised PRE/1 test*) OR (standardized PRE/1 test*) OR (budget PRE/1 tracking) OR (appraisal PRE/1 process*) OR (management PRE/1 education) OR (competency-based PRE/1 education) OR (competency PRE/1 based PRE/1 education) OR (performance PRE/1 based) OR (result* PRE/1 based) OR (outcome-based) OR (outcome PRE/1 based) OR (alternative PRE/1 assessment) OR (curriculum PRE/1 based PRE/1 assessment) OR (curriculum-based PRE/1 assessment) OR (educational PRE/1 assessment) OR (assessment PRE/1 procedure) OR (standardised PRE/1 assessment) OR (standardized PRE/1 assessment) OR (informal PRE/1 assessment) OR (assessment PRE/1 system*) OR (assessment PRE/1 mechanism*) OR (assessment PRE/1 process*) OR (educational PRE/1 quality) OR (performance PRE/1 factor*) OR (performance PRE/1 indicator*) OR (performance PRE/1 management) OR (educational PRE/1 indicator*) OR (performance PRE/1 information) OR (personnel PRE/1 evaluation) OR (program* PRE/1 monitoring) OR (progress PRE/1 reporting) OR (recordkeeping) OR (achievement PRE/1 test*) OR (assessment PRE/1 program*) OR (referenced PRE/1 tests) OR (educational PRE/1 test*) OR (high PRE/1 stakes PRE/1 test*) OR (national PRE/1 test*) OR (international PRE/1 test*) OR (competency PRE/1 test*) OR (competency-based PRE/1 test*) OR (competency PRE/1 assessment) OR (performance PRE/1 test*) OR (standardised PRE/1 assessment) OR (quality PRE/1 review) OR (results-based PRE/1 performance) OR (performance PRE/1 evaluation) OR (information PRE/1 utilization) OR (personnel PRE/1 management) OR (educational PRE/1 management) OR (educational PRE/1 administration) OR (educational PRE/1 environment) OR (educational PRE/1 finance) OR (government PRE/1 regulation) OR (quality PRE/1 assurance) OR (quality PRE/1 control) OR (organizational PRE/1 performance) OR (organizational PRE/1 effective) OR (performance PRE/1 test*) OR (management PRE/1 styles) OR (administrative PRE/1 organization) OR (national PRE/1 competency PRE/1 test*) OR (norm PRE/1 referenced

PRE/1 test*) OR (criterion PRE/1 referenced PRE/1 test*) OR (exit PRE/1 examination) OR (administration PRE/1 effectiveness) OR (administrator PRE/1 education)

 AND

(SU.EXACT.EXPLODE('Primary education') OR MJSUB.EXPLODE('Primary secondary schools') OR MJSUB.EXPLODE('Lower primary years') OR MJSUB.EXPLODE('Primary school curriculum') OR MJSUB.EXPLODE('Secondary schools') OR MJSUB.EXPLODE('Primary schools') OR MJSUB.EXPLODE('Primary school teachers') OR MJSUB.EXPLODE('Technical schools') OR MJSUB.EXPLODE('High school students') OR MJSUB.EXPLODE('School restructuring') OR MJSUB.EXPLODE('Secondary education') OR MJSUB.EXPLODE('Agricultural education') OR MJSUB.EXPLODE('Primary school students') OR MJSUB.EXPLODE('Vocational high schools') OR MJSUB.EXPLODE('Secondary school teachers') OR MJSUB.EXPLODE('Primary secondary education') OR MJSUB.EXPLODE('Secondary school students') OR MJSUB.EXPLODE('School organisation') OR MJSUB.EXPLODE('Compulsory education') OR MJSUB.EXPLODE('High schools') OR MJSUB.EXPLODE('Government schools') OR MJSUB.EXPLODE('Secondary school curriculum')) OR ((primary PRE/1 school*) OR (elementary PRE/1 school*) OR (high PRE/1 school*) OR (secondary PRE/1 School*) OR (Secondary PRE/1 Teach*) OR (secondary PRE/1 education) OR (primary PRE/1 education) OR (compulsory PRE/1 education) OR (elementary PRE/1 education))

BEI (EBSCO 24 APRIL 2014)

Publication Date: 19900101-; Language: English

#	Query
S9	S7 AND S8
S8	Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy

	Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (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 W2 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*)
S7	S3 AND S6
S6	S4 OR S5
S5	(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)
S4	DE 'MALE primary school teachers' OR DE 'PRIMARY education' OR DE 'ENGLISH language -- Study & teaching (Primary)' OR DE 'FIRST grade (Education)' OR DE 'FOURTH grade (Education)' OR DE 'INFANT school education (Great Britain)' OR DE 'LANGUAGE arts (Primary)' OR DE 'MATHEMATICS -- Study & teaching (Primary)' OR DE 'MORAL education (Primary)' OR DE 'NUTRITION -- Study &

teaching (Primary)' OR DE 'PRIMARY school teaching' OR DE 'READINESS for school' OR DE 'SCHOOL supervision, Primary' OR DE 'SCIENCE -- Study & teaching (Primary)' OR DE 'SECOND grade (Education)' OR DE 'SOCIAL studies (Primary)' OR DE 'THIRD grade (Education)' OR DE 'UNIVERSAL Primary Education (Education initiative)' OR DE 'PRIMARY school teachers' OR DE 'INFANT school teachers (Great Britain)' OR DE 'MALE primary school teachers' OR DE 'PRIMARY school teaching' OR DE 'PRIMARY schools' OR DE 'FIRST schools (Great Britain)' OR DE 'FROEBEL schools' OR DE 'INFANT schools (Great Britain)' OR DE 'NATIONAL schools (Ireland)' OR DE 'ELEMENTARY school environment' OR DE 'ELEMENTARY school teachers' OR DE 'KINDERGARTEN teachers' OR DE 'MALE elementary school teachers' OR DE 'PRIMARY school teachers' OR DE 'MALE elementary school teachers' OR DE 'MALE primary school teachers' OR DE 'ELEMENTARY school graduates' OR DE 'ELEMENTARY school principals' OR DE 'ELEMENTARY school teaching' OR DE 'ELEMENTARY school teachers -- Selection & appointment' OR DE 'ELEMENTARY school teachers -- Salaries, etc.' OR DE 'ELEMENTARY school supervision' OR DE 'ELEMENTARY schools -- Administration' OR DE 'SECONDARY school students' OR DE 'HIGH school students' OR DE 'JUNIOR high school students' OR DE 'SECONDARY school supervision' OR DE 'SECONDARY school teachers' OR DE 'SECONDARY education -- Standards' OR DE 'HIGH school teachers' OR DE 'CATHOLIC high school teachers' OR DE 'LGBT high school teachers' OR DE 'MINORITY high school teachers' OR DE 'PREPARATORY school teachers' OR DE 'SEXUAL minority high school teachers' OR DE 'HIGH school principals' OR DE 'JUNIOR high school girls' OR DE 'JUNIOR high school boys' OR DE 'HIGH school boys' OR DE 'JUNIOR high school principals' OR DE 'HIGH school teaching' OR DE 'HIGH school teachers -- Tenure' OR DE 'HIGH school teachers -- Social conditions' OR DE 'HIGH school teachers -- Economic conditions' OR DE 'HIGH school placement test' OR DE 'HIGH schools -- Entrance examinations' OR DE 'HIGH school placement test' OR DE 'INDEPENDENT School Entrance Examination' OR DE 'SECONDARY School Admission Test' OR DE 'SPECIALIZED Science High Schools Admissions Test' OR DE 'LOW-income high school students' OR DE 'JUNIOR high school students -- Economic conditions' OR DE 'INDIAN high school students' OR DE 'HIGH school teachers -- Selection & appointment' OR DE 'HIGH school teachers -- Salaries, etc.' OR DE 'HIGH school students' OR DE 'HIGH school juniors' OR DE 'HIGH school girls' OR DE 'HIGH School Proficiency Test' OR DE 'HIGH schools -- Administration' OR DE 'JUNIOR high school students' OR DE 'JUNIOR high school boys' OR DE 'JUNIOR high school girls' OR DE 'HIGH schools -- Examinations' OR DE 'A-level examinations' OR DE 'ADVANCED supplementary level examinations' OR DE 'INTERNATIONAL General Certificate of Secondary Education' OR DE 'INTERNATIONAL baccalaureate' OR DE 'SECONDARY education' OR DE 'AIDS (Disease) education (Secondary)' OR DE 'AUTOMOBILE driver education (Secondary)' OR DE 'COLLEGE preparation programs' OR DE 'COMMUNICATION -- Study & teaching (Secondary)' OR DE 'COMPREHENSIVE high schools' OR DE 'COUNSELING in secondary education' OR DE 'DANCE -- Study & teaching (Secondary)' OR DE 'DEAF -- Education (Secondary)' OR DE 'EIGHT-Year Study' OR DE 'ELEVENTH grade (Education)' OR DE 'ENDOWED public schools (Great Britain)' OR DE 'ENGLISH language -- Study & teaching (Secondary)' OR DE 'ENGLISH literature -- Study & teaching (Secondary)' OR DE 'EVENING & continuation schools' OR DE 'FINANCIAL management -- Study & teaching (Secondary)' OR DE 'FOLK high schools' OR DE 'GIFTED children -- Education (Secondary)' OR DE 'HEALTH education (Secondary)' OR DE 'HIGH schools -- Postgraduate work' OR DE 'INDUSTRIAL arts -- Study & teaching (Secondary)' OR DE 'INDUSTRIAL management -- Study & teaching (Secondary)' OR DE 'INTERPERSONAL relations -- Study & teaching (Secondary)' OR DE 'LANGUAGE arts (Secondary)' OR DE 'MARKETING -- Study & teaching (Secondary)' OR DE 'MATHEMATICS -- Study & teaching (Secondary)' OR DE 'MORAL education (Secondary)' OR DE 'NINTH grade (Education)' OR DE 'OCCUPATIONS -- Study & teaching (Secondary)' OR DE 'POSTPRIMARY schools' OR DE 'SCIENCE --

	Study & teaching (Secondary)' OR DE 'SECONDARY school supervision' OR DE 'SOCIAL studies (Secondary)' OR DE 'TELEVISION in secondary education' OR DE 'TENTH grade (Education)' OR DE 'THIRTEENTH grade (Education)' OR DE 'TWELFTH grade (Education)' OR DE 'TYPEWRITING -- Study & teaching (Secondary)' OR DE 'JUNIOR high school teaching' OR DE 'MINORITY high school teachers' OR DE 'COMPULSORY education' OR DE 'HOME schooling'
S3	S1 OR S2
S2	Accountability OR benchmarking OR (monitoring W1 activit*) OR (monitoring W1 system*) OR (progress W1 monitoring) OR (monitoring W1 mechanism*) OR (monitoring W1 process*) OR (monitoring W1 procedure*) OR (targeted W1 monitoring) OR (inspection*) OR (inspector*) OR (supervis*) OR (EMIS) OR (education W1 management W1 information W1 system) OR (performance W1 review*) OR (financial W1 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 W1 process*) OR (management W1 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 W1 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*) OR (educational W1 quality) OR (performance W1 factor*) OR (performance W1 indicator*) OR (performance W1 management) OR (educational W1 indicator*) OR (performance W1 information) OR (personnel W1 evaluation) OR (program* W1 monitoring) OR (progress W1 reporting) OR (recordkeeping) OR (achievement W1 test*) OR (assessment W1 program*) OR (referenced W1 tests) OR (educational W1 test*) OR (high W1 stakes W1 test*) OR (national W1 test*) OR (international W1 test*) OR (competency W1 test*) OR (competency-based W1 test*) OR (competency W1 assessment) OR (performance W1 test*) OR (standardised W1 assessment) OR (quality W1 review) OR (results-based W1 performance) OR (performance W1 evaluation) OR (information W1 utilization) OR (personnel W1 management) OR (educational W1 management) OR (educational W1 administration) OR (educational W1 environment) OR (educational W1 finance) OR (government W1 regulation) OR (quality W1 assurance) OR (quality W1 control) OR (organizational W1 performance) OR (organizational W1 effective) OR (performance W1 test*) OR (management W1 styles) OR (administrative W1 organization) OR (national W1 competency W1 test*) OR (norm W1 referenced W1 test*) OR (criterion W1 referenced W1 test*) OR (exit W1 examination) OR (administration W1 effectiveness) OR (administrator W1 education) OR (information W1 utilisation)
S1	DE 'BENCHMARKING (Management)' OR DE 'EDUCATION benchmarking' OR DE 'EDUCATIONAL accountability' OR DE 'NONINSTRUCTIONAL teacher responsibilities' OR DE 'PERFORMANCE contracts in education' OR DE 'SCHOOL responsibility' OR DE 'EDUCATIONAL evaluation -- Utilization' OR DE

'INFORMATION resources' OR DE 'INFORMATION technology' OR DE 'SCHOOL responsibility' OR DE 'SCHOOL environment' OR DE 'CLASSROOM environment' OR DE 'COLLEGE environment' OR DE 'ELEMENTARY school environment' OR DE 'SCHOOL bullying' OR DE 'EDUCATION -- Finance' OR DE 'ADULT education -- Finance' OR DE 'CHURCH schools -- Finance' OR DE 'COLLEGE costs' OR DE 'DANCE -- Study & teaching -- Finance' OR DE 'EARLY childhood education -- Finance' OR DE 'EDUCATION -- Costs' OR DE 'EDUCATIONAL charities' OR DE 'EDUCATIONAL vouchers' OR DE 'FINANCIAL exigency (Education)' OR DE 'GOVERNMENT aid to education' OR DE 'HIGHER education -- Finance' OR DE 'LITERACY programs -- Finance' OR DE 'MEDICAL education -- Finance' OR DE 'PHYSICAL education -- Finance' OR DE 'PRIVATE educational finance' OR DE 'PRIVATE schools -- Finance' OR DE 'SCHOOL bonds' OR DE 'SCHOOL purchasing' OR DE 'SCHOOL taxes' OR DE 'STUDENT activities -- Accounting' OR DE 'STUDENT financial aid' OR DE 'TRAINING -- Finance' OR DE 'TUITION' OR DE 'URBAN education -- Finance' OR DE 'VOCATIONAL education -- Finance' OR DE 'DATABASE management' OR DE 'INFORMATION resources management' OR DE 'SCHOOL records' OR DE 'DIPLOMAS (Education)' OR DE 'STUDENT records' OR DE 'TEACHER'S Report Form' OR DE 'SCHOOL reports' OR DE 'TOTAL quality management in education' OR DE 'EDUCATION benchmarking' OR DE 'TOTAL quality management in higher education' OR DE 'EDUCATION benchmarking' OR DE 'CORRECTIVE action (School management)' OR DE 'EDUCATIONAL quality' OR DE 'EXAMINATIONS' OR DE 'ACHIEVEMENT tests' OR DE 'ADAPTED examinations' OR DE 'BATTERIES (Examinations)' OR DE 'COMPUTER literacy -- Examinations' OR DE 'CULTURE-fair tests' OR DE 'DOMAIN referenced tests' OR DE 'DRUG use testing' OR DE 'EQUIVALENCY tests' OR DE 'ESSAY tests' OR DE 'EXAM questions' OR DE 'EXAMINATION answer keys' OR DE 'EXAMINATION answer sheets' OR DE 'EXAMINATIONS -- Design & construction' OR DE 'HIGH schools -- Examinations' OR DE 'HIGHER grade examinations (Scotland)' OR DE 'INTELLIGENCE tests' OR DE 'MATHEMATICS examinations' OR DE 'MCCARTHY Scales of Children's Abilities' OR DE 'MEDICINE -- Examinations' OR DE 'MOTOR ability testing' OR DE 'MULTIPLE choice examinations' OR DE 'NATIONAL teacher examinations' OR DE 'NORM-referenced tests' OR DE 'OBJECTIVE tests' OR DE 'ORAL examinations' OR DE 'PRE-tests & post-tests' OR DE 'PROFESSIONAL licensure examinations' OR DE 'PROGNOSTIC tests' OR DE 'PSYCHOLOGICAL tests' OR DE 'RESPONSE styles (Examinations)' OR DE 'SITUATIONAL tests' OR DE 'TAKE-home examinations' OR DE 'TEACHER competency examinations' OR DE 'TEST interpretation' OR DE 'TEST scoring' OR DE 'TEST validity' OR DE 'TRUE-false examinations' OR DE 'UNIVERSITIES & colleges -- Examinations' OR DE 'OBJECTIVE tests' OR DE 'SCALED Curriculum Achievement Levels Tests' OR DE 'VOCABULARY tests' OR DE 'ACADEMIC achievement -- Testing' OR DE 'NATIONAL norms (Education)' OR DE 'SUMMATIVE tests' OR DE 'COMPETENCY-based teacher education' OR DE 'COMPETENCY tests (Education)' OR DE 'COMPUTER adaptive testing' OR DE 'NATIONAL competency-based educational tests' OR DE 'OUTCOME-based education' OR DE 'COMPETENCY tests (Education)' OR DE 'COMPETENCY-based teacher education' OR DE 'CRITERION referenced tests' OR DE 'SCHOOL management teams' OR DE 'SCHOOL administrators' OR DE 'ART school directors' OR DE 'ASSISTANT school principals' OR DE 'COLLEGE administrators' OR DE 'COORDINATORS (Human services)' OR DE 'HIGH school department heads' OR DE 'MINORITY school administrators' OR DE 'PARENT-administrator relationships' OR DE 'PRIVATE school administrators' OR DE 'SCHOOL admission officers' OR DE 'SCHOOL board members' OR DE 'SCHOOL business administrators' OR DE 'SCHOOL directors' OR DE 'SCHOOL principals' OR DE 'SCHOOL superintendents' OR DE 'SCHOOL supervisors' OR DE 'SPECIAL education administrators' OR DE 'WOMEN school administrators' OR DE 'SCHOOL supervisors' OR DE 'STUDENT-administrator relationships' OR DE 'TEACHER-administrator relationships' OR DE 'EXAMINATIONS' OR DE 'VALUE-added assessment (Education)' OR DE

'OUTCOME assessment (Education)' OR DE 'EMPLOYEES -- Rating of' OR DE 'COUNSELORS -- Evaluation' OR DE 'MERIT ratings' OR DE 'PEER review (Professional performance)' OR DE 'SELF-evaluation' OR DE 'TEACHERS -- Rating of' OR DE 'EDUCATIONAL evaluation' OR DE 'ABILITY grouping (Education) -- Evaluation' OR DE 'AGRICULTURAL education -- Evaluation' OR DE 'ALTERNATIVE assessment (Education)' OR DE 'ALTERNATIVE education -- Evaluation' OR DE 'BASIC education -- Evaluation' OR DE 'BILINGUAL education -- Evaluation' OR DE 'BLIND -- Education -- Evaluation' OR DE 'CURRICULUM evaluation' OR DE 'DEAF -- Education -- Evaluation' OR DE 'EDUCATIONAL productivity' OR DE 'FAMILY life education -- Evaluation' OR DE 'GENERAL education -- Evaluation' OR DE 'GIFTED & talented education -- Evaluation' OR DE 'INDEPENDENT study -- Evaluation' OR DE 'INTERNATIONAL education -- Evaluation' OR DE 'MAINSTREAMING in education -- Evaluation' OR DE 'MORAL education -- Evaluation' OR DE 'PEOPLE with visual disabilities -- Education -- Evaluation' OR DE 'PRISONERS -- Education -- Evaluation' OR DE 'PROGRESSIVE education -- Evaluation' OR DE 'RELIGIOUS education -- Evaluation' OR DE 'SCHOOL field trips -- Evaluation' OR DE 'STUDENTS -- Rating of' OR DE 'TEACHERS -- Rating of' OR DE 'TEACHING aids & devices -- Evaluation' OR DE 'TRADITIONAL assessment (Education)' OR DE 'CURRICULUM-based assessment' OR DE 'ALTERNATIVE assessment (Education)' OR DE 'AUTHENTIC assessment' OR DE 'PORTFOLIO assessment (Education)' OR DE 'ACHIEVEMENT tests' OR DE 'ACCUPLACER (Achievement test)' OR DE 'ACT Assessment' OR DE 'BASIC Achievement Skills Individual Screener (Test)' OR DE 'CALIFORNIA Basic Educational Skills Test' OR DE 'COLLEGE Level Academic Skills Test' OR DE 'COURTIS Standard Tests' OR DE 'CRITERION referenced tests' OR DE 'DANTES Subject Standardized Tests' OR DE 'DOMAIN referenced tests' OR DE 'EQUIVALENCY tests' OR DE 'FLORIDA Comprehensive Assessment Test' OR DE 'FLORIDA State Student Assessment Test II' OR DE 'GED tests' OR DE 'GEORGIA High School Graduation Test' OR DE 'GOLDEN State Examination' OR DE 'HAMMILL Multiability Achievement Test' OR DE 'HIGH School Proficiency Test' OR DE 'IOWA Tests of Basic Skills' OR DE 'KAUFMAN Test of Educational Achievement' OR DE 'MASSACHUSETTS Comprehensive Assessment System' OR DE 'METROPOLITAN Achievement Tests' OR DE 'MISSOURI Mastery & Achievement Tests' OR DE 'NATIONAL competency-based educational tests' OR DE 'NORM-referenced tests' OR DE 'NORTH Carolina Competency Test' OR DE 'OBJECTIVE tests' OR DE 'OHIO Graduation Test' OR DE 'OHIO Proficiency Test' OR DE 'PEABODY Individual Achievement Test-Revised' OR DE 'PSAT (Educational test)' OR DE 'REGENTS high school examinations (New York)' OR DE 'SCALED Curriculum Achievement Levels Tests' OR DE 'TESTS of Achievement & Proficiency' OR DE 'TEXAS Assessment of Academic Skills' OR DE 'TEXAS Educational Assessment of Minimum Skills' OR DE 'VIRGINIA Standards of Learning Tests' OR DE 'VOCABULARY tests' OR DE 'WASHINGTON Assessment of Student Learning' OR DE 'WECHSLER Individual Achievement Test' OR DE 'TESTS of Achievement & Proficiency' OR DE 'STANDARDIZED tests' OR DE 'BATTERIES (Examinations)' OR DE 'COMPREHENSIVE Ability Battery (Test)' OR DE 'EXIT examinations' OR DE 'ILLINOIS Standards Achievement Tests' OR DE 'INDIANA Statewide Testing for Educational Progress' OR DE 'NORM-referenced tests' OR DE 'EDUCATIONAL tests & measurements' OR DE 'ABILITY testing' OR DE 'ACADEMIC achievement -- Testing' OR DE 'ACHIEVEMENT tests' OR DE 'BASELINE assessment (Education)' OR DE 'CHILDREN'S Skills Test' OR DE 'COLLEGE Major Interest Inventory' OR DE 'COLLEGE entrance examinations' OR DE 'COMPETENCY tests (Education)' OR DE 'COMPOSITION (Language arts) tests' OR DE 'COMPREHENSIVE examinations' OR DE 'COMPUTER assisted testing (Education)' OR DE 'DETROIT tests of learning aptitude' OR DE 'DIAGNOSTIC tests (Education)' OR DE 'DOMAIN referenced tests' OR DE 'DYNAMIC assessment (Education)' OR DE 'EARLY Learning Skills Analysis' OR DE 'ELEVEN plus (Educational test)' OR DE 'FORMATIVE tests' OR DE 'GRADING & marking (Students)' OR DE 'GUESSING (Educational tests & measurements)' OR DE

'HIGH school placement test' OR DE 'HIGH-stakes tests' OR DE 'INDIVIDUAL educational tests & measurements' OR DE 'IOWA Tests of Educational Development' OR DE 'ITEM response theory' OR DE 'LANGUAGE & languages -- Examinations' OR DE 'LISTENING comprehension tests' OR DE 'MIDTERM examinations' OR DE 'MILLER-Yoder Language Comprehension Test' OR DE 'NATIONAL Spanish Examinations' OR DE 'NORM-referenced tests' OR DE 'O-level examinations' OR DE 'OHIO Tests of Articulation & Perception of Sounds' OR DE 'OPEN-book examinations' OR DE 'PERSONALITY tests' OR DE 'PLACEMENT testing' OR DE 'PREDICTIVE tests' OR DE 'PREDICTIVE validity' OR DE 'PRESCHOOL tests' OR DE 'PRISM (Educational test)' OR DE 'RECOGNITION of prior learning' OR DE 'SCIENCE -- Examinations, questions, etc.' OR DE 'SCOTTISH Certificate of Education' OR DE 'SEXISM in educational tests' OR DE 'SPELLING ability testing' OR DE 'STANDARD Assessment Tasks (Great Britain)' OR DE 'STANDARD Grade Examinations (Scotland)' OR DE 'STANDARDIZED tests' OR DE 'STUDENT Talent & Risk Profile' OR DE 'SUMMATIVE tests' OR DE 'TEST bias' OR DE 'TEST of Auditory Reasoning & Processing Skills' OR DE 'TEST scoring' OR DE 'TIMED tests (Education)' OR DE 'VERBAL ability -- Evaluation' OR DE 'WISCONSIN tests of testimony & reasoning assessment' OR DE 'WOODCOCK Reading Mastery Tests' OR DE 'WOODCOCK-Munoz Language Survey' OR DE 'EXIT examinations' OR DE 'NORM-referenced tests' OR DE 'CRITERION referenced tests' OR DE 'EDUCATIONAL testing services' OR DE 'HIGH-stakes tests' OR DE 'NATIONAL competency-based educational tests' OR DE 'TRADITIONAL assessment (Education)' OR DE 'STANDARDIZED tests' OR DE 'EDUCATIONAL indicators' OR DE 'SCHOOL supervision' OR DE 'SCHOOL music supervision' OR DE 'STATE supervision of teaching' OR DE 'STUDENT teachers -- Supervision of' OR DE 'TEACHERS -- Supervision of' OR DE 'STATE supervision of teaching' OR DE 'SCHOOL inspections (Educational quality)' OR DE 'SECONDARY school supervision' OR DE 'SCHOOL supervision, Primary' OR DE 'SCHOOL music supervision' OR DE 'RURAL school supervision' OR DE 'ELEMENTARY school supervision' OR DE 'PRACTICUM supervision' OR DE 'AUTHENTIC assessment' OR DE 'ALTERNATIVE assessment (Education)' OR DE 'INSTITUTIONAL autonomy' OR DE 'ON-site evaluation' OR DE 'OBSERVATION (Educational method)' OR DE 'DIFFERENTIATED supervision (Education)' OR DE 'INSPECTION & review' OR DE 'SCHOOLS -- Accounting' OR DE 'EDUCATION -- Costs' OR DE 'SCHOOL budgets' OR DE 'STUDENT activities -- Accounting' OR DE 'PARENT participation in school administration' OR DE 'TEACHER participation in administration' OR DE 'STUDENT participation in administration' OR DE 'STUDENT government' OR DE 'SCHOOL personnel management' OR DE 'COLLEGE personnel management' OR DE 'PRINCIPAL-superintendent relationships' OR DE 'TEACHER development' OR DE 'TEACHER-administrator relationships' OR DE 'TEACHER-principal relationships' OR DE 'TEACHER-school board relationships' OR DE 'TEACHER-superintendent relationships' OR DE 'URBAN schools -- Administration' OR DE 'RURAL schools -- Administration' OR DE 'PRIVATE schools -- Administration' OR DE 'SCHOOL administration teachers' OR DE 'HIGH schools -- Administration' OR DE 'SCHOOL administration -- Decision making' OR DE 'SCHOOL employees' OR DE 'SCHOOL administration' OR DE 'ABILITY grouping (Education)' OR DE 'ADULT education administration' OR DE 'AGRICULTURAL colleges -- Administration' OR DE 'AGRICULTURAL high schools -- Administration' OR DE 'ART schools -- Administration' OR DE 'ARTICULATION (Education)' OR DE 'BOARDING schools -- Administration' OR DE 'BUSINESS schools -- Administration' OR DE 'CATHOLIC high schools -- Administration' OR DE 'CATHOLIC schools -- Administration' OR DE 'CATHOLIC universities & colleges -- Administration' OR DE 'CHARTER schools -- Administration' OR DE 'CHURCH schools -- Administration' OR DE 'CLASS size' OR DE 'CLASSROOM management' OR DE 'COMMUNITY schools -- Administration' OR DE 'CONTINUING education administration' OR DE 'COOPERATIVE education administration' OR DE 'CORRECTIVE action (School management)' OR DE 'COUNTY school systems' OR

DE 'DANCE schools -- Administration' OR DE 'DISTANCE education' OR DE 'DISTANCE education administration' OR DE 'DORMITORIES -- Management' OR DE 'EDUCATIONAL acceleration' OR DE 'EDUCATIONAL counseling -- Administration' OR DE 'ELEMENTARY schools -- Administration' OR DE 'FREE schools -- Administration' OR DE 'HEALTH education administration' OR DE 'HIGH schools -- Administration' OR DE 'INTERNATIONAL schools -- Administration' OR DE 'INTERSCHOOL cooperation' OR DE 'JEWISH day schools -- Administration' OR DE 'JEWISH religious schools -- Administration' OR DE 'JUNIOR high schools -- Administration' OR DE 'LAW schools -- Administration' OR DE 'LIBRARY schools -- Administration' OR DE 'MEDICAL schools -- Administration' OR DE 'MIDDLE school libraries -- Administration' OR DE 'MONITORIAL system of education' OR DE 'NEW schools' OR DE 'PERFORMANCE contracts in education' OR DE 'PHARMACY colleges -- Administration' OR DE 'PRESCHOOLS -- Administration' OR DE 'PRIVATE schools -- Administration' OR DE 'PRIVATE schools -- Business management' OR DE 'PRIVATE universities & colleges -- Administration' OR DE 'PUBLIC schools -- Administration' OR DE 'PUBLIC schools -- Business management' OR DE 'RACE relations in school management' OR DE 'RESEARCH & instruction units (Education)' OR DE 'RURAL schools -- Administration' OR DE 'SCHOOL administrators' OR DE 'SCHOOL attendance' OR DE 'SCHOOL boards' OR DE 'SCHOOL boards -- Management' OR DE 'SCHOOL centralization' OR DE 'SCHOOL city, state, etc.' OR DE 'SCHOOL closings' OR DE 'SCHOOL credits' OR DE 'SCHOOL crisis management' OR DE 'SCHOOL decentralization' OR DE 'SCHOOL discipline' OR DE 'SCHOOL district size' OR DE 'SCHOOL districts' OR DE 'SCHOOL enrollment -- Management' OR DE 'SCHOOL entrance requirements' OR DE 'SCHOOL improvement programs' OR DE 'SCHOOL management teams' OR DE 'SCHOOL personnel management' OR DE 'SCHOOL plant management' OR DE 'SCHOOL restructuring' OR DE 'SCHOOL risk management' OR DE 'SCHOOL rules & regulations' OR DE 'SCHOOL schedules' OR DE 'SCHOOL size' OR DE 'SCHOOL supervision' OR DE 'SCHOOL-based management' OR DE 'SCHOOLS -- Admission' OR DE 'SCHOOLS for people with mental disabilities -- Administration' OR DE 'SCHOOLS of social work -- Administration' OR DE 'SECRETARIATS (Education)' OR DE 'SPECIAL education administration' OR DE 'SPECIAL education schools -- Administration' OR DE 'STATE boards of education' OR DE 'STATE departments of education' OR DE 'STATE universities & colleges -- Administration' OR DE 'STUDENT-administrator relationships' OR DE 'SUMMER schools -- Administration' OR DE 'TEACHER influence' OR DE 'TEACHER participation in administration' OR DE 'TEACHERS -- Selection & appointment' OR DE 'THEOLOGICAL seminaries -- Administration' OR DE 'TOTAL quality management in education' OR DE 'TRANSFER of students' OR DE 'URBAN schools -- Administration' OR DE 'VIDEO recording in school management & organization' OR DE 'VOCATIONAL education -- Administration' OR DE 'VOCATIONAL schools -- Administration' OR DE 'SCHOOL inspections (Educational quality)' OR DE 'SCHOOL inspectors (Educational quality)' OR DE 'SCHOOL boards' OR DE 'CITIZENS' advisory committees in education' OR DE 'PRIVATE school trustees' OR DE 'SCHOOL board-superintendent relationships' OR DE 'SCHOOL boards -- Government policy' OR DE 'SCHOOL boards -- Management' OR DE 'SCHOOL budgets' OR DE 'SCHOOL autonomy' OR DE 'SCHOOL-based management' OR DE 'INSTITUTIONAL autonomy'

ERIC (EBSCO) (05 APRIL 2014)

S13 = (S11 OR S12)

S12 = (S9 AND S10 AND S1)

S11= (S8 AND S1)

S10 = (S6 OR S7)

S9 = (S4 OR S5)

S8 = (S2 OR S3)

S7= 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'

S6= (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)

S5= DE 'Accountability' OR DE 'Benchmarking' OR DE 'Educational Administration' OR DE 'School Administration' OR DE 'Educational Environment' OR DE 'Educational Finance' OR DE 'Educational Quality' OR DE 'Government Role' OR DE 'Database Management Systems' OR DE 'Decision Support Systems' OR DE 'Management Systems' OR DE 'Information Management' OR DE 'Knowledge Management' OR DE 'Personnel Evaluation' OR DE 'Records (Forms)' OR DE 'Information Management' OR DE 'Recordkeeping' OR DE 'Quality Assurance' OR DE 'Total Quality Management' OR DE 'Achievement Tests' OR DE 'Criterion Referenced Tests' OR DE 'High Stakes Tests' OR DE 'Exit Examinations' OR DE 'Personnel Management' OR DE 'Competency Based Education' OR DE 'Performance' OR DE 'Performance Factors' OR DE 'Competency Based Teacher Education' OR DE 'Administrative Organization' OR DE 'Administrator Education' OR DE 'Administrator Effectiveness' OR DE 'Administrator Evaluation' OR DE 'Information Utilization' OR DE 'Performance Based Assessment' OR DE 'Educational Assessment' OR DE 'Curriculum Based Assessment' OR DE 'Assessment Centers (Personnel)' OR DE 'Performance Tests' OR DE 'National Competency Tests' OR DE 'Standardized Tests' OR DE 'Norm Referenced Tests' OR DE 'Progress Monitoring' OR DE 'Educational Indicators' OR DE 'Supervision' OR DE 'Audits (Verification)' OR DE 'Financial Audits' OR DE 'Budgeting' OR DE 'Program Budgeting' OR DE 'Outcome Based Education' OR DE 'Informal Assessment' OR DE 'Institutional Evaluation' OR DE 'Inspection' OR DE 'Quality Control' OR DE 'Practicum Supervision' OR DE 'Teacher Supervision'

S4= Accountability OR benchmarking OR (monitoring W1 activit*) OR (monitoring W1 system*) OR (progress W1 monitoring) OR (monitoring W1 mechanism*) OR (monitoring W1 process*) OR (monitoring W1 procedure*) OR (targeted W1 monitoring) OR (inspection*) OR (inspector*) OR (supervis*) OR (EMIS) OR (education W1 management W1 information W1 system) OR (performance W1 review*) OR (financial W1 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 W1 process*) OR (management W1 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 W1 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*) OR (educational W1 quality) OR (performance W1 factor*) OR (performance W1 indicator*) OR (performance W1 management) OR (educational W1 indicator*) OR (performance W1 information) OR (personnel W1 evaluation) OR (program* W1 monitoring) OR (progress W1 reporting) OR (recordkeeping) OR (achievement W1 test*) OR (assessment W1 program*) OR (referenced W1 tests) OR (educational W1 test*) OR (high W1 stakes W1 test*) OR (national W1 test*) OR (international W1 test*) OR (competency W1 test*) OR (competency-based W1 test*) OR (competency W1 assessment) OR (performance W1 test*) OR (standardised W1 assessment) OR (quality W1 review) OR (results-based W1 performance) OR (performance W1 evaluation) OR (information W1 utilization) OR (personnel W1 management) OR (educational W1 management) OR (educational W1 administration) OR (educational W1 environment) OR (educational W1 finance) OR (government W1 regulation) OR (quality W1 assurance) OR (quality W1 control) OR (organizational W1 performance) OR (organizational W1 effective) OR (performance W1 test*) OR (management W1 styles) OR (administrative W1 organization) OR (national W1 competency W1 test*) OR (norm W1 referenced W1 test*) OR (criterion W1 referenced W1 test*) OR (exit W1 examination) OR (administration W1 effectiveness) OR (administrator W1 education) OR (information W1 utilisation)

S3= DE 'School Accounting' OR DE 'School Administration' OR DE 'School Based Management' OR DE 'School Culture' OR DE 'School Effectiveness' OR DE 'School Organization' OR DE 'School Personnel' OR DE 'School Supervision' OR DE 'Boards of Education' OR DE 'Government School Relationship' OR DE 'School Visitation'

S2 = (school W1 evaluation) OR (school W1 efficiency) OR (school W1 governance) OR (school W1 autonomy) OR (school W1 self-evaluation) OR (school W1 self W1 evaluation) OR (school W1 accounting) OR (school-based W1 management) OR (school W1 based W1 management) OR (school W1 monitoring) OR (school W1 assessment) OR (primary W1 education W1 monitoring) OR (primary W1 education W1 assessment) OR (elementary W1 education W1 monitoring) OR (secondary W1 education W1 assessment) OR (secondary W1 education W1 monitoring) OR (school W1 effectiveness) OR (school W1 organization) OR (school W1 administration) OR (school W1 performance) OR (school W1 card*) OR (school W1 record) OR (school W1 report*) OR (school W1 management) OR (school W1 supervision) OR (school W1 inspection) OR (school W1 accountability) OR (school W1 audit*) OR (school W1 board) OR (school W1 environment)

S1 = Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR

Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (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 W2 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*)

SOCIAL SCIENCES CITATION INDEX, (WEB OF SCIENCE) (24 APRIL 2014)

12 -1,142 #10 AND #4

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

11-512 #10 AND #7

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

10- 301,054 #9 OR #8

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

9- 38,795

(TS = (developing NEAR/1 nation*) OR TS = (developing NEAR/1 countr*) OR TS = (developing NEAR/1 world) OR TS = (developing NEAR/1 econom*) OR TS = (less* NEAR/ 1 developed NEAR/1 countries) OR TS = (less*

NEAR/1 developed NEAR/1 nation*) OR TS = (less* NEAR/1 developed NEAR/1 world) OR TS = (less* NEAR/1 developed NEAR/ 1 econom*) OR TS = (underdeveloped NEAR/1 countr*) OR TS = (underdeveloped NEAR/ 1 nation*) OR TS = (underdeveloped NEAR/1 world) OR TS = (underdeveloped NEAR/1 economies) OR TS = (under NEAR/1 developed NEAR/1 nation*) OR TS = (under NEAR/1 developed NEAR/1 world) OR TS = (under NEAR/1 developed NEAR/1 economies) OR TS = (low* NEAR/1 income NEAR/1 countries) OR TS = (low* NEAR/1 income NEAR/1 nation*) OR TS = (low* NEAR/1 income NEAR/1 econom*) OR TS = (low* NEAR/1 middle NEAR/1 countr*) OR TS = (LMIC) OR TS = (LMICs) OR TS = (LLMIC) OR TS = (LLMICs) OR TS = (third NEAR/1 world) OR TS = (underserved NEAR/1 countr*) OR TS = (underserved NEAR/ 1 nation*) OR TS = (deprived NEAR/1 countr*) OR TS = (deprived NEAR/1 nation*) OR TS = (deprived NEAR/1 world) OR TS = (poor* NEAR/1 countr*) OR TS = (poor* NEAR/1 nation*)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

8-283,272

(TS = Afghanistan OR TS = Albania OR TS = Algeria OR TS = Angola OR TS = Antigua OR TS = Barbuda OR TS = Argentina OR TS = Armenia OR TS = Armenian OR TS = Aruba OR TS = Azerbaijan OR TS = Bahrain OR TS = Bangladesh OR TS = Barbados OR TS = Benin OR TS = Belarus OR TS = Byelorussian OR TS = Belarus OR TS = Belorussian OR TS = Belorussia OR TS = Belize OR TS = Bhutan OR TS = Bolivia OR TS = Bosnia OR TS = Herzegovina OR TS = Herzegovina OR TS = Botswana OR TS = Brazil OR TS = Bulgaria OR TS = 'Burkina Faso' OR TS = 'Upper Volta' OR TS = Burundi OR TS = Cambodia OR TS = 'Khmer Republic' OR TS = Kampuchea OR TS = Cameroon OR TS = Cameroon OR TS = Cameron OR TS = Cameron OR TS = 'Cape Verde' OR TS = 'Central African Republic' OR TS = Chad OR TS = Chile OR TS = China OR TS = Colombia OR TS = Comoros OR TS = 'Comoro Islands' OR TS = Comoros OR TS = Mayotte OR TS = Congo OR TS = Zaire OR TS = Costa Rica OR TS = 'Cote d'Ivoire' OR TS = 'Ivory Coast' OR TS = Croatia OR TS = Cuba OR TS = Cyprus OR TS = Czechoslovakia OR TS = 'Czech Republic' OR TS = Slovakia OR TS = Slovak Republic OR TS = Djibouti OR TS = 'French Somaliland' OR TS = Dominica OR TS = 'Dominican Republic' OR TS = 'East Timor' OR TS = 'Timor Leste' OR TS = Ecuador OR TS = Egypt OR TS = 'United Arab Republic' OR TS = 'El Salvador' OR TS = Eritrea OR TS = Estonia OR TS = Ethiopia OR TS = Fiji OR TS = Gabon OR TS = 'Gabonese Republic' OR TS = Gambia OR TS = Gaza OR TS = 'Georgia Republic' OR TS = 'Georgian Republic' OR TS = Ghana OR TS = 'Gold Coast' OR TS = Greece OR TS = Grenada OR TS = Guatemala OR TS = Guinea OR TS = Guam OR TS = Guiana OR TS = Guyana OR TS = Haiti OR TS = Honduras OR TS = Hungary OR TS = India OR TS = Maldives OR TS = Indonesia OR TS = Iran OR TS = Iraq OR TS = Isle of Man OR TS = Jamaica OR TS = Jordan OR TS = Kazakhstan OR TS = Kazakh OR TS = Kenya OR TS = Kiribati OR TS = Korea OR TS = Kosovo OR TS = Kyrgyzstan OR TS = Kirgiz OR TS = Kyrgyz Republic OR TS = Kirghiz OR TS = Kyrgyzstan OR TS = 'Lao PDR' OR TS = Laos OR TS = Latvia OR TS = Lebanon OR TS = Lesotho OR TS = Basutoland OR TS = Liberia OR TS = Libya OR TS = Lithuania OR TS = Macedonia OR TS = Madagascar OR TS = 'Malagasy Republic' OR TS = Malaysia OR TS = Malaya OR TS = Malay OR TS = Sabah OR TS = Sarawak OR TS = Malawi OR TS = Nyasaland OR TS = Mali OR TS = Malta OR TS = Marshall Islands OR TS = Mauritania OR TS = Mauritius OR TS = Mexico OR TS = Micronesia OR TS = 'Middle East' OR TS = Moldova OR TS = Moldova OR TS = Moldovan OR TS = Mongolia OR TS = Montenegro OR TS = Morocco OR TS = Ifni OR TS = Mozambique OR TS = Myanmar OR TS = Burma OR TS = Namibia OR TS = Nepal OR TS = Netherlands Antilles OR TS = New Caledonia OR TS = Nicaragua OR TS = Niger OR TS = Nigeria OR TS = Northern Mariana Islands OR TS = Oman OR TS = Muscat OR TS = Pakistan OR TS = Palau OR TS = Palestine OR TS = Panama OR TS = Paraguay OR TS = Peru OR TS = Philippines OR TS = Poland OR TS = Portugal OR TS = 'Puerto Rico' OR TS = Romania OR TS = Rumania OR TS = Russia OR TS = Russian OR TS = Rwanda OR TS = Ruanda OR TS = 'Saint Kitts' OR TS = 'St Kitts' OR TS = Nevis OR TS = 'Saint Lucia' OR TS = 'St Lucia' OR TS = 'Saint Vincent' OR TS = Grenadines OR TS = Samoa OR TS = 'Samoan Islands' OR TS = 'Navigator Island' OR TS = 'Navigator Islands' OR TS = 'Sao Tome' OR TS = 'Saudi Arabia' OR TS = Senegal OR TS = Serbia OR TS = Montenegro OR TS = Seychelles OR TS = 'Sierra Leone' OR TS = 'Slovenia' OR TS = 'Sri Lanka' OR TS = Ceylon OR TS = 'Solomon Islands' OR TS = Somalia OR TS = 'South Africa' OR TS = Sudan OR TS = Suriname OR TS = Surinam OR TS = Swaziland OR TS = Syria OR TS = Tajikistan OR TS = Tadzhikistan OR TS = Tajikistan OR TS = Tanzania OR TS = Thailand OR TS = Togo OR TS = 'Togolese Republic' OR TS = Tonga OR TS = Trinidad OR TS = Tobago OR TS = Tunisia OR TS = Turkey OR TS = Turkmenistan OR TS = Turkmen OR TS = Uganda OR TS = Ukraine OR TS = Uruguay OR TS = USSR OR TS =

Soviet Union OR TS = Union of Soviet Socialist Republics OR TS = Uzbekistan OR TS = Uzbek OR TS = Vanuatu OR TS = New Hebrides OR TS = Venezuela OR TS = Vietnam OR TS = Viet Nam OR TS = West Bank OR TS = Yemen OR TS = Yugoslavia OR TS = Zambia OR TS = Zimbabwe OR TS = Rhodesia OR TS = Africa OR TS = Asia OR TS = Caribbean OR TS = 'West Indies' OR TS = 'South America' OR TS = 'Latin America' OR TS = 'Central America') AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

7-3,843 #6 AND #5

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

6- 49,106

(TS = (primary NEAR/1 school*) OR TS = (elementary NEAR/1 school*) OR TS = (high NEAR/1 school*) OR TS = (secondary NEAR/1 School*) OR TS = (Secondary NEAR/1 Teach*) OR TS = (secondary NEAR/1 education) OR TS = (primary NEAR/1 education) OR TS = (compulsory NEAR/1 education) OR TS = (elementary NEAR/1 education) OR TS = (school NEAR/1 girl*) OR TS = (school NEAR/1 boys)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

5-161,518

(TS = Accountability OR TS = benchmarking OR TS = (monitoring NEAR/1 activit*) OR TS = (monitoring NEAR/1 system*) OR TS = (progress NEAR/1 monitoring) OR TS = (monitoring NEAR/1 mechanism*) OR TS = (monitoring NEAR/1 process*) OR TS = (monitoring NEAR/1 procedure*) OR TS = (targeted NEAR/1 monitoring) OR TS = (inspection*) OR TS = (inspector*) OR TS = (supervis*) OR TS = (EMIS) OR TS = (education NEAR/1 management NEAR/1 information NEAR/1 system) OR TS = (performance NEAR/1 review*) OR TS = (financial NEAR/1 management) OR TS = (audit*) OR TS = (budget*) OR TS = (education* NEAR/1 finance) OR TS = (Total NEAR/1 quality NEAR/1 management) OR TS = (quality NEAR/1 assurance) OR TS = (quality NEAR/1 control) OR TS = (information NEAR/1 management) OR TS = (database NEAR/1 management) OR TS = (information NEAR/1 system*) OR TS = (decision NEAR/1 support NEAR/1 system*) OR TS = (standardised NEAR/1 test*) OR TS = (standardized NEAR/1 test*) OR TS = (budget NEAR/1 tracking) OR TS = (appraisal NEAR/1 process*) OR TS = (management NEAR/1 education) OR TS = (competency-based NEAR/1 education) OR TS = (competency NEAR/1 based NEAR/1 education) OR TS = (performance NEAR/1 based) OR TS = (result* NEAR/1 based) OR TS = (outcome-based) OR TS = (outcome NEAR/1 based) OR TS = (alternative NEAR/1 assessment) OR TS = (curriculum NEAR/1 based NEAR/1 assessment) OR TS = (curriculum- based NEAR/1 assessment) OR TS = (educational NEAR/1 assessment) OR TS = (assessment NEAR/1 procedure) OR TS = (standardised NEAR/1 assessment) OR TS = (standardized NEAR/1 assessment) OR TS = (informal NEAR/1 assessment) OR TS = (assessment NEAR/1 system*) OR TS = (assessment NEAR/1 mechanism*) OR TS = (assessment NEAR/1 process*) OR TS = (educational NEAR/1 quality) OR TS = (performance NEAR/1 factor*) OR TS = (performance NEAR/1 indicator*) OR TS = (performance NEAR/1 management) OR TS = (educational NEAR/1 indicator*) OR TS = (performance NEAR/1 information) OR TS = (personnel NEAR/1 evaluation) OR TS = (program* NEAR/1 monitoring) OR TS = (progress NEAR/1 reporting) OR TS = (recordkeeping) OR TS = (achievement NEAR/1 test*) OR TS = (assessment NEAR/1 program*) OR TS = (referenced NEAR/1 tests) OR TS = (educational NEAR/1 test*) OR TS = (high NEAR/1 stakes NEAR/1 test*) OR TS = (national NEAR/1 test*) OR TS = (international NEAR/1 test*) OR TS = (competency NEAR/1 test*) OR TS = (competency-based NEAR/1 test*) OR TS = (competency NEAR/1 assessment) OR TS = (performance NEAR/1 test*) OR TS = (standardised NEAR/1 assessment) OR TS = (quality NEAR/1 review) OR TS = (results- based NEAR/1 performance) OR TS = (performance NEAR/1 evaluation) OR TS = (information NEAR/1 utilization) OR TS = (personnel NEAR/1 management) OR TS = (educational NEAR/1 management) OR TS = (educational NEAR/1 administration) OR TS = (educational NEAR/1 environment) OR TS = (educational NEAR/1 finance) OR TS = (government NEAR/1 regulation) OR TS = (quality NEAR/1 assurance) OR TS = (quality NEAR/1 control) OR TS = (organizational NEAR/1 performance) OR TS = (organizational NEAR/1 effective) OR TS = (performance NEAR/1 test*) OR TS =

(management NEAR/1 styles) OR TS = (administrative NEAR/1 organization) OR TS = (national NEAR/1 competency NEAR/1 test*) OR TS = (norm NEAR/1 referenced NEAR/1 test*) OR TS = (criterion NEAR/1 referenced NEAR/1 test*) OR TS = (exit NEAR/1 examination) OR TS = (administration NEAR/1 effectiveness) OR TS = (administrator NEAR/1 education)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

4-10,391 #3 OR #2 OR #1

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

3-6,364

(TS=(school NEAR/1 personnel) OR TS=(school NEAR/1 boards) OR TS=(school NEAR/ 1 governance) OR TS=(school NEAR/1 visitation) OR TS=(school NEAR/1 record*) OR TS=(school NEAR/1 report*) OR TS=(school NEAR/1 performance) OR TS=(school NEAR/1 self NEAR/1 evaluation) OR TS=(school NEAR/1 management) OR TS=(school NEAR/1 efficiency) OR TS=(school NEAR/1 autonomy)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

2- 4,442

(TS = (school NEAR/1 environment*) OR TS= (school NEAR/1 finance) OR TS=(school NEAR/1 administrat*) OR TS=(School NEAR/1 audit*) OR TS=(school NEAR/1 budget*) OR TS=(school NEAR/1 inspect*) OR TS=(school NEAR/1 supervis*) OR TS=(school NEAR/ 1 self NEAR/1 assessment) OR TS=(school NEAR/1 monitoring) OR TS=(school NEAR/1 account*) OR TS=(school NEAR/1 based NEAR/1 management) OR TS=(school NEAR/1 effectiveness)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

1-440

(TS=(school NEAR/1 accountability) OR TS=(education* NEAR/1 accountability)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

Indexes=SSCI, A&HCI, CPCI-SSH Timespan=1990-2014

PSYCINFO (OVID) (23 APRIL 2014)

1. educational finance.mp.

2. educational quality.mp. or exp *Educational Quality/

3. database management systems.mp. or exp *Information Systems/

4. exp *Decision Making/ or exp *Knowledge Management/ or information management.mp.

5. exp *Quality Control/ or quality assurance.mp.

6. total quality management.mp.

7. exp *Stanford Achievement Test/ or exp *Educational Measurement/ or achievement tests.mp.

8. criterion reference tests.mp.

9. high stakes tests.mp.

10. exit examinations.mp.

11. competency based education.mp.

12. exp *Performance Tests/ or performance based assessment.mp.

13. educational assessment.mp.

14. curriculum based assessment.mp. or exp *Curriculum Based Assessment/

15. exp *Standardized Tests/ or national tests.mp.

16. standardized tests.mp.

17. norm referenced tests.mp.

18. progress monitoring.mp.

19. educational indicators.mp.

20. school supervision.mp.

21. exp *Educational Standards/ or exp *Teacher Effectiveness Evaluation/ or school inspection.mp.

22. school accounting.mp.

23. school based management.mp.

24. school effectiveness.mp.

25. school governance.mp.

26. school visitation.mp.

27. school monitoring.mp.

28. education monitoring.mp.

29. Education Management Information System.mp.

30. school performance review.mp.

31. exp Budgets/ or budget tracking.mp.

32. management education.mp.

33. result based education.mp.

34. outcome based education.mp.

35. school reports.mp.

36. school report cards.mp.

37. school audit$.mp.

38. school efficiency.mp.

39. school autonomy.mp.

40. school personnel management.mp.

41. school human resource.mp.

42. inspector$.mp.

43. school supervisor$.mp.

44. education accountability.mp.

45. school accountability.mp.

46. exp *Accountability/

47. exp *'Boards of Education'/ or school board$.mp.

48. education benchmark$.mp.

49. school benchmark$.mp.

50. school environment.mp. or exp *School Environment/

51. education environment.mp.

52. exp *Educational Administration/ or education administration.mp.

53. school administration.mp.

54. school administrator$.mp. or exp *School Administrators/

55. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or

28 or 29 or 31 or 32 or 33 or 34 or 42 or 44 or 48 or 51 or 52

56. (Afghanistan or Albania or Algeria or Angola or Antigua or Barbuda or Argentina or Armenia or Armenian or Aruba or Azerbaijan or Bahrain or Bangladesh or Barbados or Benin or Belarus or Byelorussian or Belarus or Belorussian or Belorussia or Belize or Bhutan or Bolivia or Bosnia or Herzegovina or Herzegovina or Botswana or Brazil or Bulgaria or 'Burkina Faso' or 'Upper Volta' or Burundi or Cambodia or 'Khmer Republic' or Kampuchea or Cameroon or Cameroon or Cameron or Cameron or 'Cape Verde' or 'Central African Republic' or Chad or Chile or China or Colombia or Comoros or 'Comoro Islands' or Comoros or Mayotte or Congo or Zaire or Costa Rica or 'Cote d'Ivoire' or 'Ivory Coast' or Croatia or Cuba or Cyprus or Czechoslovakia or 'Czech Republic' or Slovakia or Slovak Republic or Djibouti or 'French Somaliland' or Dominica or 'Dominican Republic' or 'East Timor' or 'Timor Leste' or Ecuador or Egypt or 'United Arab Republic' or 'El Salvador' or Eritrea or Estonia or Ethiopia or Fiji or Gabon or 'Gabonese Republic' or Gambia or Gaza or 'Georgia Republic'

or 'Georgian Republic' or Ghana or 'Gold Coast' or Greece or Grenada or Guatemala or Guinea or Guam or Guiana or Guyana or Haiti or Honduras or Hungary or India or Maldives or Indonesia or Iran or Iraq or Isle of Man or Jamaica or Jordan or Kazakhstan or Kazakh or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirgiz or Kyrgyz Republic or Kirghiz or Kyrgyzstan or Lao PDR or Laos or Latvia or Lebanon or Lesotho or Basutoland or Liberia or Libya or Lithuania or Macedonia or Madagascar or Malagasy Republic or Malaysia or Malaya or Malay or Sabah or Sarawak or Malawi or Nyasaland or Mali or Malta or Marshall Islands or Mauritania or Mauritius or Mexico or Micronesia or 'Middle East' or Moldova or Moldova or Moldovan or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Burma or Namibia or Nepal or Netherlands Antilles or New Caledonia or Nicaragua or Niger or Nigeria or Northern Mariana Islands or Oman or Muscat or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Poland or Portugal or Puerto Rico or Romania or Rumania or Russia or Russian or Rwanda or Ruanda or 'Saint Kitts' or 'St Kitts' or Nevis or 'Saint Lucia' or 'St Lucia' or 'Saint Vincent' or Grenadines or Samoa or 'Samoan Islands' or 'Navigator Island' or 'Navigator Islands' or 'Sao Tome' or 'Saudi Arabia' or Senegal or Serbia or Montenegro or Seychelles or 'Sierra Leone' or 'Slovenia' or 'Sri Lanka' or Ceylon or 'Solomon Islands' or Somalia or 'South Africa' or Sudan or Suriname or Surinam or Swaziland or Syria or Tajikistan or Tadzhikistan or Tajikistan or Tanzania or Thailand or Togo or Togolese Republic or Tonga or Trinidad or Tobago or Tunisia or Turkey or Turkmenistan or Turkmen or Uganda or Ukraine or Uruguay or USSR or Soviet Union or Union of Soviet Socialist Republics or Uzbekistan or Uzbek or Vanuatu or New Hebrides or Venezuela or Vietnam or Viet Nam or West Bank or Yemen or Yugoslavia or Zambia or Zimbabwe or Rhodesia or Africa or Asia or Caribbean or 'West Indies' or 'South America' or 'Latin America' or 'Central America').mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]

57. developing countries.mp. or exp *Developing Countries/

58. (developing nation$ or developing countr$ or developing world or developing economy$ or less$ developed countries or less$ developed nation$ or less$ developed world or less$ developed econom$ or underdeveloped countr$ or underdeveloped nation$ or underdeveloped world or underdeveloped economies or under developed nation$ or under developed world or under developed economies or low$ income countries or low$ income nation$ or low$ income econom$ or low$ middle countr$ or LMIC or LMICs or LLMIC or LLMICs or third world or underserved countr$ or underserved nation$ or deprived countr$ or deprived nation$ or deprived world or poor$ countr$ or poor$ nation$).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures]

59. 56 or 57 or 58

60. 46 or 55

61. 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 30 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 43 or

45 or 47 or 49 or 50 or 53 or 54

62. 59 and 61

63. limit 62 to (english language and yr='1990 -Current')

64. exp *Junior High Schools/ or exp *Charter Schools/ or exp *High Schools/ or exp *Boarding Schools/ or exp *Middle Schools/ or exp *Elementary Schools/

65. primary schools.mp.

66. elementary schools.mp.

67. high schools.mp.

68. exp *Elementary School Students/ or exp *Elementary School Teachers/ or exp *Elementary Education/ or exp *Primary School Students/ or primary education.mp.

69. exp *High School Students/ or exp *Secondary Education/ or secondary schools.mp. or exp *High School Teachers/

70. secondary education.mp.

71. compulsory education.mp.

72. 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71

73. 55 and 59 and 72

74. limit 73 to (english language and yr='1990 -Current')

ECONLIT (PROQUEST) (23 APRIL 2014)

Set 1: three terms

((Accountability OR benchmarking OR (monitoring PRE/1 activit*) OR (monitorin PRE/1 system*) OR (progress PRE/1 monitoring) OR (monitoring PRE/1 mechanism*) OR (monitoring PRE/1 process*) OR (monitoring PRE/1 procedure*) OR (targeted PRE/1 monitoring) OR (inspection*) OR (inspector*) OR (supervis*) OR (EMIS) OR (education PRE/1 management PRE/1 information PRE/1 system) OR (performance PRE/1 review*) OR (financial PRE/1 management) OR (audit*) OR (budget*) OR (education* PRE/1 finance) OR (Total PRE/1 quality PRE/1 management) OR (quality PRE/1 assurance) OR (quality PRE/1 control) OR (information PRE/1 management) OR (database PRE/1 management) OR (information PRE/1 system*) OR (decision PRE/1 support PRE/1 system*) OR (standardised PRE/1 test*) OR (standardized PRE/1 test*) OR (budget PRE/1 tracking) OR (appraisal PRE/1 process*) OR (management PRE/1 education) OR (competency-based PRE/1 education) OR (competency PRE/1 based PRE/1 education) OR (performance PRE/1 based) OR (result* PRE/1 based) OR (outcome-based) OR (outcome PRE/1 based) OR (alternative PRE/1 assessment) OR (curriculum PRE/1 based PRE/1 assessment) OR (curriculum-based PRE/1 assessment) OR (educational PRE/1 assessment) OR (assessment PRE/1 procedure) OR (standardised PRE/1 assessment) OR (standardized PRE/1 assessment) OR (informal PRE/1 assessment) OR (assessment PRE/1 system*) OR (assessment PRE/1 mechanism*) OR (assessment PRE/1 process*) OR (educational PRE/1 quality) OR (performance PRE/1 factor*) OR (performance PRE/1 indicator*) OR (performance PRE/1 management) OR (educational PRE/1 indicator*) OR (performance PRE/1 information) OR (personnel PRE/1 evaluation) OR (program* PRE/1 monitoring) OR (progress PRE/1 reporting) OR (recordkeeping) OR (achievement PRE/1 test*) OR (assessment PRE/1 program*) OR (referenced PRE/1 tests) OR (educational PRE/1 test*) OR (high PRE/1 stakes PRE/1 test*) OR (national PRE/1 test*) OR (international PRE/1 test*) OR (competency PRE/1 test*) OR (competency-based PRE/1 test*) OR (competency PRE/1 assessment) OR (performance PRE/1 test*) OR (standardised PRE/1 assessment) OR (quality PRE/1 review) OR (results- based PRE/1 performance) OR (performance PRE/1 evaluation) OR (information PRE/1 utilization) OR (personnel PRE/1 management) OR (educational PRE/1 management) OR (educational PRE/1 administration) OR (educational PRE/1 environment) OR (educational PRE/1 finance) OR (government PRE/1 regulation) OR (quality PRE/1 assurance) OR (quality PRE/1 control) OR

(organizational PRE/1 performance) OR (organizational PRE/1 effective) OR (performance PRE/1 test*) OR (management PRE/1 styles) OR (administrative PRE/1 organization) OR (national PRE/1 competency PRE/1 test*) OR (norm PRE/1 referenced PRE/1 test*) OR (criterion PRE/1 referenced PRE/1 test*) OR (exit PRE/1 examination) OR (administration PRE/1 effectiveness) OR (administrator PRE/1 education)) AND ((SU.exact('DEVELOPING COUNTRIES LDCS') OR SU.exact('LDCS') OR LOC.exact('DEVELOPING COUNTRIES') OR SU.exact('LESS DEVELOPED COUNTRIES') OR SU.exact('LESS DEVELOPED COUNTRIES') OR ORG.exact('IDCS') OR SU.exact('DEVELOPING COUNTRIES')) OR ab(Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (developing PRE/1 nation*) OR (developing PRE/1 countr*) OR (developing PRE/1 world) OR (developing PRE/1 econom*) OR (less* PRE/1 developed PRE/1 countries) OR (less* PRE/1 developed PRE/1 nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 econom*) OR (underdeveloped PRE/1 countr*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 econom*) OR (low* PRE/2 middle PRE/2 countr*) OR (LMIC) OR (LMICs) OR (LLMIC) OR (LLMICs) OR (third PRE/1 world) OR (underserved PRE/1 countr*) OR (underserved PRE/1 nation*) OR

(deprived PRE/1 countr*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 countr*) OR (poor* PRE/1 nation*))))) AND ((primary PRE/1 school*) OR (elementary PRE/1 school*) OR (high PRE/1 school*) OR (secondary PRE/1 School*) OR (Secondary PRE/1 Teach*) OR (secondary PRE/1 education) OR (primary PRE/1 education) OR (compulsory PRE/1 education) OR (elementary PRE/1 education) OR (schools) OR (school PRE/1 girl*) OR (school PRE/1 boys) OR (school) OR (schools))

Set 2

((School-based PRE/1 management) OR (school PRE/1 effectiveness) OR (school PRE/1 governance) OR (school PRE/1 organisation) OR (school PRE/1 monitoring) OR (School-self PRE/1 evaluation) OR (School PRE/1 accounting) OR (School PRE/1 administration) OR (school PRE/1 record) OR (school PRE/1 report*) OR (school-self PRE/1 assessment) OR (school PRE/1 evaluation) OR (school PRE/1 performance) OR (School PRE/1 card*) OR (school PRE/1 management) OR (school PRE/1 supervision) OR (school PRE/1 inspection) OR (school PRE/1 accountability) OR (school PRE/1 assessment) OR (school PRE/1 audit*) OR (school PRE/1 efficiency) OR (school PRE/1 autonomy) OR (primary PRE/1 education PRE/1 monitoring) OR (primary PRE/1 education PRE/1 assessment) OR (elementary PRE/1 education PRE/1 monitoring) OR (elementary PRE/1 education PRE/1 assessment) OR (secondary PRE/1 education PRE/1 assessment) OR (secondary PRE/1 education PRE/1 monitoring) OR (school PRE/1 environment) OR (school PRE/1 boards) OR (school PRE/1 visitation) OR (school PRE/1 finance) OR (school PRE/1 personnel)) AND ((SU.exact('DEVELOPING COUNTRIES LDCS') OR SU.exact('LDCS') OR LOC.exact('DEVELOPING COUNTRIES') OR SU.exact('LESS DEVELOPED COUNTRIES') OR SU.exact('LESS DEVELOPED COUNTRIES') OR ORG.exact('IDCS') OR SU.exact('DEVELOPING COUNTRIES')) OR ab(Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR

Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (developing PRE/1 nation*) OR (developing PRE/1 countr*) OR (developing PRE/1 world) OR (developing PRE/1 econom*) OR (less* PRE/1 developed PRE/1 countries) OR (less* PRE/1 developed PRE/1 nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 econom*) OR (underdeveloped PRE/1 countr*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 econom*) OR (low* PRE/2 middle PRE/2 countr*) OR (LMIC) OR (LMICs) OR (LLMIC) OR (LLMICs) OR (third PRE/1 world) OR (underserved PRE/1 countr*) OR (underserved PRE/1 nation*) OR (deprived PRE/1 countr*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 countr*) OR (poor* PRE/1 nation*)))

SOCIOLOGICAL ABSTRACTS (04 APRIL 2014) AND SOCIAL SERVICE ABSTRACTS (23 APRIL 2014) (PROQUEST)

Set 1 (Three terms)

Searched for:

((((SU.EXACT('Management Styles') OR SU.EXACT.EXPLODE('Organizational Effectiveness') OR SU.EXACT.EXPLODE('Accountability') OR SU.EXACT.EXPLODE('Government Regulation') OR SU.EXACT.EXPLODE('Supervision') OR SU.EXACT.EXPLODE('Educational Administration') OR SU.EXACT.EXPLODE('Managers') OR SU.EXACT.EXPLODE('Achievement Tests') OR SU.EXACT('Evaluation') OR SU.EXACT('Personnel Management')) OR ab(Accountability OR benchmarking OR (monitoring NEAR/1 activit*) OR (monitoring NEAR/1 system*) OR (progress PRE/1 monitoring) OR (monitoring PRE/1 mechanism*) OR (monitoring PRE/1 process*) OR (monitoring PRE/1 procedure*) OR (targeted PRE/1 monitoring) OR (inspection*) OR (inspector*) OR (supervis*) OR (EMIS) OR (education PRE/1 management PRE/1 information PRE/1 system) OR (performance PRE/1 review*) OR (financial PRE/1 management) OR (audit*) OR (budget*) OR (education* PRE/1 finance) OR (Total PRE/1 quality PRE/1 management) OR (quality PRE/1 assurance) OR (quality PRE/1 control) OR (information PRE/1 management) OR (database PRE/1 management) OR (information PRE/ 1 system*) OR (decision PRE/1 support PRE/1 system*) OR (standardised PRE/1 test*) OR (standardized PRE/1 test*) OR (budget PRE/1 tracking) OR (appraisal PRE/1 process*) OR (management PRE/1 education) OR (competency-based PRE/1 education) OR (competency PRE/1 based PRE/1 education) OR (performance PRE/1 based) OR (result* PRE/1 based) OR (outcome-based) OR (outcome PRE/1 based) OR (alternative PRE/1 assessment) OR (curriculum PRE/1 based PRE/1 assessment) OR (curriculum-based PRE/1 assessment) OR (educational PRE/1 assessment) OR (assessment PRE/1 procedure) OR (standardised PRE/ 1 assessment) OR (standardized PRE/1 assessment) OR (informal PRE/1 assessment) OR (assessment PRE/1 system*) OR (assessment PRE/1 mechanism*) OR (assessment PRE/1 process*) OR (educational PRE/1 quality) OR (performance PRE/1 factor*) OR (performance PRE/1 indicator*) OR (performance PRE/1 management) OR (educational PRE/1 indicator*) OR (performance PRE/1 information) OR (personnel PRE/1 evaluation) OR (program* PRE/1 monitoring) OR (progress PRE/1 reporting) OR (recordkeeping) OR (achievement PRE/1 test*) OR (assessment PRE/1 program*) OR (referenced PRE/1 tests) OR (educational PRE/ 1 assessment) OR (educational PRE/1 test*) OR (high PRE/1 stakes PRE/1 test*) OR (national PRE/1 test*) OR (international

PRE/1 test*) OR (competency PRE/1 test*) OR (competency-based PRE/1 test*) OR (competency PRE/1 assessment) OR (performance PRE/1 test*) OR (standardised PRE/1 assessment) OR (standardized PRE/1 assessment) OR (quality PRE/1 review) OR (results-based PRE/1 performance) OR (performance PRE/1 evaluation) OR (information PRE/1 utilization) OR (personnel PRE/1 management))) AND ((SU.EXACT.EXPLODE('Secondary Education') OR SU.EXACT.EXPLODE('High Schools' OR 'Elementary Schools' OR 'High Schools' OR 'Junior High Schools' OR 'Polytechnic Schools' OR 'Private Schools' OR 'Public Schools' OR 'Schools' OR 'Secondary Schools') OR SU.EXACT.EXPLODE('School Districts') OR SU.EXACT.EXPLODE('High Schools' OR 'Junior High Schools' OR 'Secondary Schools') OR SU.EXACT.EXPLODE('Junior High School Students') OR SU.EXACT.EXPLODE('High School Students') OR SU.EXACT.EXPLODE('Primary Education') OR SU.EXACT('Junior High Schools') OR SU.EXACT.EXPLODE('Elementary School Students') OR SU.EXACT('Public Schools') OR SU.EXACT.EXPLODE('Elementary School Students' OR 'High School Students' OR 'Junior High School Students') OR SU.EXACT.EXPLODE('Private Schools') OR SU.EXACT.EXPLODE('Elementary Schools')) OR ab((primary PRE/1 school*) OR (elementary PRE/1 school*) OR (high PRE/1 school*) OR (secondary PRE/1 School*) OR (Secondary PRE/1 Teach*) OR (secondary PRE/1 education) OR (primary PRE/1 education) OR (compulsory PRE/1 education) OR (elementary PRE/1 education) OR (schools) OR (school PRE/1 girl*) OR (school PRE/1 boys) OR (school)))) AND (ab(Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR

'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (developing PRE/1 nation*) OR (developing PRE/1 countr*) OR (developing PRE/1 world) OR (developing PRE/1 econom*) OR (less* PRE/1 developed PRE/ 1 countries) OR (less* PRE/1 developed PRE/1 nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 econom*) OR (underdeveloped PRE/1 countr*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/ 1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 econom*) OR (low* PRE/2 middle PRE/2 countr*) OR (LMIC) OR (LMICs) OR (LLMIC) OR (LLMICs) OR (third PRE/1 world) OR (underserved PRE/1 countr*) OR (underserved PRE/1 nation*) OR (deprived PRE/1 countr*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 countr*) OR (poor* PRE/1 nation*)) OR su(developing countries))) AND yr(1990-2019)

Set 2

(ab(Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (developing PRE/1 nation*) OR (developing PRE/1 countr*) OR (developing PRE/1 world) OR (developing PRE/1 econom*) OR (less* PRE/1 developed PRE/1

countries) OR (less* PRE/1 developed PRE/1 nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 econom*) OR (underdeveloped PRE/1 countr*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 econom*) OR (low* PRE/2 middle PRE/2 countr*) OR (LMIC) OR (LMICs) OR (LLMIC) OR (LLMICs) OR (third PRE/1 world) OR (underserved PRE/1 countr*) OR (underserved PRE/1 nation*) OR (deprived PRE/1 countr*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 countr*) OR (poor* PRE/1 nation*)) OR su(developing countries)) AND ((SU.EXACT('School Environment') OR SU.EXACT.EXPLODE('Academic Achievement')) OR ab(((School-based PRE/1 management) OR (school PRE/1 effectiveness) OR (school PRE/1 governance) OR (school PRE/1 organisation) OR (school PRE/1 monitoring) OR (School-self PRE/1 evaluation) OR (School PRE/1 accounting) OR (School PRE/1 administration) OR (school PRE/1 record) OR (school PRE/1 report*) OR (school-self PRE/1 assessment) OR (school PRE/1 evaluation) OR (school PRE/1 performance) OR (school PRE/1 boards) OR (school PRE/1 governance) OR (school PRE/1 environment))) OR SU.EXACT.EXPLODE('School Boards'))

IBSS (PROQUEST) 22 APRIL 2014

Set 1 (Three terms)

Searched for:

((((SU.EXACT('Management Styles') OR SU.EXACT.EXPLODE('Organizational Effectiveness') OR SU.EXACT.EXPLODE('Accountability') OR SU.EXACT.EXPLODE('Government Regulation') OR SU.EXACT.EXPLODE('Supervision') OR SU.EXACT.EXPLODE('Educational Administration') OR SU.EXACT.EXPLODE('Managers') OR SU.EXACT.EXPLODE('Achievement Tests') OR SU.EXACT('Evaluation') OR SU.EXACT('Personnel Management')) OR ab(Accountability OR benchmarking OR (monitoring NEAR/1 activit*) OR (monitoring NEAR/1 system*) OR (progress PRE/1 monitoring) OR (monitoring PRE/1 mechanism*) OR (monitoring PRE/1 process*) OR (monitoring PRE/1 procedure*) OR (targeted PRE/1 monitoring) OR (inspection*) OR (inspector*) OR (supervis*) OR (EMIS) OR (education PRE/1 management PRE/1 information PRE/1 system) OR (performance PRE/1 review*) OR (financial PRE/1 management) OR (audit*) OR (budget*) OR (education* PRE/1 finance) OR (Total PRE/1 quality PRE/1 management) OR (quality PRE/1 assurance) OR (quality PRE/1 control) OR (information PRE/1 management) OR (database PRE/1 management) OR (information PRE/ 1 system*) OR (decision PRE/1 support PRE/1 system*) OR (standardised PRE/1 test*) OR (standardized PRE/1 test*) OR (budget PRE/1 tracking) OR (appraisal PRE/1 process*) OR (management PRE/1 education) OR (competency-based PRE/1 education) OR (competency PRE/1 based PRE/1 education) OR (performance PRE/1 based) OR (result* PRE/1 based) OR (outcome-based) OR (outcome PRE/1 based) OR (alternative PRE/1 assessment) OR (curriculum PRE/1 based PRE/1 assessment) OR (curriculum-based PRE/1 assessment) OR (educational PRE/1 assessment) OR (assessment PRE/1 procedure) OR (standardised PRE/ 1 assessment) OR (standardized PRE/1 assessment) OR (informal PRE/1 assessment) OR (assessment PRE/1 system*) OR (assessment PRE/1 mechanism*) OR (assessment PRE/1 process*) OR (educational PRE/1 quality) OR (performance PRE/1 factor*) OR (performance PRE/1 indicator*) OR (performance PRE/1 management) OR (educational PRE/1 indicator*) OR (performance PRE/1 information) OR (personnel PRE/1 evaluation) OR (program* PRE/1 monitoring) OR (progress PRE/1 reporting) OR (recordkeeping) OR (achievement PRE/1 test*) OR (assessment PRE/1 program*) OR (referenced PRE/1 tests) OR (educational PRE/ 1 assessment) OR (educational PRE/1 test*) OR (high PRE/1 stakes PRE/1 test*) OR (national PRE/1 test*) OR (international PRE/1 test*) OR (competency PRE/1 test*) OR (competency-based PRE/1 test*) OR (competency PRE/1

assessment) OR (performance PRE/1 test*) OR (standardised PRE/1 assessment) OR (standardized PRE/1 assessment) OR (quality PRE/1 review) OR (results-based PRE/1 performance) OR (performance PRE/1 evaluation) OR (information PRE/1 utilization) OR (personnel PRE/1 management))) AND ((SU.EXACT.EXPLODE('Secondary Education') OR SU.EXACT.EXPLODE('High Schools' OR 'Elementary Schools' OR 'High Schools' OR 'Junior High Schools' OR 'Polytechnic Schools' OR 'Private Schools' OR 'Public Schools' OR 'Schools' OR 'Secondary Schools') OR SU.EXACT.EXPLODE('School Districts') OR SU.EXACT.EXPLODE('High Schools' OR 'Junior High Schools' OR 'Secondary Schools') OR SU.EXACT.EXPLODE('Junior High School Students') OR SU.EXACT.EXPLODE('High School Students') OR SU.EXACT.EXPLODE('Primary Education') OR SU.EXACT('Junior High Schools') OR SU.EXACT.EXPLODE('Elementary School Students') OR SU.EXACT('Public Schools') OR SU.EXACT.EXPLODE('Elementary School Students' OR 'High School Students' OR 'Junior High School Students') OR SU.EXACT.EXPLODE('Private Schools') OR SU.EXACT.EXPLODE('Elementary Schools')) OR ab((primary PRE/1 school*) OR (elementary PRE/1 school*) OR (high PRE/1 school*) OR (secondary PRE/1 School*) OR (Secondary PRE/1 Teach*) OR (secondary PRE/1 education) OR (primary PRE/1 education) OR (compulsory PRE/1 education) OR (elementary PRE/1 education) OR (schools) OR (school PRE/1 girl*) OR (school PRE/1 boys) OR (school)))) AND (ab(Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR Kirgiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Ifni OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (developing PRE/1 nation*) OR

(developing PRE/1 countr*) OR (developing PRE/1 world) OR (developing PRE/1 econom*) OR (less* PRE/1 developed PRE/ 1 countries) OR (less* PRE/1 developed PRE/1 nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 econom*) OR (underdeveloped PRE/1 countr*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/ 1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 econom*) OR (low* PRE/2 middle PRE/2 countr*) OR (LMIC) OR (LMICs) OR (LLMIC) OR (LLMICs) OR (third PRE/1 world) OR (underserved PRE/1 countr*) OR (underserved PRE/1 nation*) OR (deprived PRE/1 countr*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 countr*) OR (poor* PRE/1 nation*)) OR su(developing countries))) AND yr(1990-2019)

Set 2

(SU.EXACT.EXPLODE('School environment') OR SU.EXACT.EXPLODE('School administration')) AND (ab(Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Armenian OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Belarus OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Herzegovina OR Botswana OR Brazil OR Bulgaria OR 'Burkina Faso' OR 'Upper Volta' OR Burundi OR Cambodia OR 'Khmer Republic' OR Kampuchea OR Cameroon OR Cameroon OR Cameron OR Cameron OR 'Cape Verde' OR 'Central African Republic' OR Chad OR Chile OR China OR Colombia OR Comoros OR 'Comoro Islands' OR Comoros OR Mayotte OR Congo OR Zaire OR Costa Rica OR 'Cote d'Ivoire' OR 'Ivory Coast' OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR 'Czech Republic' OR Slovakia OR Slovak Republic OR Djibouti OR 'French Somaliland' OR Dominica OR 'Dominican Republic' OR 'East Timor' OR 'Timor Leste' OR Ecuador OR Egypt OR 'United Arab Republic' OR 'El Salvador' OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR 'Gabonese Republic' OR Gambia OR Gaza OR 'Georgia Republic' OR 'Georgian Republic' OR Ghana OR 'Gold Coast' OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Isle of Man OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan OR kirghiz OR Kyrgyz Republic OR Kirghiz OR Kyrgyzstan OR Lao PDR OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR Malagasy Republic OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR Marshall Islands OR Mauritania OR Mauritius OR Mexico OR Micronesia OR 'Middle East' OR Moldova OR Moldova OR Moldovan OR Mongolia OR Montenegro OR Morocco OR Mozambique OR Myanmar OR Burma OR Namibia OR Nepal OR Netherlands Antilles OR New Caledonia OR Nicaragua OR Niger OR Nigeria OR Northern Mariana Islands OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Poland OR Portugal OR Puerto Rico OR Romania OR Rumania OR Russia OR Russian OR Rwanda OR Ruanda OR 'Saint Kitts' OR 'St Kitts' OR Nevis OR 'Saint Lucia' OR 'St Lucia' OR 'Saint Vincent' OR Grenadines OR Samoa OR 'Samoan Islands' OR 'Navigator Island' OR 'Navigator Islands' OR 'Sao Tome' OR 'Saudi Arabia' OR Senegal OR Serbia OR Montenegro OR Seychelles OR 'Sierra Leone' OR 'Slovenia' OR 'Sri Lanka' OR Ceylon OR 'Solomon Islands' OR Somalia OR 'South Africa' OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tajikistan OR Tanzania OR Thailand OR Togo OR Togolese Republic OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR USSR OR Soviet Union OR Union of Soviet Socialist Republics OR Uzbekistan OR Uzbek OR Vanuatu OR New Hebrides OR Venezuela OR Vietnam OR Viet Nam OR West Bank OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR Rhodesia OR Africa OR Asia OR Caribbean OR 'West Indies' OR 'South America' OR 'Latin America' OR 'Central America' OR (developing PRE/1 nation*) OR (developing PRE/1 country*) OR (developing PRE/1 world) OR (developing PRE/1 economy*) OR (less* PRE/1 developed PRE/1 countries) OR (less* PRE/1 developed PRE/1

nation*) OR (less* PRE/1 developed PRE/1 world) OR (less* PRE/1 developed PRE/1 economy*) OR (underdeveloped PRE/1 country*) OR (underdeveloped PRE/1 nation*) OR (underdeveloped PRE/1 world) OR (underdeveloped PRE/1 economies) OR (under PRE/1 developed PRE/1 nation*) OR (under PRE/1 developed PRE/1 world) OR (under PRE/1 developed PRE/1 economies) OR (low* PRE/1 income PRE/1 countries) OR (low* PRE/1 income PRE/1 nation*) OR (low* PRE/1 income PRE/1 economy*) OR (low* PRE/2 middle PRE/2 country*) OR (LMIC) OR (laics) OR (LLMIC) OR (Llaics) OR (third PRE/1 world) OR (underserved PRE/1 country*) OR (underserved PRE/1 nation*) OR (deprived PRE/1 country*) OR (deprived PRE/1 nation*) OR (deprived PRE/1 world) OR (poor* PRE/1 country*) OR (poor* PRE/1 nation*)) OR (SU.EXACT.EXPLODE('Developing countries') OR SU.EXACT.EXPLODE('Less developed countries') OR SU.EXACT.EXPLODE('Arab countries')))

APPENDIX 2.3: SOURCES AND KEY WEBSITES SEARCHED

BIBLIOGRAPHIC AND SPECIALIST EDUCATION, SOCIAL, AND ECONOMIC DATABASES

Australian Education Index (AEI)

British Education Index (BEI)

Econlit

Education Resources Information Centre (ERIC)

International Bibliography of the Social Sciences (IBSS)

PsycINFO

Social Sciences Citations Index (SSCI)

Social Service Abstracts (SSA)

Sociological Abstracts

SPECIALIST DATABASES

3ie Database of impact evaluations: www.3ieimpact.org/database_of_impact_evaluations.html

Africa Journals Online (AJOL): www.ajol.info/

Bioline International: www.bioline.org.br/

The Campbell Library: https://campbellcollaboration.org/campbell-library/campbell-library/the-campbell-library

East View Information Service Online Databases: www.eastview.com/

EPPI-Centre: eppi.ioe.ac.uk

IDEAS Economics and Finance Database (RePEc): http://ideas.repec.org/

Indian Citation Index (ICI): www.indiancitationindex.com/

JOLIS library catalogue: http://external.worldbankimflib.org/external.htm

Nepal Journals online (NepJOL): www.nepjol.info/

OpenGrey: www.opengrey.eu/

SciDev Net (Science and Development Network): www.scidev.net/en/

Social Science Research Network (SSRN): http://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm

REFERENCES FROM EXISTING SYSTEMATIC REVIEWS AND PAPERS RELEVANT TO THE REVIEW QUESTION

Anderson (2005); Barakat et al. (2012); Bruns (2011); Carr and Leggatt-Cook (2011); Carrón and De Grauwe (2007); Clifford et al. (2013); De Grauwe (2005, 2007, 2008); DFID (undated); Gershberg and Gonzalez (2012); Guerrero et al. (2012); Hatch (2013); Hooge et al. (2012); Joshi et al. (2011); Kingdon et al. (2013); Klerks (2013); Lynch et al. (2013); Orr et al. (2013); Patrinos et al. (2007); Petrosino et al. (2013); Rosenkvist (2010); Scheerens (1999, 2000); Westhorp et al. (2012); Yu (2007).

WEBSITES

African Development Bank: http://www.afdb.org/en/

Asian Development Bank: http://www.adb.org

Association for the Development of Education in Africa: http://www.adeanet.org/

AusAID: http://dfat.gov.au/aid/pages/australias-aid-program.aspx http://aid.dfat.gov.au/Pages/home.aspx

Australian Council for Education Research: http://www.acer.edu.au/

British Library for Development Studies: http://blds.ids.ac.uk/

ELDIS: http://www.eldis.org/

The Future of Children: http://futureofchildren.org/

Google Scholar

Institute for Fiscal Studies: http://www.ifs.org.uk/

Institute of Development Studies: http://www.ids.ac.uk

Inter-American Development Bank: http://www.iadb.org/en/inter-americandevelopment-bank,2837.html

International Institute for Education Planning: http://www.iiep.unesco.org

National Bureau of Economic research: http://www.nber.org

Overseas Development Institute: http://www.odi.org.uk/

Poverty Action Lab: http://www.povertyactionlab.org/

UNDP: http://www.undp.org/content/undp/en/home.html

UNESCO: http://www.unesco.org

USAID: http://www.usaid.gov/

World Bank: http://www.worldbank.org/

APPENDIX 2.4: ADVISORY GROUP MEMO AND FEEDBACK

RESPONSES FROM THE ADVISORY GROUP

Four of the five Advisory Group members have sent us their responses at the time of writing. These are:

- Anton De Grauwe, IIEP-UNESCO [AdG]
- Thomas Hatch, Teachers College, Columbia University [TH]
- Pantalee Kapichi, UNICEF Tanzania [PK]
- Dennis Shirley, Lynch School of Education, Boston College [DS].

In our summary of responses below, the views shared by all those responding are not attributed. Attributed comments are indicated by initials in brackets. All comments are paraphrased unless we use quotation marks to indicate a direct quote. Complete responses appear in the table at the end of this Appendix.

- Clarification of scope: Does the approach we are proposing, to focus on four regions (East and South Asia, Latin America, sub-Saharan Africa) appear sound?

All concurred that these four regions are appropriate for this review.

AdG and PK raised questions about how countries will be grouped in the synthesis, and suggested considering: (a) level of income, (b) physical and administrative distance between central administrations and the school, (c) differing urban/rural poor access to education services.

- Does the decision to constrain the date range to 2001-present make sense to you?

There was general agreement that this was appropriate, and provides 'relatively similar baselines' across regions [PK]. Relevant date range could vary considerably depending on country [AdG], and exceptions should be considered in order to include reports prior to 2001 as indicated by frequency of reference in contemporary sources [DS].

- What are the limitations of these approaches in your view?

There was a wide variation within region, especially of economic stage of development [AdG, PK]. There was also potential for wide variation within country in terms of provinces and districts [DS]. Suggestion: identify similar groupings of countries within regions and then compare sub-regional grouping across regions [PK, DS]. 'Comparative analyses between a small number of nations … with roughly similar circumstances most helpful' [DS]. DS also suggested that it would be useful to include the experience of implementing accountability in 'failed' states and compare these with countries with reasonably intact accountability systems in order to understand better institutional breakdowns.

- Request for additional information: What additional sources (people, websites, centres, etc.) would you suggest we contact and/or include, given our intention to focus on East and South Asia, sub-Saharan Africa and Latin America?

AdG and PK mentioned several specific sources that we are now screening for inclusion.

- Which specific countries in these regions, in your opinion, would provide the most insight into understanding the systemic processes of accountability and outcomes for schools and learners?

AdG, PK and DS mentioned specific countries. Latin America: Chile (2),[6] Brazil, Mexico; East Asia: Indonesia; South Asia: Sri Lanka; sub-Saharan Africa: Ghana, Uganda (2), Tanzania, Kenya, Rwanda, Burundi.

- Do you have any other thoughts or suggestions about how we might effectively limit the scope of our review given available resources and time?

AdG argued that 'inspection/supervision' and 'assessment' are somewhat contrasting forms of monitoring, and 'monitoring' on its own is too broad to be useful. Consider eliding 'monitoring' and focusing on more specific tools of 'assessment' and 'inspection'.

CONCLUSIONS FROM ADVISORY GROUP FEEDBACK

a. *Regional focus:* Based on our systematic mapping of the sources we have identified, we propose to focus the review on specific regions that are most relevant to DfID priorities in improving educational outcomes for the poorest and most marginalised. This entails limiting the review to four regions that have the highest concentration of studies: East and South Asia, sub-Saharan Africa and Latin America. (Latin America is not a region of focus for DfID; however, because of the number of relevant studies available and the potential for offering robust comparison and contrast with other regions, we propose to include Latin America.)

b. *Temporal focus:* We also propose to constrain the date range of the review. Our initial searches were conducted from 1990 to the present. Systemic national and international focus on accountability policies in LMICs did not get underway until the mid-1990s, and we would expect that studies from 2001 would be sufficient to capture this early period of policy sharing and national implementation. This later cut-off also appears justified by the small number of documents from 1990-2000.

c. *Additional sources*: We are pursuing all leads, including those generated from the Advisory Group as well as other contacts, especially those whom the Co-PI has recently contacted. A considerable impediment to Latin American sources is the lack of translation. A large proportion of the most important sources are in Spanish.

d. *Country focus:* The iterative process of data extraction will allow us to identify specific countries of greatest interest, as well as develop important characteristics for grouping countries for comparison within and across regions. We will focus initially on country recommendations from the Advisory Group and informed by our ongoing research.

e. *Accountability scope*: At this point, we *do not* intend to narrow our focus to only two of the three accountability elements as a way of delimiting scope. We believe that our definition of monitoring differentiates this element adequately from the others, and that this will be an important element to consider alongside the other two.

---

[6] Numbers in parentheses correspond with times mentioned by different AG members.

Table 2.4.1: A summary of Advisory group feedback

Question	de Grauwe	Kapichi	Shirley	Hatch
Does the approach we are proposing to focus on four regions (East and South Asia, Latin America, sub-Saharan Africa) appear sound?	As the title of the review refers to 'low- and middle-income countries', this geographical limitation is appropriate. One group may be missing: Caribbean countries. They are, for a reason I will return to, an interesting group. A different question is: how will the countries be grouped (if there will be a grouping) when analysis is done? I would not use a grouping by region. Two other groupings may make more sense: -Level of income. Not because income in itself is that important, but because it tends to correlate with effectiveness of the State, and this is an important characteristic when examining the effectiveness of inspection, monitoring and assessment, which are tools of the	I concur with the logic of limiting the review to the three DFID focused regions and including Latin America for comparison purposes. The four regions share similar histories (colonial conquest with resultant education systems; Freedom and emancipation and attempts at restructuring their systems and later globalisation and adoption of global charters and agendas including MDGs, EFA, etc.). However, their recent histories differ significantly especially in terms of economic growth with some parts of East and South Asia far outpacing sub-Saharan Africa. It is also important to unpack the concept of rural/ urban poor and how it differs across the 4 regions in terms of access to education services	Yes, and I like the inclusion of Latin America and would be curious to learn what improvements in some countries (Brazil and Chile especially) might entail for other countries that have experienced slower rates of development.	I had a chance to go over the memo and the progress of the review so far, and it seems to me like you are on target. You clearly laid out what you've done, the choices you've made and the reasoning behind them. In terms of the specific questions you've asked, I think your proposals for limiting the scope of the study make the most sense.

Question	de Grauwe	Kapichi	Shirley	Hatch
	State. Ideally, an indicator on 'state effectiveness' should be used, but this does not exist. -Distance between the central administration and the school. As in most countries, core policy decisions about assessment, monitoring and inspection are taken at central level, while learning outcomes arguably are under the control of the schools, the distance between the two can be an important factor. Arguably, the closer the distance, the more immediate the relationship. That distance is both a physical and an administrative one. Physical: that's why I think it would be a good idea to include some Caribbean or other small island states, because there, the ones who exercise the actual monitoring are also the ones who define the policies. It's also administrative, namely in function of the number of administrative levels between school and Ministry.			

Question	de Grauwe	Kapichi	Shirley	Hatch
Does the decision to constrain the date range to 2001-present make sense to you?	On a global level, this probably makes sense, but it may be better to adapt this to the different countries that you may want to focus on. In South-Africa, for instance, the date range should start in 1994, with the first democratic election. In Chile, in 1990, with the end of Pinochet regime. In others, important policy reforms may define the date. In Indonesia, for instance, 1999 was the beginning of the decentralisation reform.	Constraining the date range: I concur mainly due to availability of data but also due to the fact that after mid to late 1990 majority of 3rd world countries adopted similar education restructuring approaches. The cut off year from 2001 provides relatively similar baselines and indicators for comparison.	Yes, this seems justifiable given the dynamic rate of change in recent years. Occasional exceptions will likely have to be made from time to time to include major reports that still are impactful from prior to 2001 that are referenced in the post-2001 time frame.	Excluding studies before 2001 seems like a good choice, and the regions seem to reflect the areas where most of the work has been done.
What are the limitations of these approaches in your view?	See the points above.	As stated above the regions are not necessarily at similar stages of economic growth and achievements and this needs to be taken into consideration in the study design and in selection of specific countries for comparisons. Even with the regions themselves large differences still endure with implications for generalisability and validity. One way to handle this would be to identify similar blocks of countries within a	The key limitation is that given the vast scope of this review it will be difficult to get into the details of accountability processes and procedures not only in countries but also in states and districts within countries. These details often are determinative for how well accountability provisions can best serve the public good. However, the categories for this study seem to be clearly conceptualised, some good	Overall, I'm more concerned about the quality of the data and the level of detail, and both those choices on timing and region seem likely to yield higher quality data. If there was a way to screen on the basis of the quality of the data, that would be good too, but I don't see an easy way to do that.

Question	de Grauwe	Kapichi	Shirley	Hatch
		region (e.g. Eastern and southern Africa for sub-Saharan Africa) and compare those with a bloc of countries in another region that exhibit similar challenges/opportunities, etc.	preliminary work has been done, and the final report should be a major contribution that should be quite impactful. Recognising that there are continent-wide issues in different regions, I wonder if it would be helpful to do some preliminary comparative analyses between and among countries to ascertain what kinds of accountability (or absence thereof) are evident. Can one come up with new ways of categorising or understanding cultures of accountability in schools and systems from this study? Are there some cases in which there might be weak governmental accountability but strong community or professional accountability for example?	
What additional sources (people,	The sources seem quite complete. I can think of two additional sources:	Addition sources: This includes a mix of research papers, working papers and publications ranging from cross	I always find comparative analyses between a small number of nations (or states within nations) with roughly similar circumstances to be	. I'm afraid that I don't have contacts in these parts of the world that would be helpful, but I assume that you will be in reach out to the usual

Question	de Grauwe	Kapichi	Shirley	Hatch
websites, centres, etc.) would you suggest we contact and/or include?	The McKinsey report on 'How the world's most improved school systems keep getting better' is well written, quite rich, and contains interesting insights in accountability. However, it looks at school systems, and not at schools. Nevertheless, it could be useful to check with the authors if they have any country reports (which were used for their study) that comment more specifically on school accountability. (http://www.mckinsey.com/client_service/social_sector/latest_thinking/worlds_most_improved_schools) The Global Partnership for Education is now a key actor in development. One of their foci is on 'learning outcomes'. It is good to include their website along the websites of interest: http://www.globalpartnership.org/ In addition, I could share with you very informally a series of documents which countries, who have	country studies to specific country analysis. See also attached -Guoxing Yu 'Research evidence of school effectiveness in sub-Saharan Africa' EdQUAL working paper no.7; University of Bristol, 2007. -Governance, management and accountability in secondary education in sub-Saharan Africa; World bank publication, 2008 -'Are our children learning?;- literacy and numeracy across East Africa- UWEZO report 2013. www.uwezo.net -URT (2008), A performance Audit report on school Inspection program for secondary schools in Tanzania. www.nao.go.tz -Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ)- Sacmeq 111 report. www.sacmeq.org	most helpful. There are different regional leaders (Chile, Singapore, South Africa) that could be studied to determine what they have done that allowed them to provide accountability measures that are atypical for their regions and could be disseminated more broadly.	contacts, such as any members of the agencies involved in inspection/accountability in these countries to find out what if any research and literature they may draw on.

Question	de Grauwe	Kapichi	Shirley	Hatch
	participated or are participating in a distance course on reforming school supervision which I am leading, have prepared or are preparing. Each country is asked to prepare a diagnosis of their school supervision system, including highlighting recent changes. These documents are not to be published or distributed, but I can share them with you, as they may help you identify interesting country case-studies, and for that purpose only. If necessary, you can afterwards contact the authors. Twelve countries participated in 2011, and 16 are taking part now. We also did a course in French, in 2012, with nine countries.	-Assessment and Education quality in sub-Saharan Africa: prospects and pitfalls- Public seminar on assessment- Oxford University – 29 October 2012- Tshwane University of Technology		
Which specific countries in these regions, in your opinion, would	I've mentioned three earlier, which I think to be interesting cases, because of their political and educational policy reforms: **Indonesia, Chile, and South-Africa**. I find it quite difficult at present to think of other useful	Specific countries - For sub-Saharan Africa I would recommend countries from the Southern and Eastern African countries in that they many of them share a similar history and education systems as former colonies	[Email] In general Chile has a very strong research capability and it is now a member of the OECD so it officially has emerged (I guess) into the world's developed economies, although there still is a lot of poverty	

Question	de Grauwe	Kapichi	Shirley	Hatch
provide the most insight?	cases. I should spend a bit of time looking at recent documents. Among countries that come to my mind, but for no precise reason, simply because they tend to be known as countries where reform has taken place, **are Sri Lanka, Ghana, and Uganda**.	of Britain; There has been a lot of comparative studies done in the region providing available and current robust data on education; Except for south Africa they are more or less sharing similar economic conditions and growth models and are sharing economic blocks	from what I've observed. Almost all of the scholars with whom I've worked have very good English and have spent some time in the US or UK. Beatrice Avalos, a leading teacher educator in Chile, is also a good contact. Otherwise Brazil and Mexico have both been posting strong economic growth in recent years and although the schools in Mexico are still of poor quality (see the recent OECD report led by Beatriz Pont of the OECD) the country's growing economy is a sign of hope. They also are transitioning now to a more market-oriented system, importing ideas and strategies from the US and UK. Other jurisdictions provide points of comparison, including Porto Alegre in the south of Brazil, which has not done so well on testing but has many important experiments underway in	

Question	de Grauwe	Kapichi	Shirley	Hatch
			participatory democracy and education.	
Do you have any other thoughts or suggestions about how we might effectively limit the scope of our review given available resources and time?	You intend to keep the three elements – monitoring, inspection, and assessment. But it could be argued that inspection and assessment are monitoring tools, quite specific ones, and that the term 'monitoring' is too broad to be useful. In other words, it may make sense to focus only on 'inspection/supervision' and 'assessment', which are somewhat contrasting forms of monitoring (qualitative versus quantitative; involving school contact vs distant; containing some advice vs no advice).	Narrowing the review further: If we have to narrow down the number of countries then for Sub-Saharan Africa I would recommend the 5 eastern African countries- Tanzania, Kenya, Uganda, Rwanda and Burundi. They share a rich history and a similarity of education systems (Tanzania, Kenya and Uganda). Rwanda and Burundi are new comers and they have recently changed their systems from Francophone to Anglophone models.	There is a large and growing body of literature now arguing for market-driven models of accountability, essentially contending that weak and corrupt states cannot or will not develop rigorous accountability systems. These are not only driven by neoliberal agendas. Since these concerns are increasingly vocal in policy circles, I wonder if studying some of the states that are often brought up as the most corrupt and dysfunctional could be helpful, and then to compare these with countries with reasonably intact accountability systems, to better ascertain where the institutional breakdowns occur?	You could conceivably exclude studies from the higher income countries, but you might lose valuable information that way.

APPENDIX 2.5: CODING TOOL

School accountability systematic review: Draft coding tool, v.7

Review Question: 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.

Source reference:

Coded by:

CODING A: Source appraisal		

Relevance
 	
Focus: Does the document focus on one or more accountability element(s) or on accountability in general?
	a) Accountability in general
	
	b) Monitoring
	
	c) Inspection/supervision
	
	d) Assessment
	
	d) More than one element (Please specify)
		The study as a whole is:
		
		    highly relevant
		
		    somewhat relevant
		
		    not relevant
		
		to our focus on accountability in this review.

Theory-building & comparison:
 	
To what degree, does the document address (elaborate and/or contradict) some aspect of the initial rough theory that we are testing?
	Aspects of the initial rough theory elaborated and/or contradicted by this research:	The research in this study is, on the whole:
		
		    highly relevant
		
		    somewhat relevant
		
		    not relevant
		
		to expanding our understanding of the initial rough theory.

CODING B Interventions, policies, programmes – If Relevance focus score = 3 (not relevant), stop here.		
Policy or Program name Code the name of the program if specified	Please specify Not stated	

Identifying interventions What is/are the specific accountability-related intervention(s) discussed in the document?	
Summarise programme hypotheses/theories Using the questions indicated, characterise the accountability-related intervention(s) addressed.	What/why? – What accountability-related work is involved and for what reasons? What role are policy initiatives intended to play and why? What specific programme actions relate to intended and/or actual outcomes noted above? What implicit and explicit rationales are given for those actions and why? Who? – looking at dimensions around the people who are the focus of accountability-related initiatives at the level of the individual, team, organisation (e.g. the people who become developed) and so on – so it's 'who' in a specific sense (e.g. educators, students) as well as in a collective sense (e.g. teams, organisations). This idea includes the concepts of leadership, culture and context. By whom? – looking at the dimensions involved in the people doing the developing; for example, looking at their orientation to the people that they are working with (insider/outsider etc.) Any other key characteristics of the accountability-related work discussed?
Summarise intended programme Theories/hypothesis (How) How? – Elaborate on what/why above. Compare/contrast how programme is intended to work and how it does work (espoused vs. enacted) looking at: mechanisms intended and actual mechanisms involved, facilitation styles, essential theoretical orientations, how knowledge is used, how users are involved, etc.) Consider integrity of implementation, unintended effects, etc.	Please specify Pay particular attention to 'How?' For what aspect(s) of the intervention does this research provide evidence for how the programme works in practice? What evidence supports, refutes or refines the espoused (intended) programme theory? Include a brief summary of the nature of the evidence and page numbers if appropriate. Not stated
CODING C: C-M-O Outcomes	

Outcomes assessed – Service delivery 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 practice.	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. Please code outcomes as described in the document, noting any correspondence with items listed above.
Outcomes assessed – System efficiency Refers to whether school and system-level processes deliver school education services effectively and efficiently.	This may include, but are not limited to: Cost/expenditure; Access; Equity Please code as described in the document, noting correspondence with items listed above.
Outcomes – Learning outcomes	This may include, but are not limited to: enrolment; attendance; retention; year repetition; completion rate; attainment; labour market participation Please code as described in the document, noting correspondence with items listed above.
Outcomes – Other	Please note any outcomes mentioned that do not fit in categories above.
CODING C: C-M-O Mechanisms of action & Intervention chains	
Key mechanisms: With reference to your answers to the 'How' question in Coding B, Summarise Programme Theories, 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 in the document	a) Setting expectation b) Providing feedback/consequences c) Institutionalisation of norms d) Capacity development of educators e) Capacity development of local stakeholders f) Others (Please code as described in the document) g) Not stated

Scope of action/stratification: What is the level of action indicated for the mechanisms identified? Select all that apply.	individual teams within schools school provincial national regional
CODING C: C-M-O Conditions	
Conditions – pre-existing conditions/context Code explicit statements by the authors that identify conditions/contextual influences that caused mechanisms identified above to fire or not to fire. PLEASE REFER TO PROTOCOL p.14	a) Please specify Political, economic, cultural, power relations, participation features of intervention implementation that affected whether and how the program generated outcomes b) Not stated
CODING D: Study design	
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)	Quantitative methods (a-g) a) Randomised controlled trial Each participant randomly has the same chance of being in the intervention and comparison group b) Non-randomised controlled trial/controlled before and after study Study includes intervention and comparison groups, with before and after data for both groups c) Retrospective controlled before and after study Data from large repeated surveys are used to retrospectively construct intervention and comparison groups, with before and after data for both groups d) Simple comparison study Intervention and comparison groups, only one data point also referred to as with and without study e) Before and after study One group of study before and after data

	f) Non comparison evaluation Only one data point - for example, post-test only, cross-sectional study g) Modelling study Based on theoretical/modelled events not real ones h) Qualitative methods For example, interviews, focus groups, observations i) Mixed methods design A study employs more than one methods above (a-h) j) Not empirical paper (e.g. discussion piece, policy brief, conceptual paper, statistics document)
What are the broad aims of the study? Please write in authors' description	Please specify (as stated by authors)
Schooling level	a) Primary b) Secondary c) Both
Location of school (sampling)	Rural Urban Not stated
What is/are the population focus/foci of the study?	Learners Head teacher / Senior management Teaching staff Teachers as learners Non-teaching staff Inspectors

	Other education practitioners
	Government
	Local education authority officers
	Parents
	Governors
	Community leaders
	Other
If learners are the population focus of the study, what were characteristics of learners in the study?	Male only
	Female only
	Mixed sex
	Low SES
	High SES
	Living in urban
	Living in rural
	Others
	Not stated
	No students participated in the study
Sample size	Please specify
Please describe data collection methods	One-to-one interview
	Group interview/focus groups
	Survey including household survey or routine data collection
	Observation
	Field note
	School records (attendance records etc.)
	Curriculum-based assessment/exam

	Others (Please specify)
Please describe data analysis methods	Please specify
Rigour: To what degree does the document support conclusions drawn from it by the researchers or the reviewers? Comments on whether methods used to draw inferences make a methodologically credible contribution to theory building, testing and/or comparison:	In general, the design, conduct, and reporting of research is – 1) High 2) Medium 3) Low
In what ways does the information extracted from this document support, weaken, modify, supplement, reinterpret or refocus the initial rough theory outlined in the protocol?	Please specify
Sources for follow up: Whether a list of references of this study has been checked to identify potential includes?	Yes No
Other comments:	

APPENDIX 3.1: DETAILS OF STUDIES INCLUDED IN THE SYSTEMATIC MAP FOR STAKEHOLDERS' DISCUSSION

This systematic map was based on 275 papers included in the review as of 15 November, 2014. As the review progress and as part of the iterative review process, we further refined our scope and inclusion criteria, extracted and analysed data in depth. Characteristics of the final set of papers included in the review are presented in Chapter 3.

MAPPING OF ACCOUNTABILITY ELEMENTS BY YEAR, COUNTRY/REGION, AND INCOME LEVEL

CATEGORY 1: ACCOUNTABILITY ELEMENTS BY YEAR (N=275)*

Table A3.1.1: Accountability elements in year increments, 1990-present

Date range	Accountability in general	Monitoring	Inspection/ supervision	Assessment
1990-1995	7	13	6	11
1996-2000	5	11	9	15
2001-2005	19	28	13	26
2006-2010	49	45	35	36
2011-present	36	22	18	29

Figure A3.1.1: Accountability elements in year increments, 1990-present*

Accountablity elements by Year
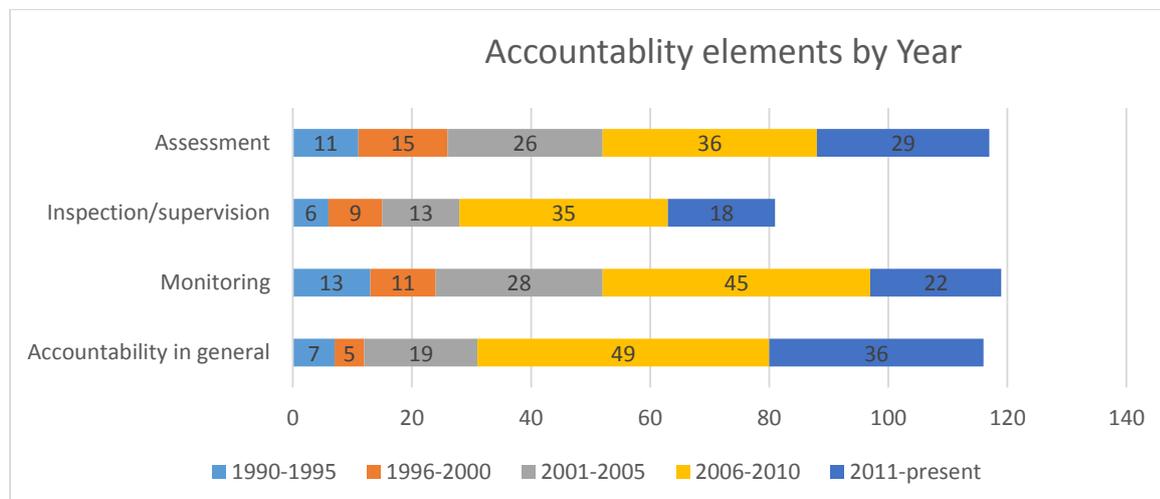
* Numbers do not total as a single document may reference multiple elements.

CATEGORY 2: ACCOUNTABILITY ELEMENTS BY COUNTRY INCOME (N=275)*

Table A3.1.2: Accountability elements by income

Income level	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Low income	15	24	15	14
Lower-middle income	27	25	24	18
Upper-middle income	34	31	21	46
Unclassified	29	31	20	26

Figure 2: Accountability elements by income*
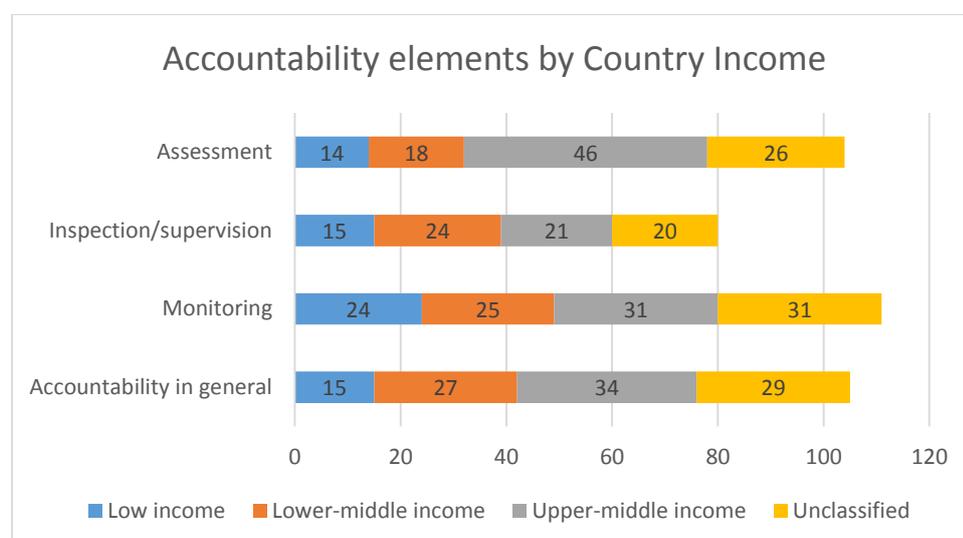
* Numbers do not total as a single document may reference more than one country.

CATEGORY 3: ACCOUNTABILITY ELEMENTS BY REGION (N=275)*

Table A3.1.3: Accountability elements by region

Region	Accountability in general	Monitoring	Inspection/ supervision	Assessment
East Asia and Pacific	12	15	10	17
Europe and Central Asia	2	3	9	7
Latin America and the Caribbean	25	21	7	15
Middle East and North Africa	3	3	0	8
South Asia	17	13	11	6
Sub-Saharan Africa	30	37	25	34

Region	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Unclassified	29	31	20	26

Figure A3.1.3: Accountability elements by region*

Accountability elements by Region
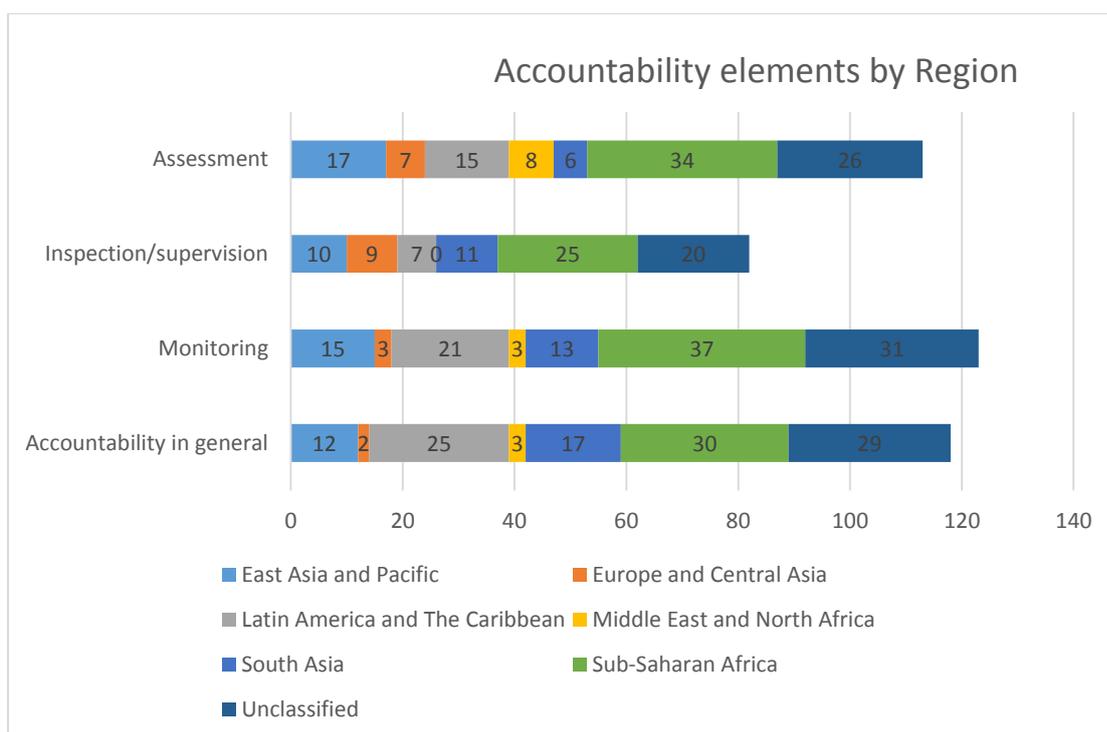
* Numbers do not total as a single document may reference multiple elements.

CATEGORY 4: 28 DFID PRIORITY COUNTRIES

Table A3.1.4: Accountability elements by DfID priority country*

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Bangladesh	6	6	2	1
Ghana	4	3	4	3

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
India	6	5	4	1
Kenya	2	4	4	3
Kyrgyz Republic	0	1	1	1
Liberia	1	1	1	1
Malawi	0	0	0	1
Nepal	0	2	0	1
Nigeria	1	3	2	3
Pakistan	5	1	5	2
Palestine	0	1	0	0
Sierra Leone	0	0	0	1
South Africa	11	7	3	14
Tajikistan	0	1	1	1
Uganda	3	6	3	1
Afghanistan	0	0	0	0
Burma	0	0	0	0
Congo	0	0	0	0
Ethiopia	0	0	0	0

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Mozambique	0	0	0	0
Rwanda	0	0	0	0
Somalia	0	0	0	0
South Sudan	0	0	0	0
Sudan	0	0	0	0
Tanzania	0	0	0	0
Yemen	0	0	0	0
Zambia	0	0	0	0
Zimbabwe	0	0	0	0

*No. of studies = 85 as of 25 September 2014 from 15 countries. No studies were available for countries in the blue rows.

Figure A3.1.4: Accountability elements by DfID priority countries*
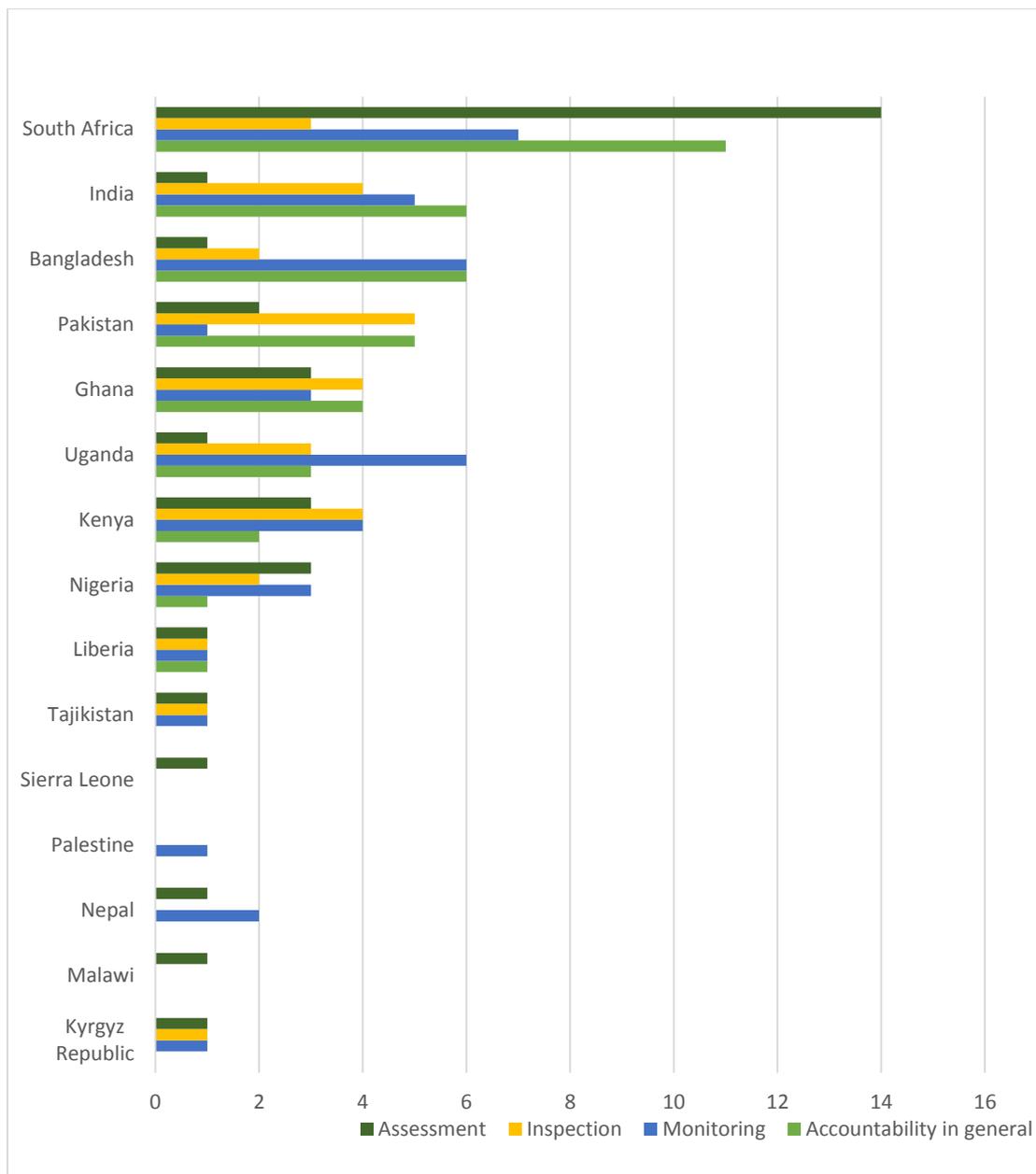
* Numbers do not total as a single document may reference multiple elements.

CATEGORY 5: LOW-INCOME COUNTRIES AND ACCOUNTABILITY ELEMENTS

Table A3.1.5: Accountability elements by low-income country

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Bangladesh*	6	6	2	1
Cambodia	0	2	0	2
Gambia	0	0	1	0
Guinea	1	0	0	0
Kenya*	2	4	4	3
Liberia*	1	1	1	1
Madagascar	1	1	1	0
Malawi*	1	0	1	1
Nepal*	0	2	0	1
Sierra Leone*	0	0	0	1
Tajikistan*	0	1	1	1
Tanzania*	0	2	1	1
Uganda*	3	6	3	1
Zimbabwe*	1	1	2	1

*Countries in which DfID works

Figure A3.1.5: Accountability elements by low-income country
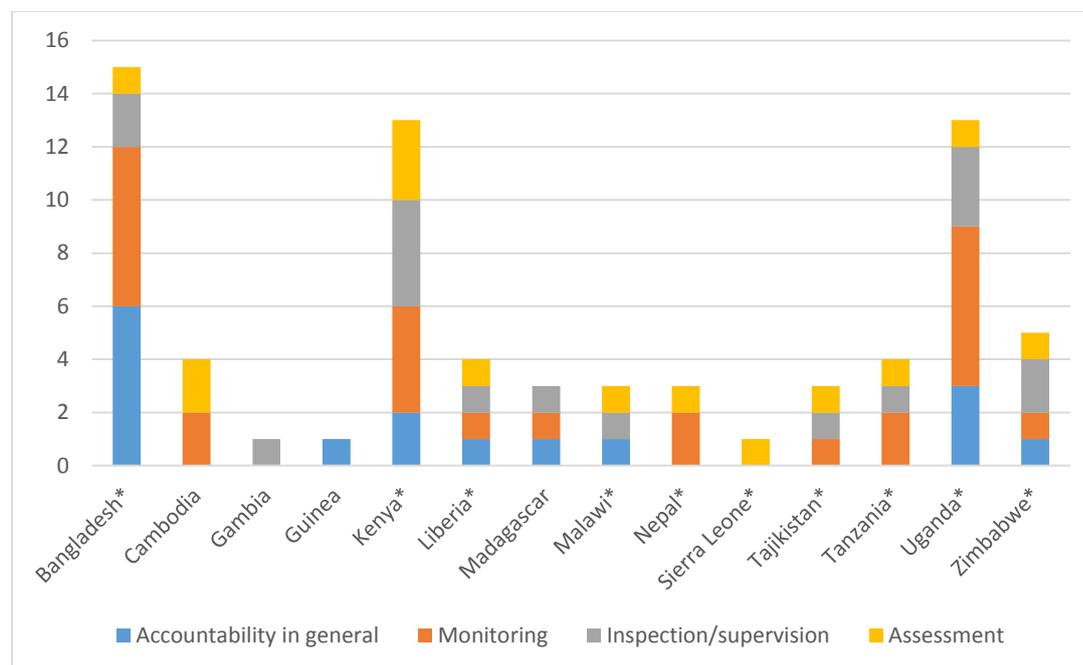
Figure A3.1.5: Accountability elements by low-income country

Table A3.1.6: Accountability elements by lower-middle-income country

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Armenia	0	0	0	1
Cameroon	1	0	1	0
Egypt	2	1	0	2
El Salvador	2	2	0	0
Ghana*	5	3	5	3
Guatemala	1	0	0	0

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
India*	6	5	4	1
Indonesia	5	6	5	1
Kyrgyz Republic*	0	1	1	1
Lao	0	1	1	0
Nigeria*	1	3	2	3
Pakistan*	5	1	5	2
Papua New Guinea	2	1	0	0
Philippines	0	1	0	1
Senegal	1	0	1	0
Sri Lanka	1	1	0	0
Uzbekistan	0	1	1	1
Vietnam	0	2	0	2
West Bank and Gaza	0	1	0	0
Zambia*	0	1	1	1

*Countries in which DfID works

Figure A3.1.6: Accountability elements by lower-middle-income country
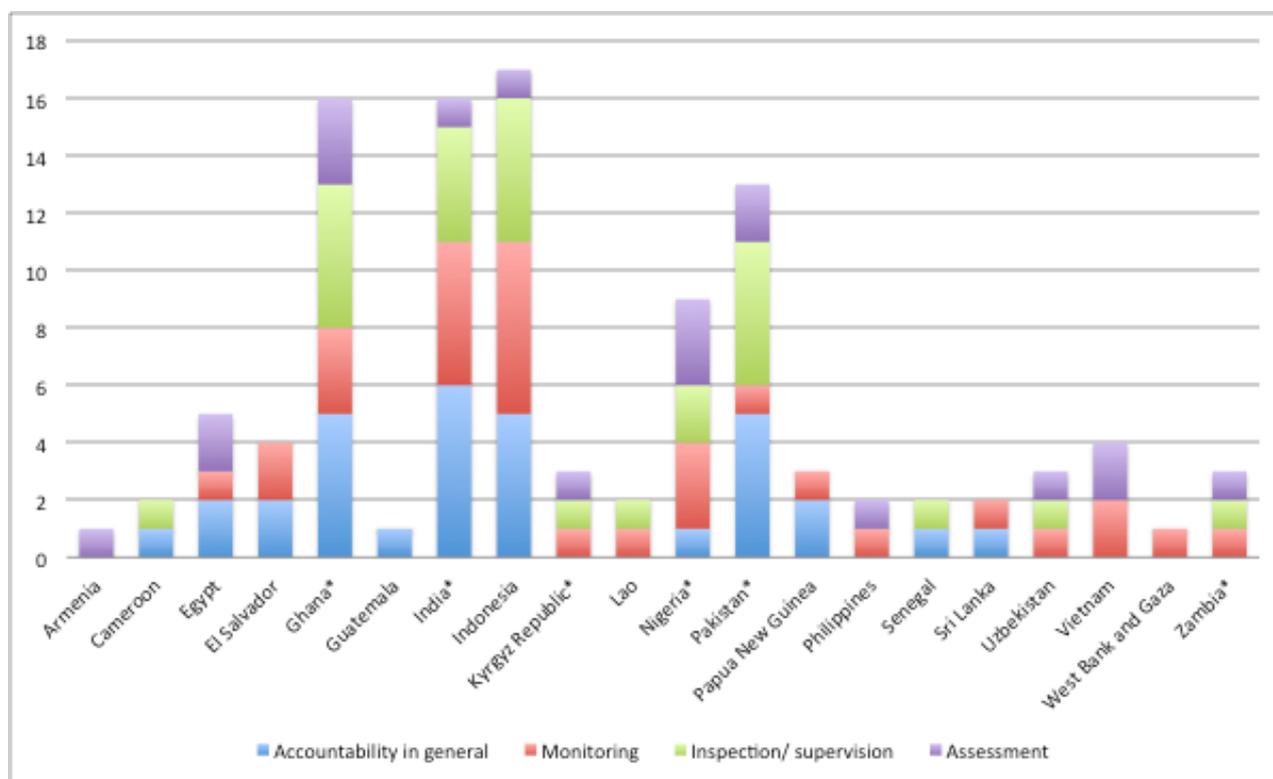
CATEGORY 7: UPPER-MIDDLE-INCOME COUNTRIES AND ACCOUNTABILITY ELEMENTS

Table A3.1.7: Accountability elements by upper-middle-income country

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Albania	1	1	1	0
Argentina	2	2	1	1
Azerbaijan	0	1	1	1
Botswana	1	2	1	2
Brazil	3	2	0	2
China	4	1	3	9

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Dominican Republic	1	1	0	0
Ecuador	1	2	1	1
Hungary	0	0	2	0
Jamaica	2	3	0	1
Jordan	1	0	0	3
Kazakhstan	0	1	1	1
Lebanon	0	0	0	1
Macedonia	0	0	0	1
Malaysia	0	0	0	1
Maldives	0	0	0	1
Mauritius	0	1	0	0
Mexico	5	3	1	3
Namibia	0	0	1	0
Peru	3	4	2	0
Romania	0	1	2	2
South Africa*	11	7	3	14
Thailand	2	2	0	1

Country	Accountability in general	Monitoring	Inspection/ supervision	Assessment
Tunisia	1	1	0	1
Turkey	0	0	3	2

*Countries in which DfID works

Figure A3.1.7: Accountability elements by upper-middle-income country
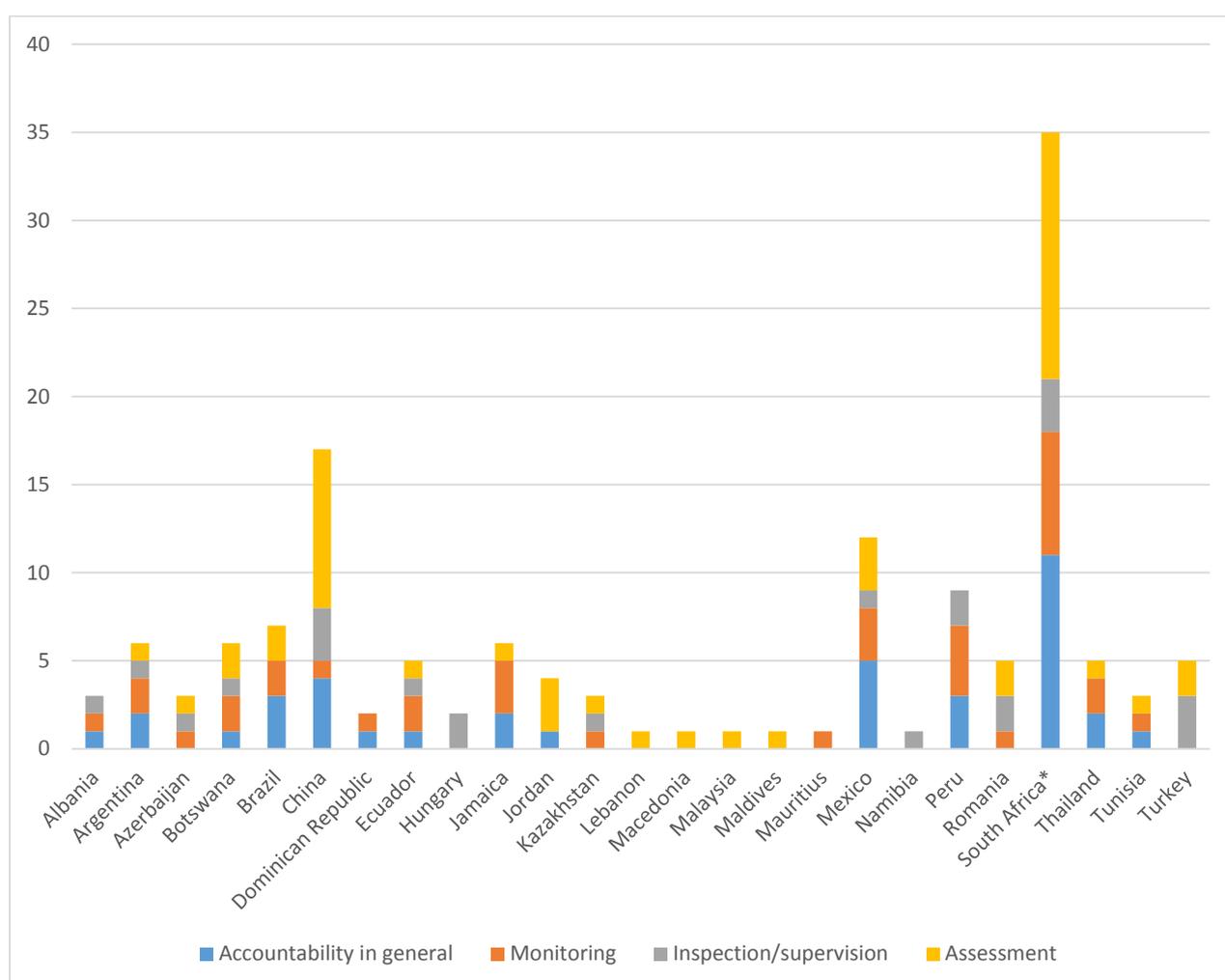
APPENDIX 4.1: DETAILS OF STUDIES INCLUDED IN THE IN-DEPTH REVIEW

MAIN CHARACTERISTICS OF IN-DEPTH REVIEW PAPERS AND C-M-O FIRST-LEVEL CODING (N=68)

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
ADEA (2001)	Accountability in general Monitoring Assessment	Sub-Saharan Africa Schooling level: Primary and secondary	Literature review	Low rigour	Level of government commitment to quality education services Existing programmes Education reform Perceptions of assessment Decentralisation/ centralisation Political structure Role of donor organisations Community involvement International or external influence Public perception	Setting expectation Providing Institutionalisation The notion of improvement Belief concerning educational assessment Demand for data

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Technological infrastructure Focus on user needs Pilot or experimental implementation Holistic implementation Quality of data	
Alcazar et al. (2006)	Inspection	Peru Schooling level: Primary	Mixed methods design	High rigour	Lack of resources	Top-down monitoring
Andrabi et al. (2013)	Monitoring Assessment	Pakistan Schooling level: Primary	Quantitative: Simple comparison study	High rigour	Pilot or experimental implementation Private schools	The notion of improvement School report cards
Attfield and Vu (2013)	Monitoring	Vietnam	Case studies	Medium rigour	Level of government commitment to quality education services	Setting expectation Providing feedback/consequences

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
		Schooling level: Primary			Existing programmes Socio-economic inequality Good access to effective, reasonable schooling	Capacity development of educators
Bansilal (2011)	Assessment	South Africa Schooling level: Secondary	Case studies	Medium rigour	Level of government commitment to quality education services Socio-economic inequality Education reform Perceptions of assessment Good access to effective, reasonable schooling	Setting expectation Capacity development of educators Understand the need of learners Belief concerning educational assessment
Barr et al. (2012)	Monitoring	Uganda Schooling level: Primary	Quantitative: Randomised controlled trials (RCTs)	High rigour	Good access to effective, reasonable schooling Community involvement	Setting expectation Providing feedback/ consequences Community-based monitoring

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Student enrolment Teacher absenteeism	
Barrera-Osorio and Raju (2010)	Accountability in general Assessment	Pakistan Schooling level: Primary	Quantitative: Non-randomised controlled trial/ controlled before and after	High rigour	Socio-economic inequality Good access to effective, reasonable schooling	Providing feedback/ consequences
Barrett (2011)	Accountability in general Monitoring Inspection	Developing countries Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Socio-economic inequality Role of donor organisations Cultural values International or external influence	Poor quality of tests and of test administration Institutionalisation of norms
Beets and van Louw (2011)	Assessment	South Africa	Non-empirical paper	Medium rigour	Lack of resources	Providing feedback/ consequences Understand the need of learners

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
		Schooling level: Primary and secondary			Good access to effective, reasonable schooling Difficult to access to school Technological infrastructure Teacher training programmes Teacher education level attained Assessment policies Structure of an education system	
Braun et al. (2006)	Assessment	Developing countries Schooling level: Secondary	Literature reviews	Medium rigour	Lack of resources Good access to effective, reasonable schooling Difficult to access to school Technological infrastructure	Communication Participation in assessment

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Teacher training programmes Teacher education level attained Assessment policies Structure of an education system	
Brock (2009)	Monitoring Inspection	China Schooling level: Primary and secondary	Case studies	Medium rigour	Education reform	Setting expectations Understand the need Empowerment
Brown et al. (2011)	Accountability in general Assessment	China Schooling level: Primary and secondary	Quantitative: Simple comparison study Case studies	High rigour	Long traditions of high-stakes examination Perceptions of assessment	Setting expectation Providing feedback/ consequences Institutionalisation of norms Understand the need of learners

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
						The notion of improvement Belief concerning educational assessment
Bruns et al. (2011)	Accountability in general Monitoring	Developing countries Schooling level: Primary and secondary	Literature review	High rigour	Level of government commitment to quality education services Socio-economic inequality Education reform Decentralisation/ centralisation Political structure Good access to effective, reasonable schooling Corruption Community involvement	Providing feedback/ consequences Capacity development of local stakeholders Empowerment Community-based monitoring Focus of inspections Involving community Parents councils School report cards Monitoring Parental involvement

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Teacher qualifications and attitudes Teacher absenteeism Low student achievement Public perception Technological infrastructure Teacher-student interaction	Quality of data collection/ management/ analysis Advocacy activities Publicising educational conditions/ data/ findings
Caddell (2005)	Monitoring	Nepal Schooling level: Primary and secondary	Case studies	Medium rigour	Role of donor organisations Corruption Focus on user needs Availability of expertise Conflict within country	Setting expectation Demand for data Incentive to report data Role of donors
Castro and Tiezzi (2003)	Assessment	Brazil	Case studies	Medium rigour	Level of government commitment to quality education services	Capacity development of educators

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
		Schooling level: Secondary				Capacity development of local stakeholders
Chen (2011)	Accountability in general Monitoring Inspection	Indonesia Schooling level: Primary	Quantitative: Non-comparison evaluation	Medium rigour	Decentralisation/ centralisation	Setting expectation Providing feedback consequences
Chisholm and Wildeman (2013)	Accountability in general Monitoring Assessment	South Africa Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Socio-economic inequality Perceptions of assessment	Providing feedback/ consequences Understand the need of learners Belief concerning educational assessment
Churches and McBride (2013)	Inspection	Developing countries	Non-empirical paper	Low rigour	Availability of expertise	Setting expectation Providing feedback/ consequences Credibility of school inspectors

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
		Schooling level: Primary and secondary				Community-based monitoring
Crouch and Winkler (2008)	Accountability in general Monitoring Inspection Assessment	Uganda South Africa Schooling level: Primary and Secondary	Case studies	Low rigour	Level of government commitment to quality education services Education reform Long traditions of high-stakes examination Decentralisation/ centralisation Community involvement	Setting expectation The notion of improvement Credibility of school inspectors Pedagogical management
Darvas and Balwanz (2014)	Accountability in general Inspection	Ghana Schooling level: Primary and secondary	Quantitative: Simple comparison study	High rigour	Level of government commitment to quality education services Existing programmes Socio-economic inequality	Providing feedback/ consequences Community-based monitoring

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Decentralisation/ centralisation Good access to effective, reasonable schooling Size of the country Cultural values Community involvement Teacher qualifications and attitudes Student enrolment Teacher absenteeism	
De Grauwe (2001)	Inspection	Namibia Zimbabwe Tanzania Schooling level: Primary	Mixed methods	High rigour	Level of government commitment to quality education services Education reform Decentralisation/ centralisation	Capacity development of educators Capacity development of local stakeholders Credibility of school inspectors

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
		and secondary			Lack of resources	
De Grauwe (2007)	Inspection	Developing countries Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Existing programmes Lack of resources Teacher qualifications and attitudes	Providing feedback/consequences
De Grauwe (2008)	Accountability in general Monitoring Inspection Assessment	Developing countries and high-income countries Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Lack of resources Political structure Difficult to access to school Teacher qualifications and attitudes	Setting expectation Providing feedback/consequences Institutionalisation of norms Capacity development of educators Capacity development of local stakeholders Community-based monitoring Centralisation Resistance to external evaluation

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
De Grauwe et al. (2007)	Monitoring	Developing countries Schooling level: Primary and secondary	Literature review	Low rigour	Socio-economic inequality Decentralisation/ centralisation Political structure Role of donor organisations Conflict within country Private schools	Providing feedback/ consequences Capacity development of local stakeholders Quality of data collection/ management/ analysis EMIS
Ferrer (2006)	Assessment	Latin America Schooling level: Primary and secondary	Case studies	Medium rigour	Level of government commitment to quality education services Socio-economic inequality Perceptions of assessment Decentralisation/ centralisation Lack of resources Role of donor organisations	Setting expectation Providing feedback/ consequences Institutionalisation of norms Capacity development of educators Capacity development of local stakeholders Belief concerning educational assessment

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
						Demand for data
						Centralisation
						Communication
						Participation in assessment
Glewwe et al. (2010)	Monitoring Assessment	Kenya Schooling level: Primary	Quantitative: Randomised controlled trials (RCTs)	High rigour	Perceptions of assessment Teacher absenteeism Teacher incentives Teacher salaries School norms Teacher surplus/shortage	Financial incentives - teachers
Gvirtz (2002)	Assessment	Latin America Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Decentralisation/ centralisation	Not stated

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
Gvirtz and Larripa (2004)	Accountability in general Monitoring Assessment	Argentina Schooling level: Primary and secondary	Literature reviews	Medium rigour	Existing programmes Perceptions of assessment Decentralisation/ centralisation Role of donor organisations Objectives of evaluation system Consequences for performance	Providing feedback/consequences Belief concerning educational assessment Centralisation Poor quality of tests and of testing Participation in assessment Poor use of test results
Harber (2006)	Inspection	Gambia Schooling level: Primary and secondary	Qualitative methods	Medium rigour	Existing programmes	Providing feedback/consequences The notion of improvement Credibility of school inspectors
Herselman and Hay (2002)	Inspection	South Africa Schooling level: Primary	Case studies	Medium rigour	Level of government commitment to quality education services Socio-economic inequality	Setting expectation Institutionalisation of norms Capacity development of educators

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Education reform	The notion of improvement
Higgins and Rwanyange (2005)	Accountability in general Monitoring	Uganda Schooling level: Primary and secondary	Literature review	Medium rigour	Level of government commitment to quality education services Existing programmes Education reform Role of donor organisations Difficult to access to school Cultural values Community involvement Teacher qualifications and attitudes Student enrolment International or external influence	Setting expectation Providing feedback/ consequences Capacity development of local stakeholders Rush to implement policies Centralisation

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Financing education Low student achievement Student attendance Public perception	
Howie (2012)	Accountability in general Assessment	South Africa Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Level of government commitment to quality education services Socio-economic inequality Good access to effective, reasonable schooling	Providing feedback/ consequences Understand the need of learners The notion of improvement
Jaffer (2010)	Accountability in general Inspection	Pakistan Schooling level: Primary and Secondary	Mixed methods design	Low rigour	Existing programmes Lack of resources Political structure	Providing feedback/consequences

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
Kapambwe (2010)	Monitoring Assessment	Zambia Schooling level: Secondary	Quantitative: Simple comparison study	Medium rigour	Level of government commitment to quality education services	The notion of improvement
Kellaghan and Greaney (2001)	Assessment	Developing countries Schooling level: Primary and secondary	Non-empirical paper	Medium rigour	Level of government commitment to quality education services	Setting expectation Providing feedback/consequences Empowerment Participation in assessment
Kellaghan and Greaney (2004)	Assessment	Sub-Saharan Africa Schooling level: Primary and secondary	Non-empirical paper	Medium rigour	Education reform Perceptions of assessment	Belief concerning educational assessment
Kingdon and Muzammil (2012)	Accountability in general Inspection	India Schooling level: Primary	Quantitative: Secondary data analysis and	Medium rigour	Political structure	Empowerment

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
	Assessment		non-comparison evaluation			
Kremer et al. (2004)	Monitoring Assessment	Kenya Schooling level: Primary	Quantitative: Randomised controlled trials (RCTs)	High rigour	Long traditions of high-stakes examination Cultural values Student enrolment Financing education Natural disaster	Financial incentives - students Parental involvement Publicising education conditions/ data/ findings
Lassibille et al. (2010)	Monitoring Assessment	Madagascar Schooling level: Primary	Quantitative: Randomised controlled trials (RCTs)	High rigour	Level of government commitment to quality education services Education reform Political structure Teacher qualifications and attitudes	Setting expectation Capacity development of educators Capacity development of local stakeholders Communication Monitoring or supervision visits Involving community

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Student enrolment Teacher absenteeism Financing education Low student achievement Poverty levels School norms Teacher surplus/shortage	School report cards
Lubisi and Murphy (2002)	Assessment	South Africa Schooling level: Primary and secondary	Case studies	Low rigour	Level of government commitment to quality education services Socio-economic inequality Perceptions of assessment Good access to effective, reasonable schooling	Providing feedback/ consequences Institutionalisation of norms Capacity development of educators

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
Luxia (2005)	Assessment	China Schooling level: Secondary	Quantitative: Non comparison evaluation	Medium rigour	Long traditions of high-stakes examination	Setting expectation Providing feedback/consequences
Macpherson (2011)	Accountability in general Inspection	Indonesia Schooling level: Primary and Secondary	Case studies	Low rigour	Difficult to access school Corruption	Capacity development of educators
Mazibuko (2007)	Inspection	South Africa Schooling level: Primary and Secondary	Qualitative methods	Medium rigour	Level of government commitment to quality education services Education reform	Providing feedback/consequences Institutionalisation of norms Capacity development of educators The notion of improvement

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
Moswela (2010)	Accountability in general Inspection	Botswana Schooling level: Secondary	Qualitative methods	Medium rigour	Decentralisation/ centralisation Difficult to access to school	Credibility of school inspectors
Mukhopadhyay and Sriprakash (2011)	Assessment	India Schooling level: Primary and secondary	Qualitative methods	Low rigour	Political structure Role of donor organisations	Setting expectation Providing feedback/ consequences Institutionalisation of norms Capacity development of educators Empowerment
Muralidharan and Sundararaman (2011)	Monitoring Assessment	India Schooling level: Primary	Quantitative: Randomised controlled trials (RCTs)	High rigour	Good access to effective, reasonable schooling Student enrolment Teacher absenteeism Low student achievement Consequences for performance	Financial incentives - teachers

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Teacher salaries	
Murimba (2005)	Monitoring	Africa	Non-empirical paper	Low rigour	Education reform	Belief concerning educational assessment
Nsibande and Modiba (2012)	Assessment	Africa Schooling level: Secondary	Qualitative methods	Medium rigour	Perceptions of assessment	Providing feedback/ consequences Understand the need of learners Capacity development of educators
Ong (2010)	Assessment	Malaysia Schooling level: Primary and secondary	Non-empirical paper	Medium rigour	Level of government commitment to quality education services Long traditions of high-stakes examination Perceptions of assessment Decentralisation/ centralisation	Providing feedback/ consequences Capacity development of educators Communication

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
Opoku-Asare (2006)	Inspection	Ghana Schooling level: Primary and secondary	Qualitative methods	Medium rigour	Existing programmes	Providing feedback/ consequences Capacity development of educators Credibility of school inspectors
Postlethwaite (2004)	Accountability in general Monitoring Assessment	Vietnam Kenya Schooling level: Primary and secondary	Literature reviews	Medium rigour	Socio-economic inequality Education reform Lack of resources Role of donor organisations Good access to effective, reasonable schooling Student enrolment Low student achievement Teacher training programmes	Providing feedback/ consequences Monitoring Quality of data collection/ management/ analysis

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Teacher education level attained	
Powell (2006)	Accountability in general Monitoring	Developing countries Schooling level: Primary and secondary	Case studies	Medium rigour	Role of donor organisations International or external influence Quality of data collection system	Belief concerning educational assessment Demand for data Centralisation Poor quality of tests and of test administration Quality of data collection/ management/ analysis EMIS Financial incentives - schools
Prew and Quaigrain (2010)	Monitoring	Ghana Schooling level: Primary and secondary	Case studies	High rigour	Difficult to access to school	Providing feedback/ consequences Empowerment Demand for data

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
Pryor and Lubisi (2002)	Assessment	South Africa Schooling level: Primary	Case studies	Low rigour	Education reform Long traditions of high-stakes examination Perceptions of assessment Lack of resources Cultural values	Institutionalisation of norms Capacity development of educators
Ravela (2002)	Assessment	Latin America Schooling level: Primary and secondary	Literature reviews	Medium rigour	Access to information	Providing feedback/ consequences
Ravela et al. (2001)	Monitoring Assessment	Latin America Schooling level: Primary and secondary	Non-empirical paper	Medium rigour	Existing programmes Education reform Perceptions of assessment Role of donor	Belief concerning educational assessment

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
					Community involvement Public perception Availability of expertise Objectives of evaluation system Quality of data collection system	
Reyneke et al. (2010)	Assessment	South Africa Schooling level: Secondary	Quantitative: Non comparison evaluation	Medium rigour	Education reform Perceptions of assessment Lack of resources School-based networks	Capacity development of educators Understand the need of learners Pedagogical management
Santiago et al. (2012)	Accountability in general Monitoring Inspection	Mexico Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Existing programmes Availability of expertise	Setting expectation Providing feedback/ consequences

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
	Assessment					
Scherman et al. (2011)	Accountability in general Monitoring Assessment	South Africa Schooling level: Secondary	Quantitative: Non-comparison evaluation	Medium rigour	Education reform	Setting expectation Institutionalisation of norms
Taylor (2009)	Accountability in general Monitoring Assessment	South Africa Schooling level: Secondary	Quantitative: Non comparison evaluation	Medium rigour	Level of government commitment to quality education services Socio-economic inequality Education reform	Providing feedback/ consequences Capacity development of educators The notion of improvement Belief concerning educational assessment
USAID (2006)	Monitoring	Developing countries Schooling level: Primary and secondary	Systematic reviews	Medium rigour	Decentralisation/ centralisation Community involvement	Providing feedback/ consequences Capacity development of educators The notion of improvement Community-based monitoring

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
						Circuit support teams Involving community Parents councils School report cards
USAID (2007)	Monitoring	Indonesia Schooling level: Primary and secondary	Qualitative methods	Medium rigour	Decentralisation/ centralisation Role of donor organisations	Setting expectation Providing feedback/ consequences Capacity development of local stakeholders Demand for data
Uwazi (2009)	Inspection	Africa Schooling level: Secondary	Qualitative methods	Medium rigour	Lack of resources	Providing feedback/ consequences Capacity development of educators Communication Focus of inspections

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
Wanzare (2002)	Inspection	Kenya Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Lack of resources	Setting expectation Providing feedback/consequences Capacity development of local stakeholders Capacity development of educators Credibility of school inspectors
Winkler and Herstien (2005)	Montoring	Nigeria	Policy-relevant document	Low rigiour	Decentralisation	Community involvement Information use
Winkler (2005)	Monitoring	Brazil Schooling level: Primary and secondary	Literature reviews	Low rigour	Socio-economic inequality Decentralisation/ centralisation Community involvement School norms	Understand the need of learners Parents councils School report cards

Study	Accountability elements	Location and setting	Study approach	Quality	Context	Mechanisms
World Bank (2008)	Accountability in general Monitoring	Sub-Saharan Africa Schooling level: Primary and secondary	Case studies Literature reviews Non-empirical paper	Medium rigour	Perceptions of assessment Decentralisation/ centralisation	Belief concerning educational assessment Credibility of school inspectors Community-based monitoring
World Bank (2010)	Monitoring	Sub-Saharan Africa Schooling level: Primary and secondary	Non-empirical paper	Low rigour	Education reform	The notion of improvement

Table A4.2.1: Assessment outcomes

A: High-stakes examinations – setting expectations					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Unintended shifts related to teachers, teaching and curricular focus					
Braun et al. (2006)	Developing countries	Expectations of high-stakes consequences/rewards for schools, school leaders and teachers for performance	Focus on academic disciplines, not practical subjects of most value to majority of learners	Med/Med	CL
Howie (2012)	South Africa	Threat of negative consequences for low-performing schools	Manipulation of examination process and results Exclusion of low-performing students	Low/Med	CL
Kellaghan and Greaney (2001)	Developing countries	Expectations of high-stakes consequences (unintended or intended) for schools, school leaders and teachers for performance	Focus on 'borderline' students Emphasis on selection of higher achieving students for admission	Med/Med	CL

A: High-stakes examinations – setting expectations					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
			Increased use of disability designation to exclude lower-performing students from exam Increased grade retention for low-achieving students		
		Threat of consequences (sanctions) for overall school performance	Students coached on assessment instruments Teaching focus on what is measured Curricular emphasis on assessed subjects Teaching focus on successful completion of items included in assessment instrument Emphasis on test performance not deeper learning		
		Promise of individual teacher bonus for student performance	Teacher recruitment and retention difficulties in low-performing schools		

A: High-stakes examinations – setting expectations					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Kellaghan and Greaney (2004)	Sub-Saharan Africa	Expectations of high-stakes consequences/rewards for schools, school leaders and teachers for performance	Focus on higher-achieving students Increased use of drill, rote memorisation, short-term strategies in teaching Adoption of short-term teaching strategies (e.g., multiple choice tests) in lower, non-examined, primary school grades Narrowing of curriculum in lower, non-examined grades	Med/High	CL
Luxia (2005)	China	Expectations around student examination performance of school and community	Teacher self-evaluation of performance in terms of student test results	Med/High	FI
Mukhopadhyay and Sriprakash (2011)	India	Promise of 'Learning Guarantee Award' based on school performance	Increased emphasis on successful test completion in regular classroom teaching Increased 'deficit assumptions' (negative teacher attitudes) towards students from lower castes	Low/High	FI
Unintended shifts related to student learning					

A: High-stakes examinations – setting expectations					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Kellaghan and Greaney (2004)	Sub-Saharan Africa	Expectations of high-stakes consequences/rewards for schools, school leaders and teachers for performance	Student use of short-term learning strategies Student emphasis on extrinsic reward Students' decreased motivation for mastering higher-order thinking skills	Med/High	CL
Intended shifts related to teachers, teaching and curricular focus					
Castro and Tiezzi (2003)	Brazil	Voluntary examination emphasis on higher-order thinking skills (ENEM)	Acceptance by schools, parents, students of legitimacy of examination Acceptance of examination results for admission to higher education Identifies what 'should be taught' Gives school leaders, teachers concrete image of desired performance	Med/Med	CL

A: High-stakes examinations – setting expectations					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Ferrer (2006)	Latin America	Use of high-stakes exam results for accreditation and incentives	High-performing schools not 'teaching to the test' but emphasising formative assessment and higher-order thinking	Med/Med	CL
Kellaghan and Greaney (2004)	Sub-Saharan Africa and Carribean	Kiswahili and practical subjects introduced to the Kenyan Certificate of Primary Education in the 1980s	Increase in coverage of assessed subjects (Kiswahili and practical subjects) despite lack of resources, textbooks, teacher competence	Med/High	CL
		Shifting emphasis in high-stakes assessment from multiple choice questions to essay writing in Trinidad and Tobago	Increase in writing tasks assigned by teachers Increase in teacher satisfaction with teaching Student exposure to written argumentation and problem solving		
		Changes in design of exam items from recognition of correct response to inference accompanied by provision of guidance to teachers and professional development	Greater emphasis on higher-order thinking skills in classroom teaching		

A: High-stakes examinations – setting expectations					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
		around understanding examination demands			

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

B. High-stakes examination: Feedback through dissemination of results					
Studies	**Countries**	**Intervention details**	**Outcome**	**Quality of study (Rig/Rel)***	**Type of evidence***
Ferrer (2006)	Latin America	Sophistication of results reporting formats for teachers and students	Utility for teaching, refining teaching strategy and focus Utility for students, identifying strengths and gaps	Med/Med	CL

B. High-stakes examination: Feedback through dissemination of results					
Studies	**Countries**	**Intervention details**	**Outcome**	**Quality of study (Rig/Rel)***	**Type of evidence***
Kellaghan and Greaney (2001)	Uganda	Workshops and seminars for teacher trainers, school inspectors and teachers including assessment results	Recognition of need to include all teachers Recognition of need for continuing support for teachers [No reported teacher or student outcomes]	Med/Med	CL
	Chile	Provision of pedagogical materials and professional development through school-based workshops to schools identified underperforming	Reduction of achievement gap between underperforming and other schools		
Kellaghan and Greaney (2004)	Swaziland	Teacher professional development and guidance around exam preparation	Teaching emphasis on higher-order thinking	Med/High	CL
Santiago et al. (2012)	Mexico	Implementation of ENLACE (National Assessment of Academic Achievement in Schools), a national evaluation and assessment framework, in Mexico	Lack of attention to facilitating use of available data at the school-level	Med/Med	CL

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

242

C: High-stakes assessment: Incentives as consequence					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Barrera-Osario and Raju (2010)	Pakistan	School- and teacher-level group incentives

Subsidy for schools reaching minimum student pass rate on the Quality Assurance Test

Competitive bonus for highest-ranking school

Group-based bonuses for teachers | Schools motivated to achieve minimum pass rate to stay in programme

Teachers not motivated to earn bonuses

No effective incentives to continuously raise standard of learning | High/High | FI |
| Glewwe et al. (2010) | Kenya | School-level group incentives | No effect on teacher attendance, teaching practice

Increase in classroom time, class work and homework devoted to exam preparation

Short-term gains in student results for programme schools not sustained after programme ends

No influence on rates of student dropout or retention | High/High | FI |

C: High-stakes assessment: Incentives as consequence					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Kremer et al. (2004)	Kenya	Merit scholarships for top-performing sixth-grade girls in schools in challenging circumstances	Increase in student school participation Increase in teacher attendance No increase in frequency of test preparation sessions Large test score gains Parents 'check up' on teachers to encourage more effort from students	High/High	FI
Mukhopadhyay and Sriprakash (2011)	India	Annual assessment, Karnataka School Quality Assessment Organisation (KSQAO), with cash incentives for high-performing schools; individual incentives for students and teachers	No additional effort to improve performance from adequately performing schools KSQAO perceived as targeting low-performing schools Increased test preparation, teaching focus on assessment items, sanctioned malpractice in low-performing schools System officials sceptical of results	Low/High	FI

C: High-stakes assessment: Incentives as consequence					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
			Teacher devaluation of potential of low-performing students (deficit assumptions)		
Muralidharan and Sundararamen (2011)	India	School-level group incentives versus teacher-level individual incentives	Individual incentive schools outperform group incentive after 2 years Increased teaching effort focused on exam preparation (homework, class work, after-hours sessions) Increased attention to lower-performing students No increase in teacher attendance.	High/High	FI

C: High-stakes assessment: Incentives as consequence					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Muralidharan and Sundararamen (2011)	India	School-level group incentives versus teacher-level individual incentives	Individual incentive schools outperform group incentive after 2 years Increased teaching effort focused on exam preparation (homework, class work, after-hours sessions) Increased attention to lower-performing students No increase in teacher attendance.	High/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

D: High-stakes assessment: Capacity development of educators through school-based performance assessments					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Bansilal (2011)	South Africa	Implementation of school-based component of national examination (SBA) in one school	Lack of provincial support Professional development disorganised and/or facilitated by an educator lacking adequate experience Context and framing of task predetermined by national design and not appropriate to level of student understanding	Med/High	FI
Beets and van Louw (2011)	South Africa	Implementation of school-based assessment (continuous assessment or CA) as component of national public examination	Teacher lack of understanding of underlying principles Teacher lack of competence in translating lesson objectives into assessment tasks	Low/Med	
Kapambwe (2010)	Zambia	Implementation of school-based Assessment (Continuous Assessment or CA) as component of public examination in pilot schools	Pupil performance in CA pilot schools significantly higher than control schools due to CA interventions.	Med/High	FI

D: High-stakes assessment: Capacity development of educators through school-based performance assessments					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Kellaghan and Greaney (2004)	Sub-Saharan Africa	Implementation of school-based assessment	No change in teacher competence Emphasis on summative, not formative, aspects of assessment Unable to overcome existing conditions (e.g., lack of resource materials, large class size) Practical difficulties in implementation result in failure or limited implementation	Med/High	CL
Lubisi and Murphy (2002)	South Africa	Implementation of school-based component of national examination (School-based assessment or SBA)	Teacher 'parroting' assessment tasks from guidance/previous examinations Lack of teacher guidance on implementation Teacher 'repackaging' items from previous examinations as SBA tasks	Low/High	CL

D: High-stakes assessment: Capacity development of educators through school-based performance assessments					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Nsibande and Modiba (2012)	Swaziland	Implementation of school-based component of national examination (School-based assessment or SBA)	Teacher lack of competence in translating lesson objectives into assessment tasks Teacher 'parroting' assessment tasks from guidance/previous examinations Emphasis on summative, not formative, aspects of assessment	Med/Med	FI
Ong (2010)	Malaysia	Implementation of school-based assessment as component of national public examination	Emphasis on summative, not formative, aspects of assessment Lack of internal and external support for implementation	Med/Med	
Pryor and Lubisi (2002)	South Africa	Implementation of school-based component of national examination (SBA)	Teacher professional development focus on procedural/bureaucratic aspects Teacher professional development does not reflect understanding of underlying principles Teacher alienation from interactive pedagogies	Low/High	FI

D: High-stakes assessment: Capacity development of educators through school-based performance assessments					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Reyneke et al. (2010)	South Africa	Implementation of school-based component of national examination (school-based assessment or SBA)	Teacher-set tasks lack validity, reliability Teacher lack of understanding of underlying principles Teacher perception of SBA as 'playing around' not learning Teacher perception of SBA as a technical procedure (compliance) unrelated to professional judgement Decreased time for other teaching, increased workload due to record-keeping Reinforcing teachers' 'deficit assumptions' about lower-achieving students SBA-inflated internal results lead students and parents to overestimate performance on external exam	Med/High	FI
Scherman et al. (2011)	South Africa	Implementation of school-based component of national examination	Current recording and reporting protocols lack validity – suggested as an explanation for uneven distribution of results	Med/Med	FI

D: High-stakes assessment: Capacity development of educators through school-based performance assessments					
Studies	**Countries**	**Intervention**	**Outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
		(SBA), analysis of performance levels for mathematics	Trial protocols confirmed very few if any pupils in upper levels of performance for mathematics		

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

E: Low-stakes assessment: Setting expectations: The establishment of curriculum standards					
Studies	**Countries**	**Interventions**	**Intermediate outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Ferrer (2006)	Colombia, Ecuador, Uruguay; subnational units in Aguascalientes, Mexico, and	Establishment of autonomous units for designing and implementing national assessment	Some impact on specifying curricular targets and concrete outcomes of learning emphasising higher-order skills (quasi-standards) Teacher involvement in validation of test items incites debate about national curriculum	Med/Med	CL

E: Low-stakes assessment: Setting expectations: The establishment of curriculum standards					
Studies	**Countries**	**Interventions**	**Intermediate outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
	Bogotá, Columbia				
Ferrer (2006)	Argentina	Establishment of autonomous units for designing and implementing national assessment	Undesirable impact on schools and educational improvement due to lack of co-ordination between assessment unit and education system	Med/Med	CL
Gvirtz (2002)	Argentina	Development of guidelines, Common Basic Contents (CBCs), to serve as basis for national assessment	Adoption of CBCs as de facto curriculum standard Development of textbook and guidance targeted at schools and classroom practice	Low/Med	CL
Kellaghan and Greaney (2004)	Namibia Eritrea	Implementation of National Learner Baseline Assessment (Namibia) Implementation of national assessment in Eritrea as part of Education for All 2000 Assessment	**Namibia:** Results highlight problematic aspects of curriculum Allocation of resources to low-performing schools **Eritrea:** Results highlight problematic aspects of curriculum	Med/High	CL

E: Low-stakes assessment: Setting expectations: The establishment of curriculum standards					
Studies	**Countries**	**Interventions**	**Intermediate outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
			Results indicate gender imbalance (boys outperform girls) Results highlight implications for teacher training and teaching strategies		
Kellaghan and Greaney (2001)	Brazil	Implementation of national assessments	Associated with curricular change, pedagogical innovation, influence on teacher training and financing schools in disadvantaged areas	Med/Med	CL
Kellaghan and Greaney (2001)	Thailand	Inclusion of diverse measures (affective outcomes, practical skills, social perception) in national assessment	Teaching strategies and assessment tasks reflect diversity of outcomes measured	Med/Med	

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

F: Low-stakes assessment: Capacity development of educators: Guidance and support for schools and teachers					
Studies	**Countries**	**Intervention**	**Intermediate outcomes**	**Quality of study (Rig/Rel)***	**Type of evidence***
Ferrer (2006)	Latin America	Public dissemination of school-level results with detailed analyses of student responses and comparison of results of schools in similar socioeconomic circumstances	[No outcomes reported]	Med/Med	CL
	Uruguay Dominican Republic Bolivia	Confidential dissemination of school-level results with detailed analyses of student responses and comparison of results of schools in similar socio-economic circumstances	[No outcomes reported]	Med/Med	CL
Ravela (2002)	Argentina	Publication of detailed analyses of student responses to national assessment	Identification of variation in proficiency in different content areas Concrete illustrations of mastery for teachers [No report of actual service delivery outcomes]	Med/Med	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

Table 4.2.1: Evidence per type of activity - high-stakes assessment

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
Braun et al. (2006)	Developing countries	To focus on academic disciplines to the exclusion of more practical subjects, such as typing or woodwork, that are of interest and value to substantial numbers of learners. Ideally, separate examinations should be set for different purposes, but this is usually not practical for developing nations. P.33	Medium rigour	CL
Castro and Tiezzi (2003)	Brazil	The ENEM [National High School Examination] has made it possible to gain a more palpable understanding of the pillars structuring secondary education reform: an interdisciplinary approach, putting learning into context and solving problems; it has allowed teachers and education specialists to visualise clearly the desired performance of young people, as is required by each of the subjects. In that sense, it is a powerful instrument to induce change insofar as it expresses what should be taught through what it assesses. One of the main results of this has been the acceptance of the voluntary exam by schools through teachers and students. The ENEM is now considered an important element to understand the competences of secondary school finalists and … the number of universities and other higher education institutions that	Medium rigour	CL

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
		make use of its results as a criterion for the selection of candidates for graduate study is increasing. P.14		
Ferrer (2006)	Latin America	Other research administered by the author in Latin American countries that use such high-stakes tests indicates that accreditation and incentives mechanisms do not necessarily mean that classroom efforts are reduced to 'teaching to the test.' In fact, many schools with high scores on standardised tests have made their own education and curricula proposals that place a marked emphasis on the integral, formative role of their students. P.50	Medium rigour	FI
Howie (2012)	South Africa	In response to some of the identified problems in the system, a multitude of national intervention strategies were initiated including the provision of learning and teaching materials, increased monitoring, targeted support programmes for schools, targeted guidelines for specific subjects on a national level, in addition to a variety of provincial strategies (DoE 2010). P.90	Low rigour	CL
Kapambwe (2010)	Zambia	The objectives of the CA programme are twofold: firstly, to promote the use of formative assessment so as to improve the quality of learning and teaching and secondly, to establish a regular system of managing cumulative pupils' performance marks for purposes of using them in combination with final examination marks for selection and certification. P.100	Medium rigour	FI

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
Kellaghan and Greaney (2001)	Developing countries	A further consequence of attaching high stakes to performance is that, in an effort to improve the mean level of performance of a school, teaching resources and strategies may be focused on 'borderline' students (to increase the number classified as proficient), while lower and higher achieving students may be neglected. Schools may also adopt procedures to reduce the number of students of low achievement who sit the assessment tasks. This may involve a number of strategies: being more selective in the students that the school enrols, retaining students in grades in which students are not assessed, or increasing the number of students classified as having a disability if regulations permit their exclusion from the assessment. Each of these strategies will reduce the number of students with low levels of achievement taking an assessment, and this, of course, will impact on the school's average performance. (P.80)	Medium rigour	CL
		Finally, high-stakes testing has been associated with problems in the recruitment and retention of teachers. Teachers leave schools in which student performance is poor, and are attracted to schools in which they will receive a bonus for good performance. P.81		
		Undesirable consequences have also been identified when high stakes (in the form of some kind of sanction) are attached to a school's performance in a national or state assessment. As has been documented in the case of external (public) examinations, students will be coached on the assessment instruments, and teachers will focus their efforts on what is measured, leading to a		

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
		narrowing of the curriculum and concentration on achieving high levels of test performance at the expense of general cognitive development and deeper forms of learning. P.80		
Kellaghan and Greaney (2004)	Sub-Saharan Africa and Carribean	In Kenya, the introduction of Kiswahili and practical subjects to the Kenyan Certificate of Primary Education in the 1980s is reported to have resulted in a dramatic increase in the coverage of these subjects in schools, despite difficulties relating to facilities, textbooks, and teacher competence (Eisemon 1990). Also in the 1980s, Trinidad and Tobago amended its Common Entrance Examination, taken at the end of primary schooling, replacing a multiple-choice test on sentence style and structure by an essay writing component. This had the effect of increasing the amount of writing tasks assigned by teachers, thus giving students experience in formulating arguments and applying their knowledge to problem solving.

London (1997) reported that 'essay writing has now been actively taught in the schools for almost a decade … [M]ost teachers … express a sense of relief that essay-writing … is being given its fair share of time within day-to-day classroom exercises' P.16 | Low rigour | CL |
| | | There is little empirical evidence to support or to challenge the claim that a change in examinations will result in an improvement in the level of student achievements. P.16 | | |
| | | In a study carried out in standard 8 in Nairobi primary schools, teachers were asked to prepare pupils for two sets of mock examination questions (Eisemon 1990). One set had been prepared | | |

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
		by the Kenya National Examinations Council, while the other was designed specifically to test higher-order cognitive skills, requiring students to make inferences rather than to recognise the correct answer. The latter paper resulted in significant changes in the way in which teachers prepared their pupils, and these pupils ultimately performed better on both examinations than did those students who had been prepared specifically for the former examination. P.19		
		There can be little doubt that assessment data published in league tables can affect the behavior of schools. In the 1990s, Senegal introduced a results oriented management system, in which information on school performance was published in the press. Between 1995 and 1998, the success rate for the examination at the end of primary school rose from 30 percent to 48 percent. Furthermore, the enrolment rate of girls rose from 40 percent to 77 percent (ADEA 2002). These improvements cannot be attributed solely to the publication of results, however, as Senegal simultaneously introduced other reforms, including the introduction of job descriptions, more school inspections, and seminars and open days. P.20		
		A further disadvantage of examinations to which high stakes are attached is that they tend to affect teaching strategies, learning strategies, student involvement in learning tasks, and student attitudes to learning. Teachers will tend to rely on drill, and may require their students to use strategies that are superficial or short-term, such as memorising, rehearsing, and rote learning. P.23		

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
		When high stakes are attached to performance, students tend to be less successful in acquiring higher-order and transferable skills; learning tasks are perceived as not inherently interesting; and, if a reward is removed, students will be less likely to engage in a task (Kellaghan, Madaus, and Raczek 1996). P.23		
		At lower grades also, the subjects in which the examinations are taken are likely to be given greater emphasis, at the expense of other curriculum goals. Even the format of examinations may affect teaching. For example, use of the multiple-choice format is observable not only in classroom tests but also in the teaching methodology applied in the early grades of primary school. P.24		
		Teachers, whose reputations may depend on how well their pupils perform in examinations, may focus their efforts on those pupils who are most likely to succeed. When this happens, it is likely to inhibit attainment of the Education For All goal that all pupils should complete a basic education of good quality. P.25		
		In Lesotho, where four out of five pupils passed the Primary Certificate Examination, fewer than one in six scored at the minimum level of mastery in a national assessment of literacy. In Malawi, close to four out of five pupils passed the Primary Certificate Examination, but in a national assessment, only one in five achieved minimum mastery. In Uganda, about 70 percent		

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
		passed the certificate examination, but in a national assessment, only about one-third achieved minimum mastery. The figures for the examinations and national assessments are not based on the same cohorts of students, but the discrepancies are so large that it is unlikely that they do not represent real differences in the standards applied in public examinations and national assessments. P.38		
Luxia (2005)	China	Moreover, some teachers evaluate their own work on the basis of students' performance in the NMET [National Matriculation English Test]. Teacher E said: I would have a sense of achievement if the NMET mean score of my class is higher than that of another class or than the class I taught in the previous year. P.153	Medium rigour	FI
Mukhopadhyay and Sriprakash (2011)	India	The schools, which opted for the evaluation, were found to be striving hard to prepare their children to clear the tests in order to achieve 'Learning Guarantee Award' for the school. For this purpose, children have been put to rigorous regular testing by the teachers… P.317	Low rigour	FI
		Another unintended outcome of the KSQAO in schools was more troubling. The construction of students from lower castes and classes as 'backward' by teachers is well-documented in India, particularly in rural government schools which serve the majority of marginalised populations…. The institutionalised need for remedial teaching in government schools was seen to strengthen such deficit assumptions of students. P.322		

A: High-stakes assessment: school- and student-level anticipation of consequences				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study***	**Type of evidence***
Reyneke et al. (2010)	South Africa	Some believed that there was 'Far too much focus on daily (and) weekly assessment if one implemented it, there would hardly (be) any time for teaching', that assessment for learning ' is designed in the way that it caters for classes with reasonable learners - not black schools' and that to get the learners involved in learning activities was nothing but 'playing around'. Remarks like these portray a misunderstanding of assessment for learning that needs to become part of classroom practice. P.286	Medium rigour	FI
Taylor (2009)	South Africa	Forceful measures were taken against underperforming schools in the final year examinations. This accountability resulted in a rapid rise in the examination results achieved by manipulating the results by: 'eliminating high-risk candidates, encouraging candidates to register at a lower examination level of standard grade, lowering the standard of examination questions and raising raw scores during the moderation process' (p. 341).	Medium rigour	CL

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

B: High-stakes assessment: Consequences of dissemination of results				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Ferrer (2006)	Latin America	In recent years, countries that use high-stakes tests have devised more sophisticated reporting formats that are more useful from a pedagogical and curriculum perspective for teachers and students. As noted earlier, the most progress in this regard has been made by systems geared to accrediting students when they leave high school or for admission to higher education. P.73	Medium rigour	CL
Kellaghan and Greaney (2001)	Developing countries	In Uganda, dissemination focused on assisting district educational authorities in developing strategies to apply, in the classroom, information derived from national assessments carried out by the Uganda National Examinations Board. The process comprised a series of workshops and seminars for teacher trainers, school inspectors, and teachers in which the results of an assessment were presented. Implications for teaching were considered, as well as how to use information to adapt the instructional process to improve learning. Teachers were expected to shift their emphasis in the way they presented curriculum material, and to pay more attention to areas of knowledge and skill that the national assessment identified as being relatively weak. It was decided to involve all teachers in schools, not just teachers at the grade level targeted in the assessment, first, because the roots of problems that might be identified were likely to be found in classes other than the one in which students were assessed, and, secondly, it was considered desirable that problems should be addressed in a consistent way throughout a school. The need for continuing support to teachers as they attempted to implement change was recognised. PP.57-58	Medium rigour	CL

B: High-stakes assessment: Consequences of dissemination of results				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		In Chile, 900 schools (about 10 per cent of schools in the country) are identified as being in need of assistance on the basis of their performance in language and mathematics assessments. Efforts are then made to improve performance in a variety of ways: by providing textbooks, classroom libraries, and pedagogical materials, and by arranging for teacher professional development through school-based workshops. Provision is made for after-school activities for students. There is evidence that the achievement gap between the schools and other schools diminished over time... P.72		
Kellaghan and Greaney (2004)	Sub-Saharan Africa	… there is, however, some evidence that in-service provision could be effective in changing the approaches adopted by teachers. Guidance provided to teachers in the preparation of students for examination, coupled with the development of teacher understanding of the demands of examinations, can lead to greater emphasis on the classroom teaching of higher-level skills. P.18	Low rigour	CL
Santiago et al. (2012)	Mexico	Information systems and sample-based national assessments … have been continuously refined over the last decade. The key challenge, however, is to ensure that stakeholders across the system make effective use of the available data. System-level data are not well exploited to inform the development of policies. There seems to be limited capacity and/or interest at the state and national levels to engage in deeper analysis and interpretation of results. Another challenge is to facilitate the use of data by professionals at the school level. This calls for the development of strategies to optimise the use of existing system-level data by stakeholders	Medium rigour	CL

B: High-stakes assessment: Consequences of dissemination of results				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		across the system. Also, there are some areas where the collection of data should be further developed: individual student and teacher trajectories in the school system; the monitoring of inequities in learning outcomes between specific student groups; the socio-economic and demographic backgrounds of students; and the perceptions of stakeholders regarding the teaching and learning environment. Finally, EXCALE (Educational Quality and Achievement Tests, sample-based standardised student assessment for national monitoring) should be continuously reviewed to ensure their relevance to national education goals (p.12).		

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

C: High-stakes assessment: Motivation through incentives				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Barrera-Osario and Raju (2010)	Pakistan	In return for receiving the subsidy benefit, the program school has to, among other things, waive tuition and fees for all students and ensure that the school achieves a minimum student pass rate in the Quality Assurance Test (QAT). Program schools that satisfy the above conditions are also	High rigour	FI

C: High-stakes assessment: Motivation through incentives				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		eligible for other substantial cash benefits offered on an annual basis: group-based bonuses for teachers in schools that achieve high QAT pass rates/mean scores and competitive bonuses for schools that rank highest in the QAT in each main program district. P.3		
		Sharp regression discontinuity (RD) estimates show that the threat of program exit on marginal first-time failures induces large learning gains. The large change in learning between the first two test rounds is likely importantly attributable to this accountability pressure given that a large share of new program entrants failed in the first test round. Schools also qualify for substantial annual teacher bonuses if they de facto achieve a minimum score in a composite measure of student test participation and mean test score. Sharp RD estimates however do not show that the prospect of future teacher bonus rewards induces learning gains for marginal bonus non-qualifiers. Thus, the evidence collectively suggests that, apart from the pressure from below to maintain a minimum level of learning for program participation, program schools do not face any effective incentives to continuously raise learning. P.38		
Glewwe et al. (2010)	Kenya	'Teacher attendance was not affected by the incentive program.' 'Prior to the program, schools that would later be selected to be program schools have slightly higher teacher attendance, although the difference was insignificant' P.20	High rigour	FI

C: High-stakes assessment: Motivation through incentives				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		'Teacher behavior was not significantly different between the incentive and comparison schools.' 'results for two objective measures (blackboard use and teaching aid use) and two subjective ones (teacher caring and energy) are presented here.' 'There was no significant difference in pedagogy between the incentive and comparison schools for any of the classroom observations prior to the program … We also find no significant difference during the intervention period between the two school groups in any of the pedagogical practices … The point estimates are close to zero for each observation type. P.21		
		'Incentive schools conducted more preps than comparison schools.' 'Prior to the program, incentive schools were slightly less likely to offer preps … but after the introduction of the program, treatment schools started to conduct more preps'. P.22 The program had little impact on dropout and repetition rates, but increased student participation in exams. P.22		
		During the period the program was in place, student scores increased, significantly so on some test measures. There is some suggestive evidence that the effect was larger in the subjects more vulnerable to coaching. After the end of the program the effect on test scores did not persist. Students who had been in program schools during the program scored no higher than their counterparts who had been in comparison schools. P.23		

		C: High-stakes assessment: Motivation through incentives		
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Kremer et al. (2004)	Kenya	'The scholarship program does not appear to have led students to focus on test performance at the expense of other dimensions of learning. This stands in sharp contrast to another project conducted by the same non-governmental organisation which provided incentives for teachers based on students' test scores. That teacher incentive program had no measurable effect on either student or teacher attendance, but increased the frequency of test preparation sessions known as 'preps' (Glewwe et al. 2003).' 'In contrast, in the merit scholarship program we study, both student school participation and teacher school attendance increased in program schools, test score gains remain large in the year following the competition, and there is no increase in the frequency of test preparation sessions. PP.2-3	High rigour	FI
		The June 2003 structured interviews with teachers provide some evidence on how parental support may have contributed to program success in Busia. For instance, one teacher mentioned that after the program was introduced, parents began to 'ask teachers to work hard so that [their daughters] can win more scholarships.' A teacher in a different Busia school asserted that parents visited the school more frequently to check up on teachers, and to 'encourage the pupils to put in more efforts'. P.13		
Mukhopadhyay and Sriprakash (2011)	India	One of the primary motivators behind the development of the KSQAO was the Azim Premji Foundation, a large-scale corporate foundation working in close collaboration with the state government. The Learning Guarantee Programme, piloted in 2002–3, was one of the earliest joint initiatives of the Azim Premji Foundation and the education department in Karnataka. This	Low rigour	FI

C: High-stakes assessment: Motivation through incentives				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		programme set out pre-specified criteria for 'learning achievements'; high-performing schools were rewarded with cash incentives and there were also awards for students and teachers at an individual level. The participation of schools in the programme was seen to be voluntary and was decided by each headteacher. The design of the programme positioned education and 'learning achievement' in explicit market-oriented terms. P.316		
Muralidharan and Sundararaman (2011)	India	School-level group incentives and teacher-level individual incentives perform equally well in the first year, but the individual incentive schools outperformed the group incentive schools after 2 years of the program. At the end of 2 years, the average treatment effect was 0.28 SD in the individual incentive schools compared to 0.15 SD in the group incentive schools, with this difference being significant at the 10 per cent level. P.41 Our results suggest that the main mechanism for the impact of the incentive program was not increased teacher attendance but greater (and more effective) teaching effort conditional on being present. P.41	High rigour	FI
		The interviews indicate that teachers in incentive schools are significantly more likely to have assigned more homework and class work, conducted extra classes beyond regular school hours, given practice tests, and paid special attention to weaker children. P.68		

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

D: High-stakes assessment: Process improvement through school-based performance assessment				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Bansilal (2011)	South Africa	Vanitha's experience of support from the education department is not encouraging. She said that in the current year, there was one course that was offered to all the grade 9 teachers in her area by the department, but the notification of that workshop was faxed to her school at 4 pm on the day of the workshop. The only other course run by the department that she attended was in the previous year and that workshop was facilitated by a grade 7 teacher who (in Vanitha's opinion) did not display sufficient insight into mathematics and could not address the questions that they posed to her. P.97 The lesson excerpt above reveals some of Vanitha's instructional choices about what to promote in her classroom, based on her analysis of the algebraic demands of the task, as well as her expectations of her learner's readiness for the task. P.104	Medium rigour	FI
		An additional constraint to the setup of the ZFM emerged from the restrictions of the assessment protocol. Although grand and sweeping statements are articulated in the policy…, the teacher actually had very little choice. The context and framing of the task was pre-determined by the national task designers, where the extensive use of language in the task was a challenge to her learners who struggled with words such as 'identical'. P.105		

D: High-stakes assessment: Process improvement through school-based performance assessment				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Beets and van Louw (2011)	South Africa	…, teachers, in an effort to ensure fairness, believe that assessments had to be uniformly administered and are consequently reluctant to conduct more intensive individualised assessments with only below-grade-level readers. These problems in South Africa are further exacerbated if it is taken into account that illiteracy rates are around 24% for learners over 15 years old and that many teachers in township schools are poorly trained (p.311).	Low rigour	CL
Kapambwe (2010)	Zambia	The results from the quantitative evaluation study on the comparison in performance between the pupils in the CA pilot schools and controls schools showed that the CA pupils' performance on the post test were higher compared to their results on the baseline tests. The difference between the baseline mean scores and the post mean scores were significant and this was attributed to the CA interventions. P.103	Medium rigour	FI
		Pointed to the fact that CA would bring about improvements in the teaching and learning processes. This implies that continuous assessment provided useful feedback on the teaching and learning processes and enabled teachers to be more involved in teaching and assessing. The results also show the positive influence that appropriate assessment has on instruction. P.104		

D: High-stakes assessment: Process improvement through school-based performance assessment				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Kellaghan and Greaney (2004)	Sub-Saharan Africa	On the formal aspects of assessment in schools, rather than on the informal aspects. As a result, the schoolroom assessment may end up based on written tests or item banks administered to pupils in a formal test situation that essentially mimics the external examination. P.53	Low rigour	CL
		It is hardly surprising in light of these observations that the implementation of school-based assessment as a component of public examinations, in countries including Lesotho, Namibia, Nigeria, Swaziland, Tanzania, and Uganda, has proved problematic. While the aspiration and motivation to introduce it have been high, the practical difficulties have on more than one occasion resulted in the failure, postponement, or limitation to a token amount of the school-based element. P.52		
		Many students currently perform poorly on examinations, but this clearly is not due solely to the quality of the examinations. Much more significant is the prevailing lack of teacher competence and lack of resource material, the large size of classes, and the difficulty of teaching higher-order skills. It is unrealistic to expect that new examinations can override the influence of these factors. P.60		
Lubisi and Murphy (2002)	South Africa	A lot of 'advice' has been given to teachers on how best to implement CA. While the KwaZulu-Natal Department of Education and Culture year mark implementation Guideline Document 1 stipulates that the year mark should not be seen to be predictive of the pupil's performance in the	Low rigour	CL

D: High-stakes assessment: Process improvement through school-based performance assessment				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		year end exam, teachers are urged to 'ultimately base their assessments on their own perceptions of the standards that are set by the KwaZulu-Natal Senior Certificate examination itself'… While this statement is understandable as it refers to 'standards', it runs the danger of being interpreted as an invitation to parrot assessment tasks/activities/exercises similar to (if not the same as) those set in previous examinations, which has always been the case with many teachers. P.265

Another element related to capacity is that of suitable guidance to teachers. Apart from listing 'classroom tests, projects, homework [and] co-operative learning activities' …, policy does not give guidance as to how 'classroom-based assessment' should be conducted and structured. This silence is likely to lead to teachers using the same tasks and exercises found in old test/exam papers (and in traditional textbooks) and re-packaging them as 'homework', 'classwork' or 'assignment'… P.265 | | |
| Nsibande and Modiba (2012) | Swaziland | Even though teachers were made aware that the curriculum documents provided were to be used as a guide, they were unable to reflect and display critical understanding of the lesson objectives and how they could be translated effectively into assessment items in the context they had to teach. P.19 | Medium rigour | FI |
| Ong (2010) | Malaysia | In Malaysia, pressure on teachers to produce high test performance results in much teaching to the test and the adoption of teaching methods designed to prepare students for the test so as to achieve high test performance. | Medium rigour | CL |

D: High-stakes assessment: Process improvement through school-based performance assessment				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		… an assessment system that places greater responsibility in the hands of teachers would necessarily demand a good deal of training and support. The success of the assessment system hinges on the professional development of, and the support provided to, the teachers (p.101).		
Pryor and Lubisi (2002)	South Africa	Where INSET has been available it has tended to focus on procedural or bureaucratic functions, such as how to fill in and calculate official mark sheets, rather than helping teachers to understand the rationale behind CA and its formative potential. CA has therefore been seen as a technical solution to the educational problem of the 'one shot' examination, which has served to alienate teachers and distract them from more interactive pedagogy... P. 674	Low rigour	FI
Reyneke et al. (2010)	South Africa	Some believed that there was 'Far too much focus on daily (and) weekly assessment if one implemented it, there would hardly (be) any time for teaching', that assessment for learning 'is designed in the way that it caters for classes with reasonable learners - not black schools' and that to get the learners involved in learning activities was nothing but 'playing around'. Remarks like these portray a misunderstanding of assessment for learning that needs to become part of classroom practice. P.286	Medium rigour	FI
		Seeing CASS as a technical procedure (something that must be done to satisfy the bureaucrats) rather than a matter of professional judgement (something that should be done to help learners),		

D: High-stakes assessment: Process improvement through school-based performance assessment				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		reduces parts of the curriculum to a set of administrative requirements that must be followed without understanding the principles on which those procedures are based... P.287 Because of extensive record keeping and monitoring of individual learners, CASS in the South African system leads to an increase in teacher workload. P.287		
		The 2008 SBA moderation report by Umalusi (the Council for Quality Assurance in General and Further Education and Training in South Africa) highlighted 'the poor quality and standard of the tasks set by educators; the low validity of internally set assessment tasks; the unreliability of marking instruments and the discrepancies in allocation of marks; and the unbalanced weighting of the cognitive demand and difficulty of the tasks' (2009 p.10). Umalusi's conclusion regarding the quality of the 2009 National Certificate assessment and examination was that much more needs to be done to improve the quality of SBA. P.278		
		One would expect that such a system of continuous SBA for account- ability would give the Grade 12 learners a realistic picture of their own competence. Umalusi however found that inaccurate continuous assessments were sending the wrong signals to learners and parents throughout the FET band, resulting in a large number of under-prepared students entering the matriculation examination... P.279		

D: High-stakes assessment: Process improvement through school-based performance assessment				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Scherman et al. (2011)	South Africa	[S]chools and teachers are required to use the protocols report according to the levels of achievement. It would also appear as if there is no statistically sound justification for why these categories and corresponding percentages exist. This is possibly why there is an uneven distribution of percentages. P. 516 the recording and reporting protocols seemed to lack statistical validity as the documentation does not provide an indication of how these levels of achievement were constructed P.521 What is clear from the analyses is that the recording and reporting protocol is not adequate, as not all of the levels could be represented in terms of difficulty of items as well as ability of persons. However, what was highlighted, and in line with national and international assessments, was that very few if any pupils are represented on the upper levels. This is a serious cause for concern, especially as this is an assessment of basic mathematical skills and abilities and does not augur well for the pupils' continuation of mathematical study in the Senior Phase P.521.	Medium rigour	FI

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

Table A4.2.3: Pathways to impact of low-stakes assessment

A: Low-stakes assessment: The establishment of curriculum standards				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Ferrer (2006)	Latin America	Autonomous assessment entities may have a higher degree of functional independence and technical legitimacy than those associated with ministries, but they can also create problems. The greatest risk is that they might become disconnected from ministry information needs and turn into programs that, while of a high technical quality, have little impact on policy decisions geared to improving educational quality. P.18	Medium rigour	FI
		In the region have professional assessment staff dedicated to establishing or specifying clear curricular targets that allow them to design more focused tests, while they continue to work (and sometimes further the debate) on what concrete outcomes of learning are expected as a priority from the students. Given the lack of concerted efforts to develop content and academic performance standards, the assessment agencies' endeavours are a significant step in the right direction.		

A: Low-stakes assessment: The establishment of curriculum standards				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		Colombia, Ecuador, and Uruguay illustrate this trend, as do the subnational assessment systems in Aguascalientes, Mexico, and Bogotá, Columbia. P.21		
		assessment questions call for demonstration of knowledge and cognitive aptitudes that are relevant and desirable, but that are not explicit in the national curriculum. Validation of the test items, especially by teachers, has spurred substantial thinking about the established curriculum and the way it is implemented in the schools. P.22		
		A set of skills consistent with national curricular guidelines was also established for the district-level assessments undertaken by Bogotá's education secretariat. These tests, like the national-level State Examination, provide an explicit conceptual framework and operational definitions for each of the assessed skills. As a result, different levels of student performance can be reported and illustrated more clearly than would be possible using the national curricular guidelines. Both technically and politically, the basic skills that have been proposed can be regarded as curricular standards, since they offer a means of effective communication among all the actors in the sector as to what students in the system are expected to learn. P.22		
		Argentina's educational assessment system include a lack of communication between the assessment unit and the offices responsible for curricular development and teacher training, and resistance to quantitative assessment on the part of some academics and technical specialists within the ministry. P.58		

A: Low-stakes assessment: The establishment of curriculum standards				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Gvirtz (2002)	Argentina	'Common Basic Contents' (CBCs) served as the basis for the development of a national assessment. Textbook publishers developed textbooks that elaborated the CBCs and focused on areas targeted in the national assessment. 'CBCs were adopted as the new school curriculum' even though they were intended to serve only as guidelines. P. 465	???	CL
Kellaghan and Greaney (2001)	Developing countries	In Brazil, the findings of national assessments have also been associated with curricular changes and pedagogical innovations, and have influenced human resource training and policies regarding the financing of schools serving disadvantaged areas. P.59	Medium rigour	CL
		In Thailand, measures of affective outcomes, practical skills, and social perception were included in a national assessment in an effort to dislodge teachers' preoccupation with cognitive development in specific content areas (a preoccupation that was reinforced by end-of-school examinations for university entrance). This led teachers to place greater emphasis on these outcomes in their teaching and assessments (Pravalpruk, 1996). P.79		
Kellaghan and Greaney (2004)	Sub-Saharan Africa	The assessment was designed to help policymakers allocate resources to underachieving schools; its results suggested that the expectation of competence in English was too high, and that curriculum materials might need to be revised. P.35	Low rigour	CL

A: Low-stakes assessment: The establishment of curriculum standards				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		The assessment identified aspects of the curriculum that were causing particular problems (such as place value, word problems in mathematics); found that boys generally outperformed girls; and identified implications for teacher education and teaching methodologies. P.35		
		There is less evidence that the information from national assessments has affected school practice, but this is not surprising. National assessments have not been in existence for long, and one would expect that it would take some time before they have an impact. Furthermore, using results to affect school practice is not easy; the effort to achieve this has been made in several countries, but its impact has not been assessed. P.63		

* FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

B: Low-stakes assessment: Guidance and support for schools and teachers				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Ferrer (2006)	Latin America	In the Dominican Republic, the institutional reports given to the schools disaggregate results by course or section and by knowledge area. They note the percentages of students who move forward a grade and who are kept behind, and the correlation between final school score (internal) and the score in the standardised test; they also provide a comparison of results with similar schools, as well as with all schools in the country, region, and district. P.73 In Uruguay, data disaggregated by school are given confidentially to each establishment; they present student results and those of students in schools in similar socioeconomic circumstances. P.73 Bolivia's System for Measuring and Evaluating the Quality of Education (SIMECAL), for both its census-based and sample-based tests, gives the results to schools in the form of an institutional report on their students' performance, one that includes the average institutional score, an operational description of achievement levels by area, and the percentage of students in the school at each level. This information is followed by a description of the strengths and weaknesses of the entire student population by core topics in each area of the curriculum. P.73 The improvement projects the schools are asked to implement call for the use of the available statistics and analysis of the in-school and out-of-school factors that have been shown to affect performance. Emphasis is placed on the in-school factors, since it is here that schools can have a direct effect by devising new teaching and curricular strategies. P.73	Medium rigour	FI

B: Low-stakes assessment: Guidance and support for schools and teachers				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Ravela (2002)	Latin America	Since its national assessments began, Argentina has probably been the country that has put the greatest stress on analysing the skills most and least acquired in the different areas assessed. It does this through 'methodological notebooks' that have been published systematically since the first national assessment in 1993. P.34 These notebooks, which to some extent have served as models for other countries in the region, seek to foster didactic and disciplinary reflection among teachers on the basis of concrete examples of items and results. Emphasis is placed on those activities with a higher level of achievement and those that proved most difficult. P.34	Medium rigour	CL

FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

C: Low-stakes assessment: Exploring absence of evidence of systemic impact				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Ferrer (2006)	Latin America	Assessment data are alarmingly underused in designing strategies to improve educational quality; sometimes, too, they prompt unwarranted conclusions or invalid generalisations. P.27	Medium rigour	FI
		For both national and international tests, it is apparent that information exchange between assessment entities and system users remains weak, despite a variety of available reporting mechanisms. Better communication strategies are needed to enable information users - including policymakers, teachers and principals, parents, the media, professional associations, and representatives of the general public - to take part in devising tests, designing reports, and defining strategies for dissemination and use. P.48		
		Assessment results do not seem to have found a place in either public debate or the agenda of the education system. Only international test results seem to have some impact on the media and public opinion. P.57		
		Bolivia: Assessment results have had only a limited impact on the policy agenda due to communication problems in the ministry and, probably, the lack of agreement within the education sector regarding the validity and importance of the achievements being tested. The main difficulties of the Bolivian assessment system include a lack of sectoral policies that stress monitoring of educational quality and that make systematic use of SIMECAL information to that end. There is a firmly held belief that SIMECAL data could be used as a basis in intersectoral		

C: Low-stakes assessment: Exploring absence of evidence of systemic impact				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		policymaking, but poor communication and inadequate links among government agencies prevent this from happening. Relatedly, there is inadequate monitoring of how schools use SIMECAL data. P.62 Dominican Republic: The local press publishes the assessment results and comments on them every year, but it is difficult to ensure that the media engage in substantive discussion of the matter rather than simply presenting overall national averages. Organisational problems in the Education Secretariat hamper the proper flow of information and synergies among the various management units. There is usually an overlap of assessment efforts among the various agencies, and they have only a limited impact on decisions geared to devising comprehensive or complementary policies on quality improvement. P.62		
		In general, the results have had little impact on public opinion and education policymaking. P.101		
		The main difficulties in the Honduran assessment system are a lack of communication and weak agreements between UMCE and SEP regarding the targets and uses of the assessments; consequently, the data are underused for policymaking purposes. P.105		
		As of this writing, the main difficulty has been that the many assessments carried out have not given rise to a substantive and comprehensive appraisal of educational quality. Because of the lack		

C: Low-stakes assessment: Exploring absence of evidence of systemic impact				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		of communication among SEP units, and the still deficient mechanisms to disseminate results, data remain unanalyzed and fail to have a greater impact on policymaking. P.110		
Howie (2012)	South Africa	It has been difficult to discern decision making taking place that is based upon the international studies and national assessments. Whilst there are decisions made and events that follow the international studies and their outcomes, it is not always easy to categorically link these to the studies themselves. P.91	Low rigour	CL
Kellaghan and Greaney (2001)	Developing countries	In other studies, the issue of how information might be used received little attention. There are cases in which data have not been made available to educational planners and managers, while in at least one country collected data have not been analyzed. P.55	Medium rigour	CL
		While such efforts are to be commended, the difficulty of conveying what is meant by centrally specified standards and information derived from assessments to those who have to act on the information should not be underestimated. In an evaluation of the national assessment in Brazil, it was found that reports were not written in a user-friendly language, and that only a few people at state level had the critical capacity to interpret results. It would be surprising if this was not also the case in other countries. P.58		

C: Low-stakes assessment: Exploring absence of evidence of systemic impact				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Kellaghan and Greaney (2004)	Sub-Saharan Africa	The present study posed a series of questions on the use of national assessment results to senior education personnel in six countries (Ethiopia, Malawi, Niger, Nigeria, South Africa, and Uganda). The respondents reported that while the findings of national assessments sometimes were covered in the media, in none of the six countries did they feature in parliamentary debate. P.36	Low rigour	CL

* FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

Table A5.1.1: Monitoring outcomes

Providing feedback: Uses of EMIS for management decisions that affect the school level					
Studies*	**Countries**	**Intervention**	**Intermediate outcomes**	**Quality (Rig/Rel)***	**Type of evidence***
Lassibille et al. (2010)	Madagascar	Provision of summary reports from EMIS that corresponded with management level (control + treatment), along with management toolkits and guides as well as training (treatment only)	Significant impacts on manager, teacher and student behaviours for treatment that included guides and training No significant impact on student learning	High/High	FI
Murimba (2005)	Southern and Eastern Africa (15 countries)	Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) monitoring, evaluation, and reports for ministries' monitoring process: reporting impacts observed as part of reporting on progress, monitoring visits, and experiences/anecdotes	Enabled countries to assess quality of education systems Allowed ministries to assess performance against own standards and other countries No school-level service delivery outcomes reported	Low/Med	CL
World Bank (2010)	Madagascar	AGEPA (Amélioration de la Gestion dans les Pays Africains)	High-quality, country-tailored technical support Capacity development to define and conduct analytical work to diagnose management and accountability gaps to impact local, school, and classroom levels to address gaps	Med/High	CL

Providing feedback: School report cards					
Studies	**Countries**	**Intervention**	**Intermediate outcomes**	**Quality (Rig/Rel)***	**Type of evidence***
Andrabi et al. (2013)	Pakistan	School report cards	Parental beliefs about school quality corresponded with student achievement scores	High/Med	FI
Barr et al. (2012)	Uganda	Participatory scorecards and standard scorecards assigned to treatment schools and teachers	Reduced student and teacher absenteeism in participatory scorecard schools Improved sense of ownership among school stakeholders	High/High	FI
Bruns et al. (2011)	Brazil Nigeria	Centrally-provided school report cards for local school, parental and community decision-making	Increased parental voice 'The report cards also acted as a management tool at the school level and as a driver of wider education reforms…' (p. 45) 'largely qualitative and anecdotal evidence suggests … positive impacts' (p. 49)	Med/High	CL

Providing feedback: School report cards					
Studies	**Countries**	**Intervention**	**Intermediate outcomes**	**Quality (Rig/Rel)***	**Type of evidence***
USAID (2006)	International	School report cards: national, sub-national, participatory	Analytical sophistication varies widely Anecdotal evidence of improved reporting of school data (Uganda); parental mobilisation (Namibia); increased teacher and parental focus on improvement of learning outcomes (Brazil); rationing scarce resources (Nigeria)	Med/High	CL
Winkler (2005)	Brazil	Low-stakes school report card	Empowering school and parents' councils	Low/Med	CL

Setting expectations: Uses of EMIS with local school development planning					
Studies	**Countries**	**Intervention**	**Intermediate outcomes**	**Quality (Rig/Rel)***	**Type of evidence***
Attfield and Vu (2013)	Vietnam	School self-audit	Gained ability to demonstrate change Improved school planning Track equity of investments	Med/High	CL
Brock (2009)	China	School development planning	Improved school operational planning	Med/Med	CL
Caddell (2005)	Nepal	Use of data in school development planning	Communication focused on data 'extraction' to meet reporting requirements Limited opportunities at local and school levels to influence priorities	Med/Med	CL
Crouch and Winkler (2008)	Nigeria	Broaden input on EMIS data collection and reporting; creation of multidimensional reports	Improved information on comparative allocation of resources to schools, teacher recruitment Transparency and equity in system management Community better informed of school quality	Low/High	CL

Setting expectations: Uses of EMIS with local school development planning					
Studies	**Countries**	**Intervention**	**Intermediate outcomes**	**Quality (Rig/Rel)***	**Type of evidence***
Powell (2006)	Nigeria Mozambique	Sub-national operational plans	Outputs from EMIS used to support development of operational plans and budgets at the district level (P. 18) – Kano, Nigeria EMIS outputs have not played significant roles in planning due to lack of funds and capacity – operational plans developed separately from strategic plans, different sets of indicators. P.19 - Mozambique	Med/Med	CL
Prew and Quaigrain (2010)	Ghana	EMIS school-level reports on teacher attendance	Put in mechanisms, including teacher incentives, to address teacher attendance	Low/Med	FI
Winkler and Herstein (2005)	Ghana Guinea Nigeria	Increasing capacity at school-level for self-audit	Supplying information fails in absence of efforts to increase information demand EMIS needs to be sensitive to underlying interests and goals of stakeholders	Low/Med	CL

		Setting expectations: Uses of EMIS with local school development planning			
Studies	**Countries**	**Intervention**	**Intermediate outcomes**	**Quality (Rig/Rel)***	**Type of evidence***
			Increase in demand possible through collective stakeholder commitment to improve and holding local school leaders accountable for resource decisions		

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

Table A4.3.2: Evidence per type of activity - Monitoring

		A: Uses of EMIS for management decisions that affect the school level			
Studies	**Countries**	**Evidence (quotations)**		**Quality of the study**	**Type of evidence***
Attfield and Vu (2013)	Vietnam	The system enabled accurate, quantitative benchmarking of a concept, with a sufficiently rich data set that has been exploited in a diverse range of ways to demonstrate change, track equity of investments, plan on both the macro and micro level and test hypotheses on the variables and factors that hide within the 'black box' of a classroom and somehow determine learning. P.83		Medium rigour	CL

A: Uses of EMIS for management decisions that affect the school level				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
Lassibille et al. (2010)	Madagascar	The results show that interventions at the school level, reinforced by interventions at the subdistrict and district levels, succeeded in changing the behavior of the actors toward better management of key pedagogical functions. In terms of education outcomes, the interventions improved school attendance, reduced grade repetition, and raised test scores (particularly in Malagasy and mathematics), although the gains in learning at the end of the evaluation period were not always statistically significant. P.322	High rigour	FI
Murimba (2005)	Southern and Eastern Africa	The mechanism for monitoring the impact of SACMEQ's research programme, as well as the training component associated with it, contains three elements. Firstly, SACMEQ ministries are periodically requested to report on any impacts observed as part of reporting on progress. Secondly, the information provided by ministries is complemented by observations made through monitoring visits made by the SACMEQ Director, national research co-ordinators (NRCs) and members of technical teams. The third element comprises the experiences (some of them anecdotal) shared at the various forums that bring the different SACMEQ players together. The impacts reported by different SACMEQ countries are summarised under the headings: (a) monitoring and evaluating quality; (b) capacity-building; (c) enhancing the quality of statistical and non-statistical information systems; (d) policy-making and systems-improvement processes; and (e) choosing pathways to the achievement of Education for All (EFA).' 'As implied by its name, SACMEQ's key role is to monitor and evaluate the quality of education. The linkages within the different data sets collected by SACMEQ allow for comparisons against country-specific norms or expectations, across countries and over time. Furthermore, the use of classical item analysis and modern item response theory facilitates a	Low rigour	CL

293

A: Uses of EMIS for management decisions that affect the school level				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	**Type of evidence***
		descriptive account of learners' performances. There has been a general dearth of data on the quality of education in member countries because there was no systematic, rigorous method of measuring and evaluating it. The data SACMEQ generated has enabled countries to have a fairly good idea of the quality of education that their systems offer (from an input, process and outcome perspective). Ministries can therefore make assessments of their systems' performance (in terms of learning outcomes) against other countries, and against standards that they have independently set for themselves. P.92		
World Bank (2010)	Selected African countries	Country demand for technical assistance in the area of education and school management has been continuously growing. P.3	Low rigour	CL
		Since its launch, AGEPA has been providing high-quality, country-tailored technical support to countries to help them define and conduct analytical work to diagnose management and accountability gaps, develop practical interventions at the local, school and classroom levels to address these gaps, and to implement and field-test the solutions developed. P.7		

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

B: Uses of EMIS with local school development planning				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
Brock (2009)	China	Average Net Enrolment Rate (NER) increased from 79% (1999) to 91% (2005). Girls net enrolment increased by as 26% in one county (the lowest was 17%). The biggest increases were at teaching points (primary schools in remote areas only going up to third or fourth grade) and among minority girls. P.456	Medium rigour	CL
		The effects of SDP have been very positive. It has introduced a level of real operational planning to schools and has shown County Education Bureau (CEB) staff how delegation of responsibility to schools and headteachers, while running the risk of abuse, does in most cases result in increased ownership of local issues – thereby reducing the demands on the... ...SDP is now being promoted in at least 10 other provinces in China. P.457		
Caddell (2005)	Nepal	In practice, however, the [District Education Plan] process remained largely focused on the extraction of data to meet national planning, monitoring and evaluation objectives as opposed to districts or schools being able to set their own agendas. Rather than offering opportunities for local voices to be heard, these attempts sought to transfer central-level concerns and responsibilities, including the pressure to meet EFA-related objectives, to the district and sub- district level. P.462	Medium rigour	CL

		B: Uses of EMIS with local school development planning		
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
Crouch and Winkler (2008)	Nigeria	Decentralised Information in Kano State, Nigeria, is undergoing a transformation of its education management information system – from one that serves very few people and very few purposes to one designed to meet the needs of all stakeholders and go beyond merely counting students and teachers. A new collection tool and a data management system were created after soliciting input from stakeholders state-wide, including central planners and managers; local education officers; PTA and teachers union members; and members of the legislature, the governor's office, testing authorities, and the Ministry of Finance. In addition to capturing information relevant to all stakeholders, the new information system also generates multidimensional reports targeting different issues and different users, as designed by the end-users themselves. For example, state personnel managers can easily review information about the number of teachers eligible for promotion and the subject areas with the greatest need for teachers, both of which inform recruiting. P.27	Low rigour	CL
Powell (2006)	Developing countries	In Ghana the EMIS is also beginning to play an important role in supporting the process of decentralisation. The outputs from the EMIS are being used to support the development of operational plans and budgets at the district level. P.18	Medium rigour	CL
		However, at the decentralised level EMIS outputs have not played significant roles in planning due to the lack of funds and capacity constraints. Moreover, at the decentralised level operational plans are		

B: Uses of EMIS with local school development planning				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
		developed separately from strategic plans, and as a consequence both have different sets of indicators. P.19		
Prew and Quaigrain (2010)	Ghana	the most senior district official present identified teacher attendance as the core problem. As a result he took it on himself to lead a campaign to put in place mechanisms to ensure that teachers did attend school. This included spending municipal money on teacher housing for schools where they are most needed. P. 739	High rigour	FI

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

C: School report cards				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
Andrabi et al. (2013)	Pakistan	We also confirm that the intervention changed parents' beliefs about school quality: after the report cards, the gradient between parental perceptions and school test scores steepens in treatment villages, consistent with information increasing the precision of the quality signal for parents. P.4	High rigour	FI
Barr et al. (2012)	Uganda	Results show statistically and economically significant effects of the participatory design scorecard, across a range of outcomes. The participatory design scorecard reduced pupil and teacher absenteeism by and 8.9 and 13.2 percent, respectively. The participatory scorecard had a commensurate impact on pupil test scores of approximately 0.19 standard deviations; such an impact would increase a pupil from the 50th percentile to the 58th percentile of the distribution. P.ii-iii	High rigour	FI
		These results suggest that the participatory design component of community-monitoring interventions may be important to their success. Delegation of this process appears to have fostered a stronger sense of ownership among school stakeholders. P.iii		

C: School report cards				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
Bruns, et al. (2011)	Brazil	While no rigorous evaluations of this experiment exist, anecdotal evidence suggests positive effects. Parents engaged in discussions with teachers about how they might improve school performance and, through school councils, increased their voice in policy debates about education... The report cards also acted as a management tool at the school level and as a driver of wider education reforms... One of the key aspects highlighted in reviews of the intervention is that the low-stakes nature of the report cards helped make them politically feasible despite a strong teachers' union... Nevertheless, a change of state government at the end of 2002 – when a new state secretary for education was inaugurated – led to abandoning the innovation. P.45	High rigour	CL
		Most of what we know about the impacts of information-for-accountability reforms in developing countries comes from small pilots and case studies … This largely qualitative and anecdotal evidence suggests that information-for-accountability reforms might have positive impacts: greater collaborations and better communications between parents and teachers, improved parental participation in school matters, better and more frequent data reporting mechanisms, better resource flows, and some suggestion of improved education outcomes. P.49		
Winkler (2005)	Brazil	by giving school-level data high visibility, school and parents' councils became a small army of quality controllers, reporting discrepancies in state and national databases P.3	Low rigour	CL

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

APPENDIX 6.1: SYNTHESIS TABLES: INSPECTION

We have ordered the key outcomes in Table A6.1.1 in three stages to understand the chain of events that would lead to (a lack of) impact of school inspections. Immediate outcomes include a direct impact of school inspections on actions at the school or classroom level; short-term outcomes arise in relatively short periods, while intermediate outcomes occur over time.

Table A6.1.1: Hierarchy of outcomes

Study	Country	Immediate	Short-term	Intermediate
Alcazar et al. (2006, p.130)	Peru			In short, there is little evidence for the inspection story
Brock (2009, p.457)	Four counties in Gansu, China: Dongxiang, Jishishan, Hezheng and Kangle		The process of SDP was also enhanced and given prominence as an important process that set out school goals which could be measured by inspectors.	
Crouch and Winkler (2008, p.15)	Uganda			The lack of an effective inspection system at the district level.

Study	Country	Immediate	Short-term	Intermediate
De Grauwe (2007, p.711)	Africa			Supervision's feeble impact on quality.
De Grauwe (2008, p.4)	Various (Asia, Africa)	Supervision reports which are shelved without any action being taken.		
Herselman and Hay (2002, p.244)	South Africa (Eastern Cape)		It also became apparent that facilitators emphasised mechanisms and procedures in their classrooms to enhance the quality of teaching and learning, while managers used school policy, subject policies, book controls, class visitations and subject-standard staff meetings as mechanisms and procedures for quality assurance.	
Jaffer (2010, p.376)	Pakistan			Neither the system of evaluation, nor the quality of education provision has improved, despite much effort. The existing monitoring and supervision system is deficient and inconsistent, rendering it ineffective and of little help

Study	Country	Immediate	Short-term	Intermediate
				in improving the quality of the education system.
MacPherson (2011, p.201)	Timor Leste			This suggests that the scale of corruption in the misuse of school grants is probably limited and may well be contained by school inspectors policing transparency in the collection and disbursement of school grants, while not engaging in the processes themselves.
Mazibuko (2007, p.i)	South Africa (KwaZulu-Natal)			The study revealed that whole-school evaluation is not being implemented. The study shows that findings and recommendations of the supervisors have not been addressed.
Mazibuko (2007, p.275)	South Africa (KwaZulu-Natal)	Their first response was to conduct meetings to discuss the reports and to decide on the way forward. They maintained that they formed teams in their schools to deal with the logistics for drawing up the school improvement plans. These teams read the reports and recommendations made by the	All principals and educators interviewed in this study maintained that they learnt something from whole-school evaluation.	But all principals and educators who participated in this study maintained that although their schools have drawn up improvement/ development plans, nothing has happened in terms of addressing areas identified by the

Study	Country	Immediate	Short-term	Intermediate
		supervisors. According to them, these plans detail how they were going to address the recommendations of the supervisors, as well the areas that they identified during school self-evaluations. All principals interviewed maintained that after the whole-school evaluation was conducted they gave reports of the findings and recommendations of the supervisors to the parents and guardians of the learners. According to the principals, this was done to ensure that all stakeholders have an input in the school for the benefit of the learners.		supervisors as areas that need improvement.
Mazibuko (2007, p.227)	South Africa (KwaZulu-Natal)			The study shows that findings and recommendations of the supervisors have not been addressed.
Mazibuko (2007, p.227)	South Africa (KwaZulu-Natal)			One supervisor said that educators often give themselves high scores in self-evaluations. Both supervisors alleged that educators may give themselves high scores because remuneration is involved.

Study	Country	Immediate	Short-term	Intermediate
Mazibuko (2007, p.206)	South Africa (KwaZulu-Natal)	In schools where educators have a negative attitude towards the whole-school evaluation and the supervisors, principals use whole-school evaluation to threaten educators. These principals create an impression that whole-school evaluation is there to punish educators who do not do their work effectively. An educator in school A confirmed that most principals use whole-school evaluation to threaten them if things are not going well in the school.	As a result, by the time the external evaluators come to the school to conduct the whole-school evaluation educators have already developed a negative attitude towards the external evaluators.	
Opoku-Asare (2006) (p. 112)	Ghana	School inspections are often pre-announced and lenient. 'The tip-off, they said, enables the teachers concerned to prepare adequately for the observation lessons. This enables those teachers to arm themselves with all the teaching materials they can possibly lay hands on and sometimes, rehearse the lessons they intend to teach for the exercise'.		

Study	Country	Immediate	Short-term	Intermediate
Uwazi (2009)	Tanzania			These inspections are considered to be ineffective as national performance of students isn't improving.

Table A6.1.2: C-M-O configurations: School inspections - providing feedback information

Study	Condition	Mechanism	Outcome
Chen (2011, p.13)	[Lack of evidence]	[Lack of evidence]	[Lack of evidence]
Darvas and Balwanz (2014, p.136)	[Lack of evidence]	[Lack of evidence]	[Lack of evidence]
Darvas and Balwanz (2014, p.136)	Lack of authority	[Lack of evidence]	[Lack of evidence]
De Grauwe (2001, pp.16, 72, 79, 129, 130)	Lack of an adequate system of performance management of school inspectors; failure to recruit inspectors with adequate expertise; low pay grade; lack of training; assessing inspectors on quantity of visits	Principals refusing advice	Lack of impact

Study	Condition	Mechanism	Outcome
De Grauwe (2007, pp.710, 711)	Lack of resources to visit schools and limited staff, conflicting roles which leads to conflict with teachers and lack of satisfaction of teachers	[Lack of evidence]	Lack of impact
De Grauwe (2008, pp.3, 5)	Inspectors are overloaded with tasks and have to cover many schools	[Lack of evidence]	Lack of impact
Harber (2006, p.621)	Lack of training of supervisors	[Lack of evidence]	Lack of impact
Jaffer (2010, pp.376, 380)	Lack of systematic approach to collecting, analysing and reporting data	[Lack of evidence]	Lack of impact
Mazibuko (2007, p.307)	Poor communication between district office and schools to address areas identified by schools in improvement plans; unclear role of circuit office in school evaluation	Regular communication about school improvement	Inspection recommendations not addressed by schools
Moswela (2010, p.71)	[Lack of evidence]	No impression is made on teachers	No improvement of teaching standards
Opoku-Asare (2006, p. 113)	[Lack of evidence]	[Lack of evidence]	[Lack of evidence]

Study	Condition	Mechanism	Outcome
Santiago et al. (2012, p.155)	Lack of solid information about numbers, positions and roles of supervisors, leading to superficiality and subjectivity in strategic discussions about supervision. The lack of a strong core of professionally trained supervisors	Taking time away from the improvement of outcomes	[Lack of evidence]
Uwazi (2009, pp.1, 5, 6)	School inspectorate programme is driven by allegations of irregularities in use of resources, need to check construction of buildings and related utilities, and allegations of misconduct by teachers and/or students, and fails to prioritise quality of instruction and poor performance of students in the inspection cycle	Waste of useful public resources and time	Failing to safeguard quality of instruction and its improvement
Wanzare (2002, pp.6, 9, 11, 12)	Plans for inspection of schools have been over-ambitious and, consequently, they are seldom carried out Inspectors have limited time and can only inspect superficially Lack of professionalism of inspectors. Conflicting inspection standards	[Lack of evidence]	Lack of impact

Study	Condition	Mechanism	Outcome
	Reports are not distributed to teachers and parents		
Wanzare (2002, p.10)	[Lack of evidence]	Poor relationship between teachers and inspectors	[Lack of evidence]
Wanzare (2002, p.6)	Professionalism; attitudes and commitment; foci of inspection; inspectorate autonomy; inspectorate-university partnerships; inspector recruitment, selection, and deployment; adequacy of inspection; Inspectorate titles	Integrity of inspection officials is questioned	[Lack of evidence]

Table A6.1.3: CMO configurations: School inspections; consequences from inspection feedback

Study	Condition	Mechanism	Outcome
Brock (2009, p.457)	Inspections measure school goals	Giving prominence to school development planning, and setting out school goals	Enhancing school development planning

Study	Condition	Mechanism	Outcome
Chen (2011, p.13)			Weak system
Churches and McBride (2013, p.21)	[Lack of evidence]	[Lack of evidence]	[Lack of evidence]
Crouch and Winkler (2008, p.15)	[Lack of evidence]	Lack of incentives for teachers and headteachers to be present at school and perform effectively	[Lack of evidence]
De Grauwe (2001, p.143)	Lack of co-ordination between supervision service and other services which work towards pedagogical improvement, such as teacher training, teacher resource centres, curriculum development and examinations. Lack of management capacity at district and school level	Frustrated school staff and supervisors	Recommendations remain words in the wind
Jaffer (2010, p.378)	Lack of clarity of procedure for recruitment and selection, inspectors on the same pay scale as teachers, lack of training, expansion of system and increased complexity	[Lack of evidence]	Adverse impact on efficiency of inspection system

Study	Condition	Mechanism	Outcome
Opoku-Asare (2006, p.112)	Close relationships between inspectors and teachers/ head teachers	[Lack of evidence]	Adverse reports do not go beyond school gates, and are not put on teachers' files

Table A6.1.4: CMO configurations: School inspections, setting expectations

Study	Condition	Mechanism	Outcome
Brock (2009, p.457)	[Lack of evidence]	Supporting bottom-up school development, bringing schools and local communities together to create a unified approach to development	Change/more developed school
Churches and McBride (2013, p.40)	Allowing differentiation in framework design	Acceptance of review and focus of improvement on relevant priorities	[Lack of evidence]
De Grauwe (2001, p.17)	[Lack of evidence]	Teachers are considering inspectors as sources of support, rather than criticism	[Lack of evidence]
De Grauwe (2008, p.5)	Teachers' resistance to external evaluation	Resistance to inspections	[Lack of evidence]

Study	Condition	Mechanism	Outcome
	Teachers' professional autonomy and privileges.		
De Grauwe (2008, p.9)	Resistance of teacher unions to inspections Power position of stakeholders	[Lack of evidence]	[Lack of evidence]
De Grauwe (2008, p.14)	Homogeneous country with few disparities	Standardisation of quality	Positive impact
Moswela (2010, p.79)	[Lack of evidence]	[Lack of evidence]	Teachers improve their teaching
Santiago et al. (2012, p.152)	Sending schools limited options and approaches to self-evaluation (to prevent confusion)	Schools engage in more detailed self-evaluation of particular aspects of schooling such as the school environment or relations with parents	[Lack of evidence]
Wanzare (2002, p.21)	Inspection handbook needs to be realistic and practical (and not too detailed, bureaucratic, and rigid)	Schools improve their self-evaluation capabilities and use inspection handbook as a reference	[Lack of evidence]

Table A6.1.5: CMO configurations: School inspections - capacity-building of educators

Study	Condition	Mechanism	Outcome
De Grauwe (2008, p.8)	[Lack of evidence]	Strengthening linkages between schools and breaking isolation of schools, participation and commitment of teachers in sustainable change, schools learning from visits	[Lack of evidence]
De Grauwe (2008, p.15)	Weak capacity of ministry to effectively regulate system International agencies/NGOs set agenda for school improvement	Widening the gap between teachers and supervisors as they use different frameworks and criteria in school evaluation	[Lack of evidence]
Herselman and Hay (2002, pp.240, 241)	[Lack of evidence]	Installing a quality culture in schools in which teachers reflect on their practices and improve their practices (instead of focusing on compliance)	Raise standards of performance and improve learners' achievement
Moswela (2010, p.79)	[Lack of evidence]	[Lack of evidence]	Teachers improve their teaching
Wanzare (2002)	Inspection handbook is a valuable resource and not too detailed, bureaucratic, and rigid	Schools use the handbook to build their self-evaluation capacity and develop new perspective on	[Lack of evidence]

Study	Condition	Mechanism	Outcome
		educational quality and school improvement	

Table A6.1.6: CMO configurations: School inspections, system alignment

Study	Condition	Mechanism	Outcome
De Grauwe (2001, p.143)	[Lack of evidence]	School staff and supervisors are frustrated and no follow-up on inspection recommendations	Inspection recommendations are not implemented
De Grauwe (2001, p.44)	[Lack of evidence]	[Lack of evidence]	[Lack of evidence]
De Grauwe (2007, p.711)	Demanding and incoherent job description, complexity of decision making in bureaucratic education administration	[Lack of evidence]	Lack of impact
De Grauwe (2008, pp. 5, 7)	Agencies and inspectors have a sense of independence	Teachers are confused and schools do not implement inspection feedback	[Lack of evidence]

Study	Condition	Mechanism	Outcome
Herselman and Hay (2002, p.241)	[Lack of evidence]	Improving general understanding of conditions of effective schools	[Lack of evidence]
Jaffer (2010, p.386)	[Lack of evidence]	Inspection feedback is ignored	No action is taken to implement inspection feedback
Kingdon and Muzammil (2013, pp.259-260)	Strong teacher unions and strong role of teachers in elected bodies	Teachers resisting change and improvement from inspections	No quality improvements from inspections
Mazibuko (2007, p.271, 273)	Lack of clarity of roles in follow-up from inspections and in evaluation of schools Schools are not allowed to go to the District Office directly and District Office is far away Schools always look to the Circuit Offices for guidance		Whole-school evaluation is not implemented and inspection findings are not followed-up/ implemented

Table A6.1.7: Evidence according to type of activity - inspections

A: Availability of financial and material resources to visit schools (lack of resources resulting in limited visits to schools, particularly in remote areas)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Alcazar et al. (2006)	Peru	Proximity to a Ministry of Education office as a proxy for intensity of inspections and monitoring by the education bureaucracy. P.129-130	High/Med	CL
Chen (2011)	Indonesia	Schools are frequently visited by school inspectors from the district education office. Schools on average receive nearly 6 visits by the district school inspectors per year. This is somewhat verified by the district's response that quarterly and monthly school visits are common by the districts. P.13	Med/Med	FI
De Grauwe (2001)	Botswana, Namibia, Tanzania, Zimbabwe	Everywhere, the distance between the office and the schools is difficult to bridge. Transportation and travel allocation issues therefore become weighty considerations in the management of these services. The manageability of the number of schools and teachers for which supervisors are responsible depends on the distance between schools and the scarcity of transport. P.29	High/High	FI
De Grauwe (2007)	Botswana, Namibia,	Research on school supervision in Africa shows the lack of satisfaction among teachers and supervisors with the impact of supervision on the classroom. The most evident reason - and the	Low/High	CL

A: Availability of financial and material resources to visit schools (lack of resources resulting in limited visits to schools, particularly in remote areas)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
	Tanzania, Zimbabwe, Mali	one that supervisors regularly quote - concerns the lack of resources. Many supervisors do not have the necessary vehicles nor the funds to travel, while at the same time the number of schools per officer has grown. P.710		
De Grauwe (2008)	Various (Asia, Africa)	Teachers who are left unsupervised for many years. P.4	Low/High	CL
Herselman and Hay (2002)	South Africa (Eastern Cape)	The quality assessment organisations mentioned the following issues and concerns in the facilitation of quality assurance at the 10 piloted schools: short time frames and notices, unavailability of transport, shortage of instruments/forms, shortage of manpower, non-payment of claims for transport. p.242	Low/Med	FI
Macpherson (2011)	Timor Leste	The Inspectorate was established in this context in 2008, with 65 school inspectors directed to sustain the quality and accountability of between 20 and 30 schools each, some so remote they took all day to reach by motorbike and on foot. Many of the schools in the remote inland mountain ridge of Timor Leste are cut off during the rainy season. Inspectors called for laptops, fuel for motorbikes and vehicles, and accommodation for regional office personnel moving to take up their posts. P.190	Low/Med	FI

A: Availability of financial and material resources to visit schools (lack of resources resulting in limited visits to schools, particularly in remote areas)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Mazibuko (2007)	South Africa (Kwazulu-Natal)	Shortage of supervisors in the province, lack of educator training, particularly principals in whole-school evaluation and lack of support from the Department of Education, particularly District Offices, are regarded as the major factors that impede the proper implementation of whole-school evaluation. P.i	Low/High	FI
Moswela (2010)	Botswana	Owing to the isolation, remoteness and a poor road network, inspectors could not visit the schools frequently. P.71	Med/High	CL
Uwazi (2009)	Tanzania	Due to inadequate personnel, lack of transport, office space, equipment and housing, a school is inspected about once in every two years. P.3	Med/High	FI
Wanzare (2002)	Kenya	Numerous problems are associated with the present Kenya's system of school inspection. (8) transport; (9) planning inspection; (12) cost of inspection; (15) resourcing. P.6	Low/Med	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

B. Workload of school inspectors/ambiguity about main functions (and as a result limited visits to schools, particularly in remote areas)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study (Rig/Rel)***	**Type of evidence***
De Grauwe (2001)	Botswana, Namibia	The number of supervisors has not kept pace with the numbers of schools and teachers; as a result, the school/supervisor and teacher/supervisor ratios are high. P.14	High/High	FI
		The inspection service is still relatively young and its present structure lacks clarity or logic. Supervisors are asked to handle a load of administrative and pedagogical tasks. They face a heavy workload with few resources which leads to a lack of motivation. Workload is particularly an issue at secondary level where school inspectors are generally subject-specific. Many countries (e.g. Botswana) do not have a full staffing of subject supervisors. P.143		
De Grauwe (2007)	Africa	The obligation for many supervisors to offer support and exercise control, two contrasting activities, has led to (i) an internal role conflict and (ii) to regular conflict with teachers. P.711	Low/High	CL
De Grauwe (2008)	Various (Africa, Asia)	There is a profound conflict between the mandate of the service and its resources. The mandate is very demanding: to exercise control over and offer support to all schools and teachers, while informing schools of ministry policies and bringing school realities to the attention of decision-makers. The expansion in the numbers of schools and teachers has not been accompanied by an equal expansion in the numbers of supervisors, the evident result being that each supervisor has	Low/High	CL

B. Workload of school inspectors/ambiguity about main functions (and as a result limited visits to schools, particularly in remote areas)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study (Rig/Rel)***	**Type of evidence***
		so many schools under his or her charge that they simply cannot visit all schools more than once or twice a year, if at all. P.3		
Mazibuko (2007)	South Africa (KwaZulu-Natal)	The two supervisors who participated in this study maintain that they are unable to cope with the number of schools that is supplied by the National Department of Education. According to these officials the National Department of Education expects them to conduct whole-school evaluation in at least 3 000 schools a year in KwaZulu-Natal Province. P.196	Low/High	FI
Moswela (2010)	Botswana	Instructional supervisors' effectiveness is constrained by the much expanded secondary education system that has seen a massive increase in schools and teachers in a relatively short time. P.71	Med /High	FI
Wanzare (2002)	Kenya	[T]he amount of observation of classroom teaching by inspectors is uneven and disturbingly small. Inspectors spent most of their time solving administrative problems with headteachers, and that teachers were never helped as adequately as they should. Due to paucity of time at the disposal of school inspectors, the school inspections, wherever held, have become superficial and a mere formality. P.11	Low/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

C: Lack of expertise, status and credibility of school inspectors/inspections				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Churches and McBride (2013)	Nigeria	Points to the need for employment/deployment of qualified and experienced inspectors, induction of new inspectors and capacity building for practising inspectors. P.21	Low/High	CL
Darvas and Balwanz (2014)	Ghana	Supervisors' lack of authority. P.136	High/Med	CL
De Grauwe (2001)	Botswana, Namibia, Tanzania, Zimbabwe	Zimbabwe, Botswana and Namibia see the need to define more clearly the tasks of the different categories of supervision and support staff, and to integrate this (in Namibia) into a performance management system which is then used to evaluate school inspectors, monitor their work and provide them with support. P.16	High/High	FI
		The need for more and better training – both at the beginning and during their career – is a recurring demand of supervisors in the Eastern and Southern African region. As supervisors, they will then be on a lower grade and salary scale than some secondary principals, a position which renders their inspection job very difficult. P.72		

C: Lack of expertise, status and credibility of school inspectors/inspections				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		Absence of coherent and motivating professional development programme. P.75		
		Supervisors are assessed mainly on the basis of the quantity of their work: the number of visits made and the number of reports written. P.79		
		Supervisors are generally recruited from among subject specialists, few have experience in school management and they occupy a post at the same grade as the secondary school principals. Many principals therefore do not consider supervisors as their superiors and refuse their advice. P.129-130		
De Grauwe (2008)	Various (Africa, Asia)	Supervisors tend to spend little time in each school. Their visits lead almost unavoidably to superficial reports, which have little credibility in the eyes of teachers. Principals and teachers criticise visits for their superficial and artificial character (can one judge the performance of a school or teacher on the basis of a single visit a year?) P.5	Low/High	CL
Harber (2006)	Eastern and Southern Africa	The need for more and better training – both at the beginning and during their career – is a recurring demand of supervisors in the Eastern and Southern African region. While a number of	Med/Med	FI

		C: Lack of expertise, status and credibility of school inspectors/inspections		
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		in-service courses take place, they are not integrated within an overall capacity-building programme, neither do they focus sufficiently on supervision issues. P.621		
		There is limited availability of instruments, such as manuals and guidelines, which help supervisors to fulfil their tasks effectively and break to some extent the feeling of isolation. Some instruments are available, but few go beyond the rather administrative forms and circulars P.621.	Med/Med	CL
Jaffer (2010)	Pakistan	As the inspection system expanded and became more complex, several issues arose in the realm of human resource planning and management, including the status of the inspectors, the issue of seniority, credibility and authority, and the mechanisms for recruitment and selection. Compounding the issues were the lack of clarity and of written policies and procedures on many of these matters. Apparently no conscious effort was made to match the applicant's expertise and experience with the functions that the position required. Second, no specified process existed for recruiting the personnel. This practice opened the door for certain individuals to use connections and political influence to get transferred to a preferred post. Complicating this situation, the inspector's position was equivalent to the teaching grade of a high school teacher, so these teachers and the inspectors were at the same grade and salary scale. Hence, individuals could not be held accountable for sub-standard performance or rewarded for good performance. No pre-service training was provided to prepare the inspectors for the specific	Low/High	FI

C: Lack of expertise, status and credibility of school inspectors/inspections				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		roles and responsibilities of the post. These lacunae further weakened the inspectors' position and authority, and also impacted adversely on the efficiency of the inspection system. P.378		
		There is still no systematic approach to collecting, analysing, or reporting data. Even when reports are made, the findings are not always valid and credible. Hence, the existing monitoring and supervision system is deficient and inconsistent, rendering it ineffective and of little help in improving the quality of the education system. P.380		
Macpherson (2011)	Timor Leste	There was strong demand for leadership within the inspectorate; these school inspectors wanted more regular meetings with their district superintendents to handle the load of investigations, in addition to the scheduled visits to schools and occasional meetings with the regional inspector to develop precedents for deciding common types of cases. P.199	Low/Med	FI
Moswela (2010)	Botswana	If instructional supervision is to help the teachers in improving their teaching then they should play an active part in the process and instructional supervision should not be carried out only	Med /High	FI

		C: Lack of expertise, status and credibility of school inspectors/inspections		
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		towards the end of the year when teaching has stopped and students are writing examinations. P.79		
Santiago et al. (2012)	Mexico	It further highlighted the lack of a strong core of professionally trained supervisors. P.155	Med/Med	FI
Uwazi (2009)	Tanzania	The School Inspectorate needs to know if and to what extent its efforts and recommendations have contributed to its objective of facilitating good performance in schools. School inspections have therefore to be monitored and evaluated as a part and parcel of the learning process about what works or doesn't and for improvement of future inspections. P.7	Med/High	FI
Wanzare (2002)	Kenya	Numerous problems are associated with the present Kenya's system of school inspection. (1) professionalism; (2) attitudes and commitment; (6) foci of inspection; (7) Inspectorate autonomy; (10) Inspectorate-university Partnerships; (13) inspector recruitment, selection, and deployment; (14) adequacy of inspection; (18) Inspectorate titles. There had been a lack of clear policy of identifying suitable candidates to be recruited as school inspectors and, consequently, unsuitable personnel find their way into the Inspectorate and put the integrity of some officials into question...	Low/High	FI

C: Lack of expertise, status and credibility of school inspectors/inspections				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		Because of apparent lack of incentives … there is a lack of commitment and initiatives on the part of school inspectors to their inspectoral roles which has further led to the inspectors performing inadequately. … because school inspectors are incompetent and are untrained, they are unable to monitor and to evaluate educational programs effectively. P.6		

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

D: Inspections focus on bureaucratic/administrative issues				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
Chen (2011)	Indonesia	Monitoring instruction inside classrooms is not commonly covered during these visits, and reviewing the school budget is also not a routine task of the supervisors.. PP.13-14	Med/Med	FI

D: Inspections focus on bureaucratic/administrative issues				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
Darvas and Balwanz (2014)	Ghana	Supervisors' often just visiting to check figures like student attendance and not check classes. P.136	High/Med	FI
De Grauwe (2007)	Africa	When supervisors visit schools, their focus is on administrative control. P.711	Low /High	CL
Jaffer (2010)	Pakistan	Some have argued that inspection judges school performance only at one point in time, focusing on compliance with administrative and legal requirements, rather than on the processes by which institutions improve. P. 376	Low/High	CL
Opoku-Asare (2006)	Ghana	'Evidence from the teacher interviews also indicates that some inspectors are more concerned with teacher attendance, preparation lesson notes and punctuality to school than with standards in teaching and learning'. P.113	Med /High	FI
Santiago et al. (2012)	Mexico	A tendency for supervisors to focus on administrative, bureaucratic and syndicate activities which take schools' time and focus away from the improvement of outcomes. P.155	Med / Med	CL
Wanzare (2002)	Kenya	…, school inspectors have the tendency to focus on school buildings and administrative systems rather than on teaching and learning, with minimal attention to the identification and improvement of educational standards. … because of conflicting inspection standards, school	Low /High	CL

D: Inspections focus on bureaucratic/administrative issues				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study**	***Type of evidence**
		inspectors have the tendency to inspect everything and sometimes they make contradictory proposals. … added that school inspectors sometimes have the tendency to over-emphasise certain areas, such as the smartness of the teacher, instead of the way the teacher teachers. On this debate, … the inspectors often seem to be checking up schools rather than trying to identify and improve standards. Plans for inspection of schools have been over-ambitious and, consequently, they are seldom carried out. P.9		

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

E: Tone of voice of school inspectors				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence ***
De Grauwe (2007)	Africa	When supervisors visit schools, their attitude can be condescending and their focus is on administrative control, which offers them a sentiment of power. P.711	Low/High	CL
De Grauwe (2008)	Various (Africa, Asia)	Principals and teachers object to the attitude of supervisors, which many feel is disrespectful of their professionalism. Supervision visits which teachers consider disrespectful, if not demeaning, rather than helpful. P.5	Low / High	CL
Wanzare (2002)	Kenya	Similarly, … that inspection of schools in Kenya has at times been marked by impromptu, irregular visits by some inspectors with the object of 'catching' the teachers doing the wrong. Some school inspectors have been criticised for being harsh to teachers and for harassing teachers even in front of their pupils (Bowen, 2001; Isolo, 2000; Kamuyu, 2001; Nakitare, 1980; Ndegwa, 2001). According to Isolo, many school inspectors have developed the following questionable habits: (a) they look down upon teachers with resentment and suspicion; (b) they demand bribes from teachers in order to make favorable reports; (c) they are dictatorial and have taken the attitude of 'do as I say or get in trouble' and (d) they work with unsmiling determination. Describing unprofessional conduct of school inspectors, Kamuyu (2001) noted that some inspectors behave like outsiders whose sole mission is to work against teachers to prove that no teacher is competent. Similarly, Masara (1987) noted that some inspectors reportedly visit schools to boss and to harass teachers instead of helping them solve professional problems. The unprofessional	Low / High	CL

E: Tone of voice of school inspectors				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence ***
		behavior of some school inspectors has had the following serious negative consequences. Poor relationship between inspectors and teachers (Masara, 1987) P.10		

* * Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

F: Lack of consequences/ lack of co-ordination with other services to follow-up				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Alcazar et al. (2006)	Peru	In short, there is little evidence for the inspections story, perhaps because inspections are believed by teachers to have no consequences. P.130	High/Med	CL
Brock (2009)	Four counties in Gansu, China:	By giving power to inspectors (to report, to propose changes, to propose support) the process of SDP was also enhanced and given prominence as an important process that set out school goals which could be measured by inspectors. P.457	Med/Med	FI

		F: Lack of consequences/ lack of co-ordination with other services to follow-up		
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
	Dongxiang, Jishishan, Hezheng and Kangle			
Chen (2011)	Indonesia	Key aspects of quality assurance and consequences are missing from inspection visits. The lack of rewards and sanctions for good or bad performance leaves the system weak. P.13	Med/Med	FI
Churches and McBride (2013)	Nigeria	Points to the need for adequate legal provisions for enforcing compliance by schools and proprietors. P.21	Low/High	CL
Crouch and Winkler (2008)	Uganda	Lack of mechanisms and consequences to hold leaders and teachers to account. The inability to enforce inspection standards creates weak incentives for teachers and headmasters to even be present at school, much less to perform effectively. P.15	Low/High	FI
De Grauwe (2001)	Namibia, Tanzania	Coordination is difficult, especially between the supervision service and other services which work towards pedagogical improvement, such as teacher training, teacher resource centres, curriculum development and examinations. Recommendations made in inspection reports and addressed to the administrative and/or pedagogical authorities, remain words in the	High/High	FI

F: Lack of consequences/ lack of co-ordination with other services to follow-up				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		wind, which frustrates the school staff as well as the supervisors. Supervisors feel frustrated and constrained by their lack of autonomy and authority to take action on their own recommendations. P.143		
Mazibuko (2007)	South Africa (KwaZulu-Natal)	The study shows that findings and recommendations of the supervisors have not been addressed. This is attributed to the poor communication between the District Office and schools to address areas identified by schools in their improvement plans. What compounds the problem is that the role of the Circuit Office in whole-school evaluation is unclear. This has a bearing on the functioning of schools because the Circuit Office is closer to the schools and should, therefore, communicate with them regularly. P.307	Low/High	FI
Opoku-Asare (2006)	Ghana	Many inspectors are close with headteachers and teachers in certain schools, depending on the relationship between a teacher and the head-teacher and, the head and the inspectors, adverse reports do not go beyond the school gate, and teachers are given a second chance to pass the test without having the report put on their file. P.112	Med/High	FI
Uwazi (2009)	Tanzania	It shows that the school inspectorate programme is not functioning properly … by not effectively communicating and following up on implementation of recommendations. P.1	Med/High	FI

F: Lack of consequences/ lack of co-ordination with other services to follow-up				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Wanzare (2002)	Kenya	Numerous problems are associated with the present Kenya's system of school inspection. (3) feedback and follow-up. P.6	Low/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

Table A6.1.8: Evidence of impact of school inspections (school improvement and unintended consequences)

Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Alcazar et al. (2006)	Peru	In short, there is little evidence for the inspections story. P.130	High/Med	CL
Brock (2009)	Four counties in Gansu, China: Dongxiang,	By giving power to inspectors (to report, to propose changes, to propose support) the process of SDP was also enhanced and given prominence as an important process that set out school goals which could be measured by inspectors. P.457	Med/Med	FI

Study	Country	Evidence (quotations)	Quality of study (Rig/Rel)*	Type of evidence*
	Jishishan, Hezheng and Kangle			
Crouch and Winkler (2008)	Uganda	The lack of an effective inspection system at the district level. P.15	Low/High	FI
De Grauwe (2007)	Africa	Supervision's feeble impact on quality. P.711	Low/High	CL
De Grauwe (2008)	Various (Asia, Africa)	While there has been rather little systematic research on the functioning and the effectiveness of supervision systems, the anecdotes in this regard are plentiful …; supervision reports which are shelved without any action being taken. P.4	Low/High	CL
Herselman and Hay (2002)	South Africa (Eastern Cape)	According to the primary school principals and facilitators: 'quality assurance mechanisms and procedures were at that stage not part of their school's strategic planning process' (Principal1). It also became apparent that facilitators emphasised mechanisms and procedures in their classrooms to enhance the quality of teaching and learning, while managers used school policy, subject policies, book controls, class visitations and subject-standard staff meetings as mechanisms and procedures for quality assurance. P.244	Low/Med	FI

Study	Country	Evidence (quotations)	Quality of study (Rig/Rel)*	Type of evidence*
Jaffer (2010)	Pakistan	Neither the system of evaluation, nor the quality of education provision has improved, despite much effort. The existing monitoring and supervision system is deficient and inconsistent, rendering it ineffective and of little help in improving the quality of the education system. P.376	Low/High	CL
Mazibuko (2007)	South Africa (KwaZulu-Natal)	The study revealed that whole-school evaluation is not being implemented. The study shows that findings and recommendations of the supervisors have not been addressed. P.i	Low/High	FI
		According to the principals and educators who participated in this study, their first response was to conduct meetings to discuss the reports and to decide on the way forward. They maintained that they formed teams in their schools to deal with the logistics for drawing up the school improvement plans. These teams read the reports and recommendations made by the supervisors. According to them, these plans detail how they were going to address the recommendations of the supervisors, as well as the areas that they identified during school self-evaluations. All principals interviewed maintained that after the whole-school evaluation was conducted they gave reports of the findings and recommendations of the supervisors to the parents and guardians of the learners. According to the principals, this was done to ensure that all stakeholders have an input in the school for the benefit of the learners. But all principals and educators who participated in this study maintained that although their schools have drawn up improvement/development plans, nothing has happened in terms of addressing areas identified by the supervisors as areas that need improvement. P.258		

Study	Country	Evidence (quotations)	Quality of study (Rig/Rel)*	Type of evidence*
		The study shows that findings and recommendations of the supervisors have not been addressed. Often the schools' self-evaluation did not tally with the supervisors' findings. For example, one supervisor said that educators often give themselves high scores. Both supervisors alleged that educators may give themselves high scores because remuneration is involved. P.227		
		All principals and educators interviewed in this study maintained that they learnt something from whole-school evaluation. P.275		
		In schools where educators have a negative attitude towards the whole-school evaluation and the supervisors, principals use whole-school evaluation to threaten educators. These principals create an impression that whole-school evaluation is there to punish educators who do not do their work effectively. An educator in school A confirmed that most principals use whole-school evaluation to threaten them if things are not going well in the school. As a result by the time the external evaluators come to the school to conduct the whole-school evaluation educators have already developed a negative attitude towards the external evaluators. P.206		
Macpherson (2011)	Timor Leste	This suggests that the scale of corruption in the misuse of school grants is probably limited and may well be contained by school inspectors policing transparency in the collection and disbursement of school grants, while not engaging in the processes themselves. P.201	Low/Med	FI

Study	Country	Evidence (quotations)	Quality of study (Rig/Rel)*	Type of evidence*
Opoku-Asare (2006)	Ghana	School inspections are often pre-announced and lenient. 'The tip-off, they said, enables the teachers concerned to prepare adequately for the observation lessons. This enables those teachers to arm themselves with all the teaching materials they can possibly lay hands on and sometimes, rehearse the lessons they intend to teach for the exercise' P.113.	Med/High	FI
Uwazi (2009)	Tanzania	School inspection reports at secondary school level. These inspections are considered to be ineffective as national performance of students isn't improving.	Med/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

Table A6.1.9: Elements of context that impact on effectiveness of school inspections

A: Lack of resources in the system				
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Alcazar (2006)	Peru	A community's remoteness and poverty level are strong predictors of higher absence among teachers in its primary school, as is lower literacy among the parents of students. These factors could affect teacher absence through various channels, including by worsening the work environment for teachers and by reducing the ability of communities to induce good teacher performance. P.132	High/Med	FI
Darvas and Balwanz (2014)	Ghana	Inputs-focused initiatives respond to part of the 'access challenge,' however these programs do not address other issues (for example, age of initial enrollment, household expectations of child labor, school culture) that also act as access-barriers. Further, none of these programs directly address the structural inequities in the current system, including getting trained teachers to pupils with the greatest need. Data from this section show that students and populations who may require the most support to meet expected outcomes (for example, learning, primary completion, access to secondary), receive, on average, disproportionately fewer resources (for example, trained teachers, textbooks) from the government than their peers. Inequitable distribution of inputs creates a negatively reinforcing loop where children with the greatest need receive the fewest resources and opportunities, thereby reproducing cycles of poverty and inequality. Children from the northern regions, deprived districts, poor and rural households and	High/Med	FI

A: Lack of resources in the system				
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		ethnic and linguistic minorities are most disadvantaged by inequities in basic education service delivery. P.9		
		Inefficient allocation of qualified teachers, teacher absenteeism and loss of instructional time during the school day are three of the greatest inefficiencies in the current system. P.13		
De Grauwe (2008)	Various	Services which specialise in pedagogical support suffer from a somewhat comparable weakness. In many case, their advice is also benefiting mainly the schools closest to where these support services are located. Research undertaken on the role of resource centres for instance in India, Kenya, Nepal and Zambia … shows that they are generally not able to reach out to a large number of schools, and even where they succeed in doing so, they offer advice which is of little relevance to the situation of schools whose resources and context are too far away from the standard one that these services know and cater for. P.25	Low/High	CL
Herselman and Hay (2002)	South Africa (Eastern Cape)	Teachers claim that the main reasons for neglecting quality at primary schools are the large proportion of learners that cannot read the text books they are provided with; big classes with no equipment; the overloaded curriculum with inappropriate learning needs; the fact that only a quarter of the children finish primary school; spending more per capita on higher education than on primary education; limited access to pre-schooling. P.240	Low/Med	CL

A: Lack of resources in the system				
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Jaffer (2010)	Pakistan	First, Pakistan has a high proportion of very small rural primary schools in which one or two teachers are expected to teach all five grades. Even urban primary schools face a disparity in the allocation of teachers; whereas some schools have five teachers for five classes, others have only two teachers for five classes, and still others can have ten teachers for five classes. Also, the professional development opportunities for teachers are insufficient and inequitable. Moreover, no concerted effort has been made to improve the quality of the physical infrastructure and facilities, particularly for rural schools. Another key factor neglected in the discourse and action on quality is the role of leadership in school effectiveness and improvement. The school head's position is widely considered to be critical for all aspects of school effectiveness. But in Sindh, the position of the head teacher in primary schools has not been sanctioned. Heads are only appointed in the secondary schools (classes 9–10), and in the elementary schools, which include primary and middle schools, classes 1–8. In primary schools, a senior teacher is usually designated as the in-charge person to look after the school's day-to-day management, under the supervision of the SPE or an ADOE. However, teachers do not take on this position very enthusiastically, because it carries with it only responsibility, but no authority. In view of the scenario presented above, one could safely conclude that, even if the inspection system in Sindh is revamped, improvements in the quality of education will still be elusive, unless the entire system is given attention. P.384	Low/High	FI

A: Lack of resources in the system				
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		Many government teachers hold other jobs to supplement their income, for example teaching in a private school or managing their own tutoring centre. This creates many issues, including teachers paying more attention to their other work. P.387		
Macpherson (2011)	Timor Leste	The absence of a common language to teach teachers with and the lack of teaching resources in any language. Other major challenges are the poor quality of education in terms of teacher capability, teacher qualifications and the curricula. There is high absenteeism of teachers and students, high attrition rates, high repetition rates, high adult illiteracy, a gender imbalance with only 30 per cent of teachers in primary schools being women, poor classroom facilities, teacher: student ratios typically about 1:40, and about one-third of the population being of school age. P.190	Low/Med	CL
Mazibuko (2007)	South Africa (KwaZulu-Natal)	In some instances, the principal of a disadvantaged school has a class to teach. As a result this principal ends up not concentrating on his/her managerial roles because he/she has to ensure that the class that he/she teaches does well in order to boost the morale of other staff members. An educator in school C points out that in most disadvantaged schools where there are support personnel, they have administration clerks but they do not have financial officers and other support personnel. As a result the principal of a disadvantaged school ends up doing everything. The principal of school A also maintained that the principal of a disadvantaged school has a	Low/High	FI

A: Lack of resources in the system				
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		problem acquiring sponsorships. Whereas, principals of advantaged schools are able to acquire sponsorships from the private sector. P.285		
		Principals and educators interviewed in this study maintained that financial constraints and other crucial issues make it difficult for their schools to address areas or issues identified by the supervisors as areas that need attention for the development of these schools. Lack of resources, overcrowding, lack of support from DfE (resources, information). Principals and educators claim that financial constraints prevent them from addressing the recommendations of the supervisors. P.229		
Moswela (2010)	Botswana	Instructional supervisors' effectiveness is constrained by the much expanded secondary education system that has seen a massive increase in schools and teachers in a relatively short time. P.71	Med/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

B: Lack of knowledge in schools				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Herselman and Hay (2002)	South Africa (Eastern Cape)	Teachers claim that the main reasons for neglecting quality at primary schools are the concept of quality and how to achieve it are exceptionally complex and difficult. P.240	Low / Med	CL
Mazibuko (2007)	South Africa (KwaZulu-Natal)	Clear understanding of whole-school evaluation and its implications can lead to proper implementation. But the study reveals that participants, particularly principals and educators, had only a general understanding of whole-school evaluation, as they had not undergone training. This implies that principals and educators do not really understand the pros and cons of whole-school evaluation. Because of this shortcoming, schools cannot conduct whole-school evaluation effectively. The principal of school D contended that most educators have a negative attitude towards whole-school evaluation. This can be attributed to the fact that most educators have not been trained on whole-school evaluation and may not fully appreciate the significance of whole-school evaluation in schools. Lack of understanding and knowledge impede most educators from fully participating with the supervisors during the external evaluation. P.201	Low / High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

C: Lack of consequences overall				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Alcazar et al. (2006)	Peru	[T]eachers appear to have few incentives to avoid absenteeism or minor misconduct, at least in practice. Hiring decisions are ostensibly made on merit but, according to informed observers, are substantially influenced by connections and bribery. Transfers to desirable locations appear also to be mediated by these non-meritocratic factors, reducing the incentive to perform well. Salary is set primarily based on tenure and characteristics of the job or location, rather than on performance in a given position. And serious disciplinary sanctions are sufficiently difficult to implement, in practice, that they appear unlikely to restrain teacher behavior. We should note that the lack of formal incentives related to salary or tenure does not necessarily mean that teachers will perform poorly. P.122	High/Med	FI
Crouch and Winkler (2008)	Uganda	An important weakness in governance is the weak incentive for teachers and headmasters to perform and the corresponding lack of accountability by schools to either parents or the education ministry or the district education office. P.15	Low/High	FI

*FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

D: (Lack of) alignment in the education/accountability system				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
De Grauwe (2001)	Botswana, Namibia, Tanzania, Zimbabwe	Relationships between supervision and the other pedagogical services are close and institutionalised in Botswana, supervisors being members of committees and panels in charge of curriculum development, teacher training and examinations. In Zimbabwe, their involvement is less institutionalised but still quite intense: supervisors serve as resource persons in training and participate in writing test items, marking examinations and preparing evaluation reports. In Namibia, however, the situation tends to the opposite, with no formal contacts and very few informal ones between supervisors and other pedagogical support staff. Supervisors are, for instance, not represented on the examination board, neither will they be on the Regional Education Forum. In Tanzania, in principle, supervisors do sit on curriculum panels and help with examinations, but in practice their involvement in pedagogical improvement is limited because of the practical constraints on their work. P.44	High/High	FI
		Co-ordination is difficult, especially between the supervision service and other services which work towards pedagogical improvement, such as teacher training, teacher resource centres, curriculum development and examinations. The follow-up to school visits suffers from this lack of coordination. Recommendations made in inspection reports and addressed to the administrative and/or pedagogical authorities, remain words in the wind, which frustrates the school staff as well as the supervisors. P.143		

D: (Lack of) alignment in the education/accountability system				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
De Grauwe (2008)	Various (Africa, Asia)	Very regularly implementation of recommendations demands coordination between different agencies and offices, which goes counter [to] the sense of independence of many such officers and especially of supervisors. P.5	Low/High	CL
De Grauwe (2008)	Lesotho, Botswana, Namibia, Tanzania, Zimbabwe, Nigeria	There is also the risk of conflicts between these groups and confusion among teachers who get contrasting advice e.g. from inspectors and pedagogical advisors, who have different opinions on the correct teaching methods. P.7	Low / High	CL
Mazibuko (2007)	South Africa (KwaZulu-Natal)	Clarity of roles in follow up from school inspections: Lack of clarity on the role of the Circuit Offices. Circuit Offices are closer to the schools and are in constant contact with the schools. The Circuit Offices' proximity with the schools makes it easier for the Circuit Offices not only to know schools better but also to understand and know the needs and strengths of the schools. But the role that should be played by the Circuit Offices in whole-school evaluation is not clearly stated. As a result, the two superintendents who participated in this study indicated that they do not know what is actually happening in the schools in terms of the implementation of whole-school evaluation. The principal of school C also confirmed that the role of the Circuit Office is not clear	Low/High	FI

D: (Lack of) alignment in the education/accountability system				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		and this makes it difficult for them to address the findings and recommendations of the supervisors. P.273 According to this principal, what compounds the problem is that the District Office is far away from KwaMashu area and schools are not always allowed to go to the District Office directly. All participants in this study are of the opinion that the role of the Circuit Offices in whole-school evaluation is not quite clear. This has a negative impact in the implementation of whole-school evaluation since schools always look to the Circuit Offices for guidance. P.271		

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

E: Culture of accountability and political context				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Chen (2011)	Indonesia	The bottom-up pressure from parents seems weak in Indonesia's public primary schools, but the top-down pressure from government supervisory bodies is slightly more significant. P.13	Med/Med	FI
Darvas and Balwanz (2014)	Ghana	Citizens' increased expectations of government have been revealed in recent political debates and media coverage on issues related to national health insurance, civil service salary reform (for example, the single-spine salary system), free senior high school and support for accelerated development of economically marginalised regions (for example, the Savannah Accelerated Development Authority). P.3 Many inequities associated with powerful constituencies appear resistant to change. Specifically, many inequities, such as the allocation of trained teachers and the insufficient provision of support to deprived districts and populations, appear perpetuated by interests associated with powerful constituencies such as teachers unions, the upper middle class and government decision-makers responsible for allocation and management of public resources. P.3	High/Med	FI
		Further, as more families enter the middle class and urbanise, many pupils are exiting the public system and paying for elite private schools. The influence of powerful interests and the exit of influential constituencies from public schools each reduce pressure on government to reform basic education and leaves poorer families worse off. P.4		

E: Culture of accountability and political context				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
De Grauwe (2008)	South Africa	In South-Africa, the government at federal level and in many provinces attempts to undertake whole school evaluations, but this encounters severe resistance from teacher unions, who feel that such evaluation should be preceded by comprehensive teacher professional development programmes... A particularly important factor in the context is the position of power of different actors, some of whom may benefit from these reforms (to make supervision more effective) while others oppose them. P.9	Low/High	CL
De Grauwe (2008)	Various (Africa, Asia)	It is pertinent here to point out that the resistance to supervision is an expression of a more general resistance to external evaluation on the part of teachers. This can be interpreted as a reflection of their professional autonomy or as a corporatist protection of their privileges. P.5	Low/High	CL
		The interventions by these school monitoring services are reflective of the State's intervention as a whole: because the State is incapable to fulfil its mandate, authorities tend to focus on those groups whose support is important to their survival. The politically less vociferous groups are to some extent abandoned and will at times, with their own scarce resources and with the help of non-governmental organisations, set up their own services. P.25		

E: Culture of accountability and political context				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Herselman and Hay (2002)	South Africa (Eastern Cape)	Teachers claim that the main reasons for neglecting quality at primary schools are uncovered information that becomes an embarrassment to the policy-makers. P.240	Low/Med	CL
Jaffer (2010)	Pakistan	Some respondents attributed the teachers' lack of accountability to political interference. Both inspectors and supervisors said there was no clearly laid out job description which provided expected standards of performance for them and could serve as a tool for appraising inspection and supervision. The respondents were of the view that even a serious offence or lapse in performance could not be punished because of an individual's political affiliation, whereas someone known for good performance might be transferred because of political pressure. P.386	Low/High	FI
Kingdon and Muzammil (2013)	Uttar Pradesh (India)	Strong role of teacher unions and strong representation of teachers in elected bodies prevents quality improvement from inspections and examinations as teachers resist change and improvement and have a strong position to do so. Evidence suggests that these accountability sanctions and probity procedures have not been effectively implemented because teachers resist them by pressurising the District Inspectors of Schools, both through their unions and via political pressure from teacher politicians. Nor did the District Inspectors of Schools and other officials exercise any authority over them as the erring teachers were often supported by powerful teachers' associations. PP.259-260	Med/High	FI

E: Culture of accountability and political context				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Mazibuko (2007)	South Africa (KwaZulu-Natal)	The other thing that creates a negative attitude towards whole-school evaluation and/or supervisors, especially among the educators, is that most educators do not want to be observed when teaching. This supervisor summed this up: 'They (educators) really do not want us (supervisors) in their classes. They do not want us to see what they are doing in the classroom'. P.206	Low/High	FI
		Most of the former Coloured schools (Black schools) show them on the first day that they are not concerned about their presence. According to the supervisors, these schools continue operating as if nothing has happened. The supervisor explained this as follows: 'You could see that these people do not care a damn'. Meanwhile, according to this supervisor, the former Indian schools always try to impress the supervisors. According to this supervisor, most of the former Indian schools pretend they appreciate the supervisors' visit. P.208		
Opoku-Asare (2006)	Ghana	Many inspectors are close with headteachers and teachers in certain schools, depending on the relationship between a teacher and the head-teacher and, the head and the inspectors, adverse reports do not go beyond the school gate, and teachers are given a second chance to pass the test without having the report put on their file'. P.112	Med/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

Table A6.1.10: Mechanisms of impact of school inspections

A: Involvement of local community/parents/ other service providers in accountability of schools/organising school inspections more locally				
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Alcazar et al. (2006)	Peru	Active oversight and involvement of the local community may improve performance through better monitoring.	High/Med	CL
Chen (2011)	Indonesia	This appears to cast doubt on whether the accountability of schools to parents would work in Indonesia, where community harmony is highly valued, and a majority of parents are reserved and do not openly complain or express dissatisfaction. P.13	Med/Med	CL
Crouch and Winkler (2008)	Uganda	Weak inspection (lack of enforcement of inspection standards, weak incentives) is not compensated for by giving PTAs and school management committees the information, capacity, and authority to take action to reduce absenteeism and improve performance. P.15	Low/High	FI
De Grauwe (2001)	Botswana, Namibia,	To allow for a closer and more regular supervision of schools, these are brought together into clusters and officers are given the responsibility for one such group of schools, with their office sometimes being located at that level … following logically from the above, there is an increased stress on in-school supervision and support. In Botswana, Staff Development Committees are	High/High	FI

		A: Involvement of local community/parents/ other service providers in accountability of schools/organising school inspections more locally		
Study	Country	Evidence (quotations)	Quality of study (Rig/Rel)*	Type of evidence*
	Tanzania, Zimbabwe	established in secondary schools: they carry out needs assessment and draw up programmes for school- based training, using expertise from within or outside the school. In Zimbabwe, headteachers of both primary and secondary schools have received or are receiving training in school management. In principle, each school has a Board, with members of the administration and teacher representatives, which supervise, assist and advise teachers, and also has to assess and recommend them for promotion. Fourthly, the civil society and certain school communities are gradually being given an increased role in monitoring the functioning and quality of schools. P.147 Many countries, in their attempts to reform and innovate supervision, are increasingly relying on in-school or community-based strategies (such as resource centres, school clusters, in-school supervision by the principal or by peers, school-based management) to complement – if not to replace – external supervision and support. P.8 The focus shifts from individual teacher inspection to school evaluation. P.17		
De Grauwe (2007)	Africa	Supervision can be targeted at the teacher as an individual or at the school as an institution. This distinction is not a matter of detail: a concentration on the teachers makes them, as individuals, responsible for quality, while a focus on the school recognises the importance of its functioning and of the interactions between principal, teachers and parents. Many French-speaking African	Low/High	CL

		A: Involvement of local community/parents/ other service providers in accountability of schools/organising school inspections more locally		
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		countries do not have school supervision, but only teacher supervision, while in much of English-speaking Africa, both can take place. P.710		
		Several strategies have as an objective to make external supervision more effective, by bringing supervision closer to the school. P.711		
De Grauwe (2008)	Various	A second condition (for effective supervision) relates to the State having sufficient resources to ensure that its supervision reaches out to all schools on a fairly regular basis, as this is needed to guarantee the respect of these standard rules and regulations. P.24	Low/High	CL
Herselman and Hay (2002)	South Africa (Eastern Cape)	The focus areas also involve all stakeholders (management, School Governing Bodies, facilitators, learners, school safety and infrastructure). It is, important, however, to note that the process of quality assurance should not only be externally driven. All stakeholders and especially the Government should be involved in the process of quality assurance by establishing a framework and an implementation plan of evaluation. Facilitators should take ownership of and be accountable for their own teaching by being part of the self-evaluating process. It would ensure that the quality assurance process is not an 'added on' approach and that it focuses more on accountability than on improvement. P.241	Low/Med	CL

A: Involvement of local community/parents/ other service providers in accountability of schools/organising school inspections more locally				
Study	**Country**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Moswela (2010)	Botswana	The quantitative growth in the secondary education and the distance between schools has meant that visits to schools by inspectors could no longer be sustained from a central office at headquarters and this necessitated the decentralisation of the inspection office. The decentralisation of the Inspection and Field Services from headquarters to the regions was therefore influenced, in the main, by the expansion secondary education. This decentralisation placed education officers in the regions while the principal education officers remained at head office. Decisions that directly affected curriculum implementers (teachers) were (with the expansion of the education system and the remoteness of schools) taken far away at head- quarters without their input. A two-way exchange between inspectors and teachers needed to be created 'so that classroom teachers do not feel abandoned or that their contribution is unimportant. P.72	Med/High	CL
Santiago et al. (2012)	Mexico	A number of key school agencies and types of personnel already exist in Mexico with the potential to support self-evaluation in all schools and undertake new roles in a more complete effective school evaluation model. P.153	Med/Med	FI
Wanzare (2002)	Kenya	However, there is no clear indication regarding accessibility of the reports by teachers, parents, and any other interested parties. P.12	Low/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

B: Strengthening internal evaluations to increase monitoring and capacity building for improvement				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study (Rig/Rel)***	***Type of evidence**
De Grauwe (2001)	Botswana, Namibia, Tanzania, Zimbabwe	The fourth point in this regard (improving the effectiveness of supervision) might well be the most significant. School heads are in a similar situation. Overcoming these constraints implies, however, a profound change in the management system and culture and – but less crucially – a strengthening of management capacities at district and school levels. P.144	High/High	CL
De Grauwe (2007)	Africa	Several strategies have as an objective to make external supervision more effective, by strengthening school-level supervision. Schools are increasingly asked to start a process of self-evaluation, for instance through the preparation of school improvement plans. However, in many countries this has been limited to a simple demand by ministries for schools to prepare a plan, without any assistance or guidance, with mixed success. P.711	Low/High	CL
De Grauwe (2008)	Mozambique, Senegal	A second series of reform trends aims at strengthening internal school evaluation processes. The key rationale for this emphasis on internal evaluation is the conviction that sustainable change in the school demands participation and commitment by the teachers. These internal evaluations can involve a cluster of neighbouring schools or the individual school. School clusters have been throughout the years a popular strategy, which has many objectives including strengthening	Low/High	CL

B: Strengthening internal evaluations to increase monitoring and capacity building for improvement				
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study (Rig/Rel)***	***Type of evidence**
		supervision within this cluster. This not only allowed schools to learn from such a visit; it also strengthened the linkages between schools and broke possible isolation. Moreover, it made school staff feel responsible for its own improvement and proud of their success. P.8		
De Grauwe (2008)	Various (Africa, Asia)	At times, there will be conflict between the internal and external evaluation processes. This is the case mainly in countries where the preparation of school development plans is being officially encouraged but not taken very seriously by the supervisors who stick to business as usual. This situation is possibly harmful because it only widens the gap between supervisors and teaching staff. Both parties might well end up by using totally different frameworks and criteria for making judgments about school practices. It could be argued that the promotion by ministries of education of school self-evaluation serves objectives which have very little to do with school improvement, but are a reaction to the Ministry's weak capacity to regulate the whole system. In those developing countries, where school improvement programs have gained in popularity, they represent more the agenda of international agencies or NGO's than a change in culture within the education system. P.15	Low/High	CL
Herselman and Hay (2002)	South Africa (Eastern Cape)	One way of installing a culture of quality is to establish selfevaluation processes at schools. Self-evaluation makes provision for reflective practices which help teachers to reflect on their own practices. It forces them to ask questions such as 'Why am I doing this?' and 'How can I improve what I am doing?' on a continuous basis. In this way they become reflective practitioners who are	Low/Med	CL

		B: Strengthening internal evaluations to increase monitoring and capacity building for improvement		
Studies	**Countries**	**Evidence (quotations)**	**Quality of the study (Rig/Rel)***	***Type of evidence**
		more interested in improving their practices than in trying to comply with the pressures of accountability. Quality assurance mechanisms and procedures form part of a continuous system of review and can be designed to serve a positive purpose in furthering the interest of the school, staff and learners. …of evaluation performance of schools and that little, if any, comprehensive data exist on the quality of teaching and learning or on the educational standards achieved in the system. P.240		
		It is anticipated that the selfevaluation process will play a major role in the quality assurance processes of schools. The underlying assumption is that quality assurance systems are dependent on whole school evaluation in order to make meaningful interventions that will hopefully raise standards of performance and improve learners' achievement. P.241		
Wanzare (2002)	Kenya	The book [new inspection handbook], no doubt, is a valuable source for schools by enabling them to improve their self-evaluation capabilities and should be a reference for teachers, headteachers, inspectors, and board members. It seems to be a remarkably fresh way to view education quality and school improvement. Although the new inspection handbook provides a framework for the inspection of schools, however, it has numerous concerns. It seems to be too detailed, bureaucratic, and rigid to be of realistic and practical use in inspection of schools. P.21	Low/High	CL

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

C: Adapting inspections to local context to feedback on relevant issues and set expectations				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Barrett (2011)	Botswana and other non-specified countries	As relevance also refers to the recognition of learners' multiple socio-cultural identities, it demands that school processes and the intrinsic benefits of education are responsive to these identities. For example, … show how school processes may recognise or overlook the histories, identities and cultural practices of indigenous groups with implications for children's engagement in learning. Tshireletso (1997) observed parents from indigenous minority groups in Botswana disowning schools that have practices counter to their own cultural values, such as the use of corporal punishment. The choice of language of instruction is one powerful way in which education systems either recognise or diminish the ethnic and/or linguistic identity of learners. Recent research has drawn attention to the gendered experiences of girls and boys in schools, including the sexual harassment of girls in particular (e.g. Leach et al. 2003), with implications for the formation of their gendered identities and emerging sexual identities. Measuring learning outcomes tells us very little about how schools respond to and influence learners' socio-cultural identities. As Alexander (2008) has forcefully observed, quantifiable measures of quality are always partial as some aspects of educational processes can only be judged through observation against qualitative indicators (see also O'Sullivan 2006). The observation and judgement of	Low / Med	CL

C: Adapting inspections to local context to feedback on relevant issues and set expectations				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		processes is the complex work of school inspectorates and other educational supervisors and managers that international targets are too blunt an instrument to tackle. P.128		
Churches and McBride (2013)	Pakistan, South Africa	The involvement of stakeholders in the design phase produces readier acceptance of review and a better framework, requiring fewer subsequent adjustments. P.40	Low / High	CL
De Grauwe (2008)	Various	When a country is relatively homogeneous and with little disparities, supervision as a standardisation tool can have a positive impact and may not worsen disparities. In such a scenario, applying the same framework and norms throughout the country may indeed make sense. P.14	Low / High	CL
Moswela (2010)	Botswana	If instructional supervision is to help the teachers in improving their teaching then they should play an active part in the process and instructional supervision should not be carried [out] only towards the end of the year when teaching has stopped and students are writing examinations. The teacher's input in instructional supervision is too important to be overlooked. P.79	Med / High	CL
World Bank (2010)	Peru	The need for local and regional adaptation is, currently, being used as an excuse for mediocrity. It is possible to develop standards that are locally adapted yet that provide both ambition and a metric for accountability. Standards should be simple, should emphasise skill, and should be meaningful particularly to teachers and parents. Service or process standards should be developed	Low / Med	FI

C: Adapting inspections to local context to feedback on relevant issues and set expectations				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		over time via observation of successful practice under difficult or average conditions. Schools that outperform others under similar conditions could be studied, and the good practices they engage in should eventually find their way into the procedural norms and standards. P.207		

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

D: Developing standards, frameworks and guidelines				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Brock (2009)	Gansu province, China	Frameworks and guidelines are made available to both schools and inspectors. Availability to schools supported school development planning as a vehicle for change in which schools and local communities are brought together to create a unified approach to the school's development, and in which the relationship between the county education bureau and the school changes from top down to bottom up development. P.457	Med/Med	FI

D: Developing standards, frameworks and guidelines				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Harber (2006)	Uganda	Before you wanted to do things but there was no reference point. I didn't have a vocabulary and the workshop provided reasons and enabled you to defend your position. I used to just turn up to school but now warn ahead. We were seen as a threat, a witch hunt, now they are ready for us and are happier as previously they were very uncomfortable. Inspectors used to go to the village in the night and then sneak into the school to spy, now it is more collegial and there is more fair play. P623 I been involved in a survey using a score card to try to get the views of students' teachers and the community about what makes a good school – they score the school in terms of priorities and this sometimes leads to a hot debate. It helps participation and is a good way of identifying problems. It helped to solve a problem concerning PTA meeting. P.624	Med/Med	FI
Herselman and Hay (2002)	South Africa (Eastern Cape)	Apart from those schools involved in the pilot project, most other school managers and facilitators had a limited knowledge of quality and quality assurance processes. Draft documents on quality assurance or examples were not distributed among all principals. P.243	Low/Med	FI
Jaffer (2010)	Pakistan	There is still no systematic approach to collecting, analysing, or reporting data. Even when reports are made, the findings are not always valid and credible. Hence, the existing monitoring and	Low/High	FI

D: Developing standards, frameworks and guidelines				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		supervision system is deficient and inconsistent, rendering it ineffective and of little help in improving the quality of the education system. P.380		
Santiago et al. (2012)	Mexico	The tools developed by INEE [national institute for educational evaluation] are a good complement as they permit schools to engage in the more detailed self-evaluation of particular aspects of schooling such as the school environment or relations with parents. However, the plethora of guides, materials and instruments, however well conceived and valuable, will undoubtedly have confused schools as they searched for the recommended approach and were faced with too many options. P.52	Med/Med	FI
World Bank (2010)	Peru	Peru needs much clearer learning standards, especially in the early grades, and needs to focus particularly on reading achievement. These standards need to be developed and disseminated. The ambition to simultaneously develop standards for the whole system should be resisted. Peru needs to start with reading (and perhaps writing) standards, and with the early grades. Standards should be simple, should emphasise skill, and should be meaningful particularly to teachers and parents. Standards should be grade specific, or perhaps even specific to semesters within the school year. P.130	Low/Med	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

E: Openness and transparency				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence ***
De Grauwe (2001)	Botswana, Namibia, Tanzania, Zimbabwe	Openness and transparency are increasingly encouraged, implying that schools will be informed beforehand of visits. As a result, where these reforms are actually being implemented, teachers are beginning to consider inspectors as sources of help rather than of criticism. P.17	High/High	CL
Uwazi (2009)	Tanzania	There is limited transparency: the information concerning audits is not published implying that it is not accessible to the media and the general public. And, no evaluations are conducted to inform on whether and to what extent information from school inspectorates is appropriately used and its aims achieved. P.6	Med/High	FI

* Rig/Rel=Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

F: Quality of the feedback and communication/focus on priorities				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Chen (2011)	Indonesia	Even though the visits are frequent, it seems that the key aspects of quality assurance are missing from these visits. For example, monitoring instruction inside classrooms is not commonly covered during these visits, and reviewing the school budget is also not a routine task of the supervisors. This leads to the question of what the feedback given to principals and teachers is mostly based on, or whether the evaluations and feedback, together with other inspection areas (i.e. school facility and administrative procedures) are superficial. P.13	Med/Med	FI
Churches and McBride (2013)	Developing countries	Whenever possible, consultation with stakeholders should take place, particularly discussions with principals, proprietors, employers and higher education providers. The involvement of stakeholders in the design phase produces readier acceptance of review and a better framework, requiring fewer subsequent adjustments. For example, the South African National Education Evaluation and Development Unit framework concentrates on literacy and numeracy because they are the immediate priorities. Various aspects of what it can mean to teach well also receive different degrees of emphasis, according to local priorities. Specifically, for example, there has had to be differentiation in framework design to account for subject knowledge, knowledge of how students learn and attention to the development of students' higher-order thinking skills. Similarly frameworks may need to be adjusted to highlight different aspects of school leadership,	Low/High	CL

F: Quality of the feedback and communication/focus on priorities				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		such as improvement planning or climate for learning. At the same time, it is important for a framework to acknowledge the distinctiveness of individual schools. This is true of all schools, but particularly of private schools. When considering a school's overall effectiveness it is essential to take into account what it is (and is not) trying to achieve. P.40		
De Grauwe (2001)	Botswana, Namibia, Tanzania, Zimbabwe	The school's need for more or less supervision is not taken into account in the distribution of schools to offices and supervisors. The distribution remains done on an administrative basis, the underlying assumption being that all schools without distinction need similar supervision. P.57	High/High	FI
De Grauwe (2008)	Various (Africa, Asia)	This lack of impact is the result of a complex series of factors, which can be organised around three key issues. Secondly, precisely because supervisors have many tasks and many schools but are expected to cover all schools (the number of schools supervised may play a part in their performance evaluation), they tend to spend little time in each school. Their visits lead almost unavoidably to superficial reports, which have little credibility in the eyes of teachers. P.3	Low/High	CL
Moswela (2010)	Botswana	The environment in which instructional supervision takes place in schools is rather hostile and intimidating to teachers to make any meaningful impression on the improvement of teaching standards. P.71	Med/High	FI

F: Quality of the feedback and communication/focus on priorities				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Opoku-Asare (2006)	Ghana	The Inspectorate's responsibility for monitoring and supervision of teaching and learning in the schools is very important for achieving and maintaining standards and quality at the pre-university education level. The quality or effectiveness of school inspection, however, depends upon the objectivity with which it is conducted. It is therefore imperative that the system of school inspection be efficient and of a standard that would ensure quality educational outcomes at the foundation level of Ghanaian education. P.110	Med/High	FI
		'Evidence from the teacher interviews also indicates that some inspectors are more concerned with teacher attendance, preparation lesson notes and punctuality to school than with standards in teaching and learning'. P.113		
Santiago et al. (2012)	Mexico	Focus on administrative, bureaucratic and syndicate activities takes schools' time and focus away from the improvement of outcomes. P.155	Med/Med	CL
Uwazi (2009)	Tanzania	If inspections are not done effectively, if communication and feedback is lacking, then school inspections can be reduced to a waste of useful public resources and time. The inspections do not specifically focus on vital aspects of importance for combating poor performance among students in secondary schools. For instance, not any of the school inspections has comprehensively examined problems concerning drop-outs, pedagogical performance and	Med/High	FI

F: Quality of the feedback and communication/focus on priorities				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		students' performance, and very few do mention them. The recommendations provided tend to be repeated over time in a rather routine manner. The same recommendations are repeated time after time, and even to the same school. P.1 It shows that the school inspectorate programme is not functioning properly and therefore fails to safeguard quality of instruction and its improvement by failing to prioritise the issues of poor performance of students in the inspection cycle. P.1		
		It appears that the large number of special investigations during this time was prompted by allegations of irregularities in use of resources, need to check construction of buildings and related utilities, and allegations of misconduct by teachers and/or students. We noted that there are no clear priorities on poor performing students in mathematics and science subjects in the School Inspectorate's own annual and operational planning for inspection. The issue of poor performing students is not addressed in issued guidelines to the school inspectors. According to these guidelines the school inspectors are supposed to go through 148 items at each inspection. But only sixteen of these items are referring to the issue of poor performing students. P.5		

F: Quality of the feedback and communication/focus on priorities				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
		The inspectorate should communicate its findings not only to government officials but also to head teachers of the inspected schools, school boards and to parents through the boards. The information should also be made available to the public through print media, for example booklets, and electronically, for example through the MOEVT [Ministry of Education and Vocational Training] web page. The audit however finds that although the inspection reports are distributed to Government representatives as well as local authorities and stakeholders, the way it is done does not promote efficient and effective communication and use of the results of the inspection. For example: There is no routine of sending the school inspection's reports to the administrative district level. P.6		
World Bank (2010)	Peru	Current practice in Peru is for these procedural norms to be based on vague theories and bureaucratic needs, rather than on school-level practice and need. P.130	Low/Med	FI

*Rig/Rel = Rigour/Relevance; FI = Finding reported in study; CL = Claim by author(s) inferred from their own findings (i.e., discussion of findings) or others' findings or reports

G: Alignment and co-ordination across the system (e.g. internal/external evaluations, co-ordination with other services/unions)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Brock (2009)	Four counties in Gansu, China: Dongxiang, Jishishan, Hezheng and Kangle	By giving power to inspectors (to report, to propose changes, to propose support) the process of SDP was also enhanced and given prominence as an important process that set out school goals which could be measured by inspectors. P.457	Med/Med	FI
Chen (2011)	Indonesia	Key aspects of quality assurance and consequences are missing from inspection visits. The lack of rewards and sanctions for good or bad performance leaves the system weak. P.13	Med/Med	FI
Crouch and Winkler (2008)	Uganda	The lack of an effective inspection system at the district level, which is partly the result of the inability of the newly created Education Standard Agency to enforce inspection standards on the districts and partly the result of district governments failing to recognise their role in ensuring quality, creates weak incentives for teachers and headmasters to even be present at school, much less to perform effectively. Weak inspection is not compensated for by giving PTAs and school management committees the information, capacity, and authority to take action to reduce absenteeism and improve performance. P.15	Low/High	FI

G: Alignment and co-ordination across the system (e.g. internal/external evaluations, co-ordination with other services/unions)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
De Grauwe (2001)	Botswana, Zimbabwe	Relationships between supervision and the other pedagogical services are close and institutionalised in Botswana, supervisors being members of committees and panels in charge of curriculum development, teacher training and examinations. In Zimbabwe, their involvement is less institutionalised but still intense: supervisors serve as resource persons in training and participate in writing test items, marking examinations and preparing evaluation reports. P.44	High/High	FI
De Grauwe (2007)	Africa	Several strategies have as an objective to make external supervision more effective, by systematising the follow-up or formulating a more coherent and therefore less demanding job description. Supervision visits seldom lead to a well-organised follow-up, by the supervisors themselves, by the administration or by the schools. This lack of follow-up, the result of the powerlessness of supervisors and of the complexity of decision-making in a bureaucracy such as the educational administration, is frustrating to teachers and discredits the supervision system. It is also a core reason for supervision's feeble impact on quality. P.711	Low/High	CL
De Grauwe (2008)	Various (Africa, Asia)	The lack of impact is directly related to the lack of attention given to the follow-up to supervision. Evidently, when reports are short and superficial or simply shelved without being distributed, it is hardly surprising that they lead to little follow-up. P.5	Low/High	CL

G: Alignment and co-ordination across the system (e.g. internal/external evaluations, co-ordination with other services/unions)				
Studies	**Countries**	**Evidence (quotations)**	**Quality of study (Rig/Rel)***	**Type of evidence***
Herselman and Hay (2002)	South Africa (Eastern Cape)	Whole school evaluation will play a key role in seeking to identify pockets of excellence within the system, which will serve as models of good practices; and improve the general understanding of what factors create effective schools. P.241	Low/Med	CL
Jaffer (2010)	Pakistan	As for the supervisors' power and authority in the school hierarchy, they can only report good work, and recommend transfer, rewards or penalties, rather than make decisions on these matters. The problem, as the respondents indicated, was that others rarely followed up on the supervisor's recommendations. As one respondent put it, ''the higher authorities just write 'seen''' on the supervisor's recommendations ''without taking any action. There is no decision on the actions that we have suggested for school improvement. They ignore our note. And so we know that nothing will come out of these reports and efforts''. P.386	Low/High	FI
World Bank (2010)	Peru	Accountability pressure built around standards will lead to improved results only if one can assume that all actors have all the information and skills needed to come up to standard. P.132	Low/Med	CL

APPENDIX 7.1 CHALLENGES IN CARRYING OUT THIS SYSTEMATIC REVIEW

A7.1.1 CHANGES IN THE REVIEW PROCESS (SUMMARY)

The literature on realist synthesis emphasises that the search for evidence occurs in an iterative manner, moving back and forth from the initial rough theory to the sources of evidence with the aim of reaching 'theoretical saturation', that is, through the constant comparison of what the evidence illuminates with various aspects of the initial theory until search efforts do not yield any new information (Pawson, et al., 2004, p. 19). However, our review topic covers a complex set of accountability elements and our initial theory was more of a broad outline than a constrained set of testable propositions. Moreover, the literature we identified did not offer consistent coverage within regions or comparable coverage across regions. Therefore, we did not carry out a comparative analysis across regions in the synthesis. We also anticipated that we might have an opportunity to reveal the mutual dependence and interaction of multiple accountability elements operating together within the same country. The literature we reviewed did not provide adequate grounding for within-country analyses of interaction among policy initiatives related to accountability because: the high-quality studies concentrated on highly-specified programme activity; policy reviews provided overviews but offered few insights into school-level implications; and case studies generally provided adequate school-level detail but little detail about interaction of multiple policy initiatives.

The absence of a well-elaborated theoretical framework and empirical evidence to inform our synthesis of the literature shifted the balance of our analysis further towards theory building. Even with full-text screening, we were still not certain what discrete elements of particular papers might yield in terms of enriching our theoretical understanding. For that reason, we employed an approach that was more closely aligned with that pursued in other configurative approaches, which involved scanning the terrain in the search for papers that might help us develop a robust theoretical framework.

Our approach to theory building began with our use of our initial rough theory to inform our initial coding scheme of the literature. Throughout the coding, we aligned evidence relating to outcomes for education and accountability to our initial rough theory, adding codes for mechanisms and context that were not included in our initial coding scheme. Also, when we conducted the syntheses, especially of Assessment and Monitoring literature, we identified particular pathways to impact for each element that corresponded with categories of initiatives and their intended outcomes. Pathways to impact are specific programmatic approaches to accountability which are founded on assumptions about how desired outcomes occur (e.g., implementation of EMIS for local management decisions; high-stakes vs low-stakes assessments). Identified pathways to impact allowed us to trace the ways that different initiatives were intended to produce certain outcomes so that we could more clearly articulate between intended mechanisms and the actual configurations of context-mechanisms-outcomes described or reported in the papers. These approaches not only allowed us to test our initial rough theory, but most important, they enabled us to build theory by comparing the configurations of context-mechanisms-outcomes derived from our syntheses of papers with our initial rough theory. We could then build theory by identifying gaps and elaborating our initial rough theory.

A7.1.2 SEARCHING

The development of search terms was a challenge due to the broad nature of the accountability elements and the fact that different terms may be used for the elements of accountability, inspection, monitoring and assessment in developed versus developing countries.

Furthermore, few studies on accountability systems in developing countries are published and accessible in traditional databases. We asked our Advisory Group for relevant papers and authors and specifically analysed reference lists of relevant papers to add studies manually that had not appeared in our search of databases and websites. We carefully searched on relevant LMICs websites and specialist databases such as African Journals online and Bangladeshi Journals online. This was time-consuming as the search functionality in these sources was limited. For example, we could use only key search terms (e.g. school accountability) without other combinations of terms, resulting in hundreds of titles identified. In many cases, relevant titles had to be manually entered to EPPI-Reviewer 4. However, to ensure transparency, we documented all manual searches of websites and citation searches on a spreadsheet, entering details such as key search terms used, relevant literature found, duration of search and information on systematic reviews found.

A7.1.3 SCREENING

A more direct outcome of the screening of literature was that we realised that members of the team differed in their interpretation of the definitions of the three accountability elements. In particular, our discussions on the papers highlighted how terminology to describe monitoring, assessment and inspection differed when looking at accountability as practised in LMICs. Moreover, our particular concern in this review was to concentrate on the ways that accountability activity influences school-level procedures and outcomes. The emphasis on the school-level required consensus about what that meant in terms of a wide range of different accountability activities.

In terms of monitoring, many studies address system-level monitoring, analysing how, for example, international surveys (e.g. PISA, SAQMEC) can be used to monitor the performance of the education system as a whole or at the provincial level. As we are interested in school-level monitoring and how monitoring systems impact on the school level, we decided to exclude these studies in the first round of data extraction. Some of these studies, however, highlighted that system-level monitoring might also lead to changes in the accountability and governance of schools, and we marked these studies as potentially relevant for a second round of data extraction in case the first round did not provide sufficient detail of the conditions and mechanisms of change.

For example, Nzomo et al. (2001) draws out the implications for school policy of data from a regional assessment in Southern Africa, the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ). SACMEQ collects a wide range of data on a selective sample of schools within its member countries, ranging from baseline input data on pupil, teacher, school and community characteristics to standardised assessments of reading performance. The authors of the Nzomo et al. (2001) report using analyses of SACMEQ data to highlight discrepancies in provincial resource allocation and reading achievement that were intended to have policy implications for the equitable allocation of resources, which would have direct implications for the three outcomes of interest for this review, notably student learning, school effectiveness and system efficiency. This article was deemed 'somewhat relevant', indicating that we might need to return to it at a later time to parse its recommendations and associated data for the contributions each made towards understanding conditions and mechanisms, that is, what is being monitored, under what conditions, by whom, how and to what effect.

Our discussions during the searching and screening phases led us to tighten our definition of monitoring, emphasising the infrastructure that supports collection, analysis and reporting of quantitative/empirical results aimed at evaluation of school-level performance against benchmark targets using agreed indicators of educational quality.

In relation to inspection, the screening process highlighted that the term 'inspector' or 'inspection' has a different meaning in LMICs and that often other terminology is used to describe the type of inspection practices we are interested in. Where high-income countries would refer to inspections as evaluations by government officials external to the school using a standardised external framework of standards, inspectors and inspections in developing countries often also refer to head teachers being required by national legislation to evaluate the performance of their teachers, or to district officials managing and evaluating a number of schools. The screening also provided proof of what we had already suspected, that the term 'supervision' is often used in LMICs when referring to inspections. These reflections led us to emphasise our focus on the actions related to inspection/supervision that have at their core an element of judgement, using an external framework, and producing results that allow for some level of comparison between schools.

Assessment appeared to be the most straightforward and clear accountability element, although our screening indicated that characterising different types of assessment was highly problematic. We distinguished between high- and low-stakes assessments, but even this broad definition was problematic in that some types of national assessment that began as low-stakes evolved to take on more aspects of high-stakes assessment as they gained legitimacy (e.g., Santiago et al., 2012). As we are interested in assessment as an accountability element, we re-emphasised our focus on outcomes from both low- and high-stakes standardised assessments that provide some indication of school-level performance.

LIST OF ABBREVIATIONS

ADEA	Association for the Development of Education in Africa
AGEPA	Amélioration de la Gestion de l'Education dans les Pays Africains
CA	Continuous assessment
CASS	Continuous assessment
CBCs	Common basic contents
C-M-O	Context-mechanisms-outcomes
DfID	UK Department for International Development
EFA	Education for All
EMIS	Educational Management Information Systems
ENEM	Exame National do Ensino Médio, or National High School Exam
EPPI-Centre	Evidence for Policy and Practice Information Centre
IIEP	International Institute for Educational Planning
KSQAO	Karnataka School Quality Assessment Organisation
LMICs	Low and middle-income countries
NGO	Non-governmental organisation
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PTA	Parent Teacher Association
RCT	Randomised controlled trial
SACMEQ	Southern Africa Consortium for Monitoring Educational Quality
SBA	School-based assessment
SDP	School development planning
SIMECAL	System for Measuring and Evaluating the Quality of Education
TIMSS	Trends in International Mathematics and Science Study
USAID	United States Agency for International Development