NEPC

NATIONAL EDUCATION POLICY CENTER

School Self-Evaluation and Inspection for Improving U.S. Schools?

Katherine E. Ryan, Tysza Gandha, and Jeehae Ahn

University of Illinois at Urbana-Champaign

November 2013

National Education Policy Center

School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
http://nepc.colorado.edu

Briefs published by the National Education Policy Center (NEPC) are blind peer-reviewed by members of the Editorial Review Board. Visit http://nepc.colorado.edu to find all of these briefs. For information on the editorial board and its members, visit: http://nepc.colorado.edu/editorial-board.

Publishing Director: **Alex Molnar**

Suggested Citation:
Ryan, K.E., Gandha, T., & Ahn, J. (2013). *School Self-evaluation and Inspection for Improving U.S. Schools?* Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/publication/school-self-evaluation.

School Self-evaluation and Inspection for Improving U.S. Schools?

Katherine E. Ryan, Tysza Gandha, and Jeehae Ahn,
University of Illinois at Urbana-Champaign

Executive Summary

No Child Left Behind (NCLB) implementation and results have stimulated vigorous debates about the benefits and harm of accountability policies. Despite robust criticisms, some form of accountability for publicly-funded education to safeguard school and teaching quality and equitable treatment of students is important for serving the public interest. The Race to the Top initiative and pending re-authorization of the Elementary and Secondary Education Act (ESEA) raise questions about alternative models of accountability that might be more effective in generating these valued outcomes.

The U.S. test-based accountability model holds schools and teachers accountable for student outcomes with little attention to school improvement processes. Many other countries enact more school-centered accountability efforts, such as school self-evaluation followed by inspection (SSE/I) to examine school quality, as in similar systems in England, Wales, Northern Ireland, and the Netherlands. SSE/I is a complex policy instrument with mixed consequences and many research questions still to be answered. Moreover, accountability models from other countries cannot be naively imported to the U.S. given the vital distinctions in sociopolitical contexts. That being said, a look at some of the purposes or principles behind SSE/I—especially its emphasis on quality improvement—can nevertheless inform efforts to redesign and improve the U.S. accountability model. The purpose of this brief is to take just such a look at this model.

Complexities of adapting an SSE/I model include questions about: appropriate purpose and scope; credentials of external inspectors; conceptions of and criteria for school quality; engagement of all stakeholders in design efforts; and adequate funding.

In relation to these concerns, we recommend that policymakers consider the following as starting points for incorporating SSE/I notions in the U.S. context:

1. Instead of sanctions-based *inspections* for making summative judgments about schools, employ ***external reviews* that focus on providing guidance and support** for school development and improvement. Proportionate reviews targeting struggling schools might provide external guidance and support where they are needed most, and they might also allow for the most efficient resource allocation.

2. Employ as external reviewers **qualified experts who meet a prescribed standard for qualifications and required experience**. Robust training should initially be compulsory and then required on an on-going basis.

3. Incorporate a **broad set of school quality criteria that goes beyond standardized test scores** to adequately represent other valued aspects of quality teaching and learning. In particular, we recommend teaching quality be defined in relationship to a specific set of criteria.

4. Include the perspectives of **multiple stakeholders** (administrators, teachers, students, parents, community leaders, and researchers) in design efforts, allowing all key stakeholder groups to participate directly.

5. **Recalibrate and revitalize existing state resources** (e.g., regional offices of education) to support external review initiatives.

School Self-evaluation and Inspection for Improving U.S. Schools?

Introduction

No Child Left Behind (NCLB) test-based accountability relies primarily on *performance measurement* (state high-stakes testing based on top-down educational standards) and *sanctions* to stimulate school improvement.[1] The NCLB 10-year experiment has resulted in small improvements in test scores, and even those may be inflated.[2] The evidence suggests that overall, U.S. students have not attained mastery in a variety of subject areas (in reading, for example), and the educational achievement gaps the law supposed to remedy persist across racial and socioeconomic groups.[3]

These achievement issues are also visible in comparisons of student achievement on international assessments such as Program for International Student Assessment (PISA). Although 2011 results from Trends in International Mathematics and Science Study (TIMSS) are more promising,[4] students in the U.S. perform at the average on reading and science and below average in math when compared to students from other countries.[5] Of course, it is well known that the performance of students in the least well-supported schools may lower the average of students in well-funded schools—whose performance is often comparable with that of students from the highest ranked countries; still, the overall average performance of American students is widely perceived as unacceptably low.

In addition to NCLB's failure to produce the promised substantive gains in achievement (as measured on standardized tests), the law has been criticized for: failing to provide information useful in improving teaching and learning; narrowing curriculum; and, disproportionately affecting the experience of students, teachers and schools in low-income areas.[6] Race to the Top (RTTT) and NCLB waivers represent attempts to address some criticisms, but the law continues to rely heavily on testing and sanctions to drive educational reform. For example, RTTT requires teacher evaluation systems that weigh students' test performance gains with other information to make judgments about teaching quality.[7]

Despite known problems in the current system, however, some form of accountability in the U.S. national educational system will likely remain.[8] As long as education is publicly-funded and mandatory, citizens in this democratic society have a right to expect that schools will be held accountable for effectively serving public interests, including ensuring that they use public funds efficiently and appropriately.[9] In addition, while no educational accountability policy instrument is robust enough to adequately address the opportunity gaps experienced by particular groups of students (e.g., low-income), efforts to safeguard equitable treatment of students—in particular, ensuring the quality of their education and

that their schools are staffed by qualified professionals—serve a critical public interest.[10] There is also growing interest in ensuring that schools are aligned with regional or national policies and standards. Thus, a question of interest to policymakers, practitioners, and researchers alike is: What model of accountability might be more effective than NCLB and other U.S. initiatives in contributing to improved teaching and learning?

Certainly many other national education systems employ some form of school accountability; in contrast to the U.S., other countries' accountability models typically incorporate some kind of quality improvement system. One popular model thought to be useful in enabling school improvement is school self-evaluation and/or inspection (SSE/I).

An SSE/I framework acknowledges the multi-dimensional nature of school quality.

While the empirical relationship between SSE/I and school improvement is not well-established, according to Education at a Glance 2011, 29 countries require some form of school self-evaluation, inspection, or a combination of both as part of their educational accountability systems.[11]

On the surface, SSE/I offers some attractive features when compared with test-based accountability. These include an emphasis on holding schools responsible for internal school processes and practices presumably under their control and leveraging more resources to enable schools to build capacity to improve student learning.[12] Such systems incorporate multiple measures (school indicators like graduation rates, test scores and direct observations of classroom teaching, for example) as well as perspectives of various stakeholders (parents and students, for example). Thus, an SSE/I framework acknowledges the multi-dimensional nature of school quality and addresses the many criticisms of NCLB test-based accountability that relies heavily on large-scale assessments.[13] In counterpoint, some of the key criticisms of SSE/I include substantial financial and time costs and potential unintended consequences of high-pressure school inspections (such as stress on teachers and school leaders, time devoted to preparing for inspection that otherwise would have been devoted to core instructional goals and practices).[14] There is also some concern about quality in the inspection process, including concerns about whether all inspector reports will be reliable, useful and fair.[15] The extent to which SSE/I can be trusted to respect educators' professional judgment and autonomy is also controversial.[16]

Even if such concerns were not in evidence, SSE/I from other countries could not be simply transferred to the U.S. because of crucial differences in sociopolitical contexts, including enormous differences in the public perception of teachers, in institutional arrangements, and in governance systems. All of that being said, however, a look at some of the purposes or principles behind SSE/I—especially its emphasis on quality improvement—can nevertheless inform efforts to redesign and improve the U.S. accountability model. The purpose of this brief is to take just such a look at this model.

We begin with an overview of SSE/I that includes definitions and a broad description of common implementation patterns. We then elaborate on the UK-style SSE/I approach (as seen in England, Wales, Northern Ireland, and the Netherlands[17]), which has received substantial attention here in the U.S.[18] In describing the theory of change underpinning the UK model, we draw on academic literature, policy documents, and other sources, adding information from in-situ observations and interviews conducted with educators in England. We also present the limited empirical evidence about the impact of this approach on educational outcomes available from studies conducted in England and the Netherlands. We conclude with questions and recommendations for policymakers to consider in their further efforts to improve U.S. educational accountability.

Overview of School Self-evaluation and Inspection Policies and Practices

While educational accountability requirements and the machinery of implementing accountability vary across nations, to date two approaches dominate the literature. Most can be classified as either high stakes test-based educational accountability (TBA) models (common in the U.S. and Latin America) or school self-evaluation and/or inspection (SSE/I) models (common in Europe and the Pacific Rim). Broadly speaking, the Organisation for Economic Co-operation and Development (OECD) describes school self-evaluation (SSE) as a systematic review of and reflection on "the quality of the instruction and education services provided and…school outcomes"[19]; as the name suggests, personnel in a school complete the evaluation. School inspection (I) is defined as "a mandated, formal process of external evaluation with the aim of holding schools accountable."[20]

Actual practices of SSE/I vary considerably among countries in relation to a variety of characteristics, including, for example: a focus on SSE, or I, or both; the use of reports; and the criteria/standards employed. Seven of 29 OECD countries require only school inspections as part of their educational accountability system. Four countries rely on only SSE, and 18 countries report utilizing both SSE and I.[21] Essentially, SSE/I approaches differ in how they balance tensions between "judgments about quality" (as when an inspector makes summative judgments about quality) and "development of quality" (as when an external reviewer makes recommendations for improvement). In England, for example, external inspectors make summative judgments, and final reports appear online as public information. In contrast, external inspections in Hong Kong are flexible and intended to be responsive to each school's individual development needs and interests; reports are meant mostly to support school's continuous improvement and are not publicly distributed.[22] Further, while inspections or external reviews have usually involved all schools, lower-achieving schools are now becoming either the primary or sole focus of external reviews or inspections in approximately nine countries.

Historically, SSE/I processes are grounded in both school effectiveness and school improvement theories. School effectiveness research focuses on the characteristics of successful schools associated with improved student performance on standardized tests

(teacher quality, for example).[23] School improvement theories promote improved student learning through improvements in teaching and learning and conditions, (the quality of school leadership, for example).[24] More recently, organizational learning—when defined as the capacity to develop, acquire, use, and reflect on knowledge to change organizational practices, routines, and outcomes—has strengthened and extended the SSE/I theoretical base.[25]

In part because of this emphasis on informing improvement, the idea of SSE/I has captured some attention in the U.S. Scholars, educational leaders, policymakers and others from a variety of settings have proposed some kind of inspection system modeled after those in England, the Netherlands, and other nations as a replacement or supplement to U.S. educational accountability.[26] For example, while noting that a variety of initiatives (like investment in early childhood education) are critical for addressing the nation's inequality, a report by Broader Bolder Approach to Education (BBA), an offspring of the Economic Policy Institute (EPI), urged this kind of SSE/I initiative also be considered. The report was signed and endorsed by numerous leading educators and social and behavioral scientists.[27] More recently, an Education Sector report proposed that the U.S. think about implementing inspections based on the English SSE/I model.[28]

Other stakeholders have, not surprisingly, voiced concerns. For example, a National Education Policy Center review soundly criticized the Education Sector report for its advocacy orientation in addition to expressing concerns about costs to states.[29] The report noted that further analysis and research would be needed to assess whether some form of SSE/I would be viable in the U.S. context.

We do not pretend to provide definitive answers to questions being raised about a potential SSE/I innovation in the U.S. Instead, we seek only to provide a look at those elements of the UK-style system that may prove useful in informing U.S. moves toward reform. To that end, in the next section we offer detail on that inspection system, which employs targeted inspections based on a school's previous inspection outcomes and risk assessment (for example, findings related to students' academic achievement over time, student safety, and teaching quality). Some kind of "proportional inspection" that concentrates on schools experiencing chronic issues with similar school problems, including safety and fragmented curriculum, might have promise for the U.S. system.

UK-style School Self-evaluation and Inspection

The UK-style SSE/I model, where external inspection follows self-evaluation (as practiced in England, Wales, Northern Ireland, and the Netherlands), is intended both to provide accountability and to foster quality improvement. Each country conducts the processes differently. For example, the agency responsible for inspection varies, as does frequency of inspections. And, while inspection reports are made publicly available across the UK, the extent and type of sanctions given for low-performing schools also differ somewhat. For example, only England has such sanctions as school closure or takeover for chronic low

performance. The Netherlands, which instituted a SSE/I system based in part on the UK system, is similar although there are important differences in the context of these countries. For example, the school inspectorate in the Netherlands is part of the Education Ministry, not an independent educational office.[30]

In spite of the distinctions, the core of what we call the UK-style SSE/I system are school self-evaluation and proportionate, risk-based external inspections. School self-evaluation, which all schools are expected to conduct, is seen as integral to the quality improvement process as well as the key basis for external inspections. The inspections are intended to evaluate the quality of self-evaluation, including ensuring that the self-evaluation processes are implemented properly and effectively.

Unlike the NCLB incentive mechanism, SSE/I expands quality improvement to incorporate processes and practices that emphasize diagnosis of barriers to improvement, such as unsafe school environment and lack of leadership, and of strategies aimed at improving educational outcomes (e.g., implementing a coherent curriculum, improved teaching). That is, schools are evaluated on how well they use their resources (human, student data, and financial resources, for example) to meet the needs of their particular students. SSE/I models assume that schools' instructional and organizational capacities must increase before learning outcomes can improve.[31]

The following segments provide detail on the core components.

School Self-Evaluation

All schools are expected to conduct their own self-evaluations and to prepare for external inspections in accordance with some common evaluation frameworks. School self-evaluation is intended to provide school personnel with an understanding of the school's overall quality and key priorities for development, and thus, serve as a catalyst for ongoing self-reflection, learning and improvement. School self-evaluation is also used as the starting point for external inspection. That is, the inspection team assesses to what extent and in what ways they agree or disagree with the school's self-evaluation. The inspection team's recommendations and feedback are intended to help school personnel be sure they have appropriately diagnosed areas for improvement and take necessary actions.[32]

Inspection Framework

In England, the Office for Standards in Education, Children's Services and Skills (Ofsted) has devised and published the framework guiding school self-evaluation. Independent from the UK government education department, Ofsted is responsible for inspecting all schools in England that are primarily state-funded. Currently, the four main areas of focus in the Ofsted framework 2012 include:

- the achievement of pupils;
- the quality of teaching;
- the behavior and safety of pupils;
- and, the quality of school leadership and management.

Inspections also address the overall development of pupils, including their social development, and the extent to which the school is meeting the needs of its diverse student populations.[33]

When scrutinizing the achievement of pupils, inspectors consider a longitudinal view of large-scale test scores over the last three years for all students and for different student groups (students with special needs, for example). Likewise, in assessing teaching quality, inspectors use multiple criteria, including whether teachers demonstrate high levels of expertise and subject knowledge, and whether they use methodologies that facilitate learning and the overall pupil development; in addition, inspectors determine the extent to which teachers meet national "Teachers' Standards." These institutional arrangements are similar in other UK countries (except Scotland). The Dutch Inspectorate of Education has a similar inspection framework to standardize inspections and judgments of school quality.[34]

Inspection Evidence and Process

Schools are inspected every few years; after each inspection, schools receive an overall quality rating and oral and written feedback. In England, for instance, schools are inspected every 3-5 years, over a two-day period, on short notice (48 hours or less). In the Netherlands, schools are inspected at least once every four years, and inspection visits may be announced or unannounced. Low-performing schools are inspected more frequently in order to closely monitor their progress. Inspectors typically observe lessons, and they consider such areas as whether the lesson presented is challenging enough for all pupils, whether it is differentiated to meet individual needs, and whether pupils' responses demonstrate meaningful learning. They also scrutinize whether teachers use appropriate methods (such as questioning, discussion, and assessment) to promote learning. Assessments of student behavior include observations in school common areas as well as in classrooms and also on a review of student behavior records.

Inspections also include discussions with pupils, parents, and staff, and case studies of individual pupils (low-achieving individuals, for example).[35] Inspectors also examine the school's own records and evaluations of the quality of teaching and learning. As appropriate, they may also provide oral feedback to teachers for further improvement. Students' performance on standardized written exams—which are not multiple-choice in the UK—are critical, but not the only data used to judge school quality.

The range of evidence gathered during inspections along with test scores provide support for judgments made. The inspection report presents an overall assessment of the school's effectiveness based on the four key inspection areas (pupil achievement, teaching quality, pupil behavior and safety, and leadership and management quality) using the following 4-point scale: *1 Outstanding*, *2 Good*, *3 Requires Improvement*, and *4 Inadequate*. Following the receipt of the inspection report issued within ten working days of the inspection, schools are expected to develop, revise, or both, their improvement plans and address the problem areas that have been identified during inspection. A school graded *Requires Improvement* or *Inadequate* will be monitored and a full return inspection will be conducted within 18-24 months. In particular, schools graded *inadequate* may either be issued a "Notice to Improve" or placed in "Special Measures." And a school placed under "Special Measures" will receive its first monitoring inspection within three months of the initial inspection. Thus, school inspections are proportionate to the needs and character of each school.

In summary, SSE/I focuses on developing organizational structures (internal processes and practices) that steer how schools organize themselves to improve student learning. At the same time, schools are held accountable to important education stakeholders (government, students, parents, and the community at large).[36] Unlike test-based accountability models that provide little guidance and support for school improvement,[37] self-evaluation and external inspection processes are designed to facilitate schools' improvement processes as well as to provide an "external or visitor's eye" for monitoring school quality in key areas.[38] Thus, SSE/I could be beneficial for low-performing schools with limited instructional and organizational capacities.

What is the Impact of SSE/I?

Whether UK-style SSE/I is a "tool for supervision" or a "tool for school improvement and equity" is a hotly debated and much discussed topic in policy and academic circles and among the public at large. Self-evaluation does require schools to collect substantial data and engage in robust performance management employing data-based decision-making—but there is less attention to a school's ability to actually engage in such self-assessments.[39] And, schools often consider external inspections a means of central government control, rather than "a critical friend" that can meaningfully support and help them improve.[40] Yet, a recent evaluation study of the Ofsted inspection found that the inspection process, where it was thorough, rigorous, and consistent, was generally perceived positively among teachers.[41] [42]

Although SSE/I is likely useful for quality assurance,[43] empirical support for the processes is modest to date. Much of the available English-language literature is about either English (Ofsted) school inspections or the Netherlands approach. This brief draws on research from both countries to increase confidence that findings primarily based on Ofsted school inspections are not necessarily unique to the English setting. We conducted a selective literature review, focusing primarily on literature about English inspections since 2005,

when Ofsted implemented a "new" inspection system characterized by shorter inspection notice and fewer but clearer recommendations.[44] A few pre-2005 studies are also included to explore the importance of new inspection characteristics as compared with previous types of SSE/I. In addition, we reviewed Dutch research conducted since the first Supervision Act was implemented in the Netherlands in 2003, outlining standardized frameworks, procedures, and protocols for implementing SSE/I.[45] Our selective review process prioritized larger-scale, longer-term, or multi-method studies to gather more reliable and generalizable evidence about SSE/I.

The findings from our review are synthesized in three sections below. First, we summarize findings about educators' actual experiences with self-evaluation and external inspection. Second, we report on the influence of SSE/I on internal school processes and practices. Finally, we describe the impact of SSE/I on student outcomes. U.S. policymakers might be most drawn to evidence about SSE/I's impact on student learning as they evaluate the model's potential for improving student outcomes. However, to determine the relationship of SSE/I to school quality and the particular mechanisms by which it might or might not contribute to desired outcomes, it is also important to examine how educators experience and perceive the process as well as its impact on school processes.

Educators' Experiences with SSE/I

School self-inspection theoretically allows schools to exercise greater professional autonomy; similarly, inspections are intended to be supportive. Nevertheless, SSE/I has been critiqued for driving a culture of external control and school compliance—shifting the power of educational evaluation from teaching professionals to visiting inspectors.[46] In addition, some researchers have raised concern about the potential stress that school personnel may experience as they anticipate and prepare for inspections.[47] Others have theorized that excessive stress and pressure may generate negative side effects, including and window dressing behaviors like intentionally leaving deep-rooted problems out of an SSE report.[48] To reduce cost, unintended negative consequences, and time demands of the process, in 2005 Ofsted introduced changes as we noted above, including shorter inspection notice and clearer recommendations.

Despite criticisms, several large-scale studies have consistently documented that educators generally have positive experiences with Ofsted's SSE/I process, even pre-2005. Around 90% of the 2800 principals and 900 teachers surveyed in 2003 reported satisfaction with Ofsted inspections. Approximately 80% of teachers considered the demand for documentation and information to be reasonable and 80% of principals indicated that benefits of inspection outweighed the costs.[49] A 2006 independent evaluation of Ofsted inspections, using survey and interview data from a random sampling of schools stratified by school level, geographic region and inspection results, also showed that around 90% out of 134 school leaders were satisfied with the inspection process; that is, they found inspectors' judgments accurate and fair and the reports helpful.[50] Two-thirds of survey respondents agreed that the new inspection was less stressful than before, validating

Ofsted's improvement efforts. A follow-up study was conducted three years after 2006 inspections that involved 96 interviews with school leaders, teachers, and teaching support staff and survey data from an additional 126 school leaders. Educators widely described ongoing efforts at their school to update their self-evaluation form even without further monitoring by inspectors. Furthermore, educators generally perceived self-evaluation as a valuable process for identifying school strengths and weaknesses, even if it was time-consuming.[51]

Still, experiences were not universally positive, revealing some complexity in the use of SSE/I to improve schools.[52] McCrone et al.'s survey analysis (2006) showed a positive and statistically significant relationship (exact statistics not reported) between school leaders' perceptions of feedback quality—fairness, completeness, and constructiveness—and their satisfaction.[53] Fewer than a quarter of the 126 respondents in that survey (exact statistics were not reported) reported negative experiences, and qualitative case studies included in the study provided converging evidence. Teachers and teaching support staff who described their observation feedback as fair and extensive also expressed high overall satisfaction; those who described concern or disappointment with observation feedback expressed low overall satisfaction. Plowright (2007) theorized that school culture might influence teachers' perceptions of SSE/I.[54] That is, schools with a learning organization culture (for example, a school routinely using collaborative efforts to solve problems) might view SSE/I as a process for school development, not just for accountability. Ehren and Visscher conducted ten in-depth case studies in Dutch elementary schools to examine potential moderating effects of school and inspection characteristics.[55] Their qualitative analysis found that inspection outcomes are influenced by a complex interaction between inspector style, perception of recommendations, and school capacity and readiness for change. Thus, while there is evidence that educators widely have positive experience with the SSE/I process, implementation quality may be an important factor in building an effective system.

SSE/I's Consequences on Internal School Processes and Practices

SSE/I theory claims that improving school policies and practices is associated with improving student learning. Empirical studies dating back to the start of Ofsted inspections in the early 1990s consistently found that SSE/I contributed positively to improvements in such internal school processes and practices as instruction and assessment.[56] In a 2009 study conducted three years post-inspection,[57] over half of school leaders surveyed (n=126) reported that inspection recommendations from three years ago had continued to help their school in setting priorities and in directing their focus in school development. Around half reported that inspection had a positive impact on the quality of teaching at their school. One-quarter to one-third reported a positive impact on assessment and progress monitoring practices at their school, which in turn, improved teachers' understanding of their students.

Earlier studies similarly found that a majority of principals and about half of teachers reported making at least some of the changes recommended by external inspection.[58] Responses including doing more of something (core curricular work, for example) or less of something (teacher-directed work, for example).[59] Although some changes were more superficial than others, analysis of inspection results found some converging evidence that schools were making changes that improved school quality and classroom instruction. Between 1994 and 2003, two-thirds of schools were judged to have improved and the proportion of observed lessons judged as unsatisfactory had decreased from one in five to one in twenty.[60]

Although the consequences as described above are positive, empirical studies consistently found that SSE/I's impact on school improvement is by no means transformative. According to a large-scale evaluation conducted by the National Foundation for Educational Research in England (an independent organization), a majority of educators reported that inspections generally did not highlight new areas of action for schools. On the other hand, SSE/I did contribute to supporting and improving processes and practices by validating school self-evaluation findings, helping the school to prioritize their actions, and leveraging support (e.g., funding) from local authorities.[61]

SSE/I's limitations are due in part to issues with the quality of inspections and local capacity for school improvement. In England and the Netherlands, studies found, in some cases, that feedback from inspectors was too superficial or vague to have an impact on teaching and learning, even after Ofsted made explicit efforts to improve the quality of inspectors' feedback.[62] However, Ofsted acknowledges its impact may be limited. A Children, School and Families Committee Report says, "Ofsted has a duty to encourage improvement in schools…not necessarily an active role to play in school improvement…Ofsted has neither the time nor resources…."[63] The responsibility for school improvement still rests primarily on schools and local governing bodies (equivalent of districts in U.S.), which may lack the resources, knowledge, and enabling factors to implement inspection recommendations.[64]

SSE/I's Impact on Student Outcomes

It is somewhat more difficult to measure the ultimate impact of SSE/I on student outcomes than it is to assess perceptions of educators and changes in school processes. Longitudinal trends in performance on the English national age-11 (KS2) tests and age-16 (GCSE) exams had shown improvement between 1997 and 2007. The percentage of age-11 students who achieved level 4 or higher in English, math, and science had increased by 15 to 19%, while the percentage of age-16 students with A* to C grades in 5 GCSE exams had increased from 45% to 61%.[65] Not surprisingly, perhaps, given the difficulty of untangling SSE/I from other influences, results in the literature are mixed.

Some studies in the Netherlands and UK have found few effects on student achievement in the overall population.[66] Other studies that used statistical methods to control for other

influences and to better estimate the effect of SSE/I on student performance were more promising.[67] Most recently, Hussain analyzed performance trends of low-performing schools (in "Notice to Improve" or "Special Measures" categories) inspected between 2006 and 2009.[68] He compared performance improvement for schools inspected at the start of the school year (with a year to act) with the performance of otherwise-similar schools

> *If schools are being held accountable for improving teaching and student learning, policymakers should also be expected to support the capacity required to produce improved teaching and learning.*

inspected at the end of the year (with no time to act before the next test date). The effect of inspection on low-performing schools was statistically significant, accounting for 10% standard deviation of age-11 national standardized test performance in mathematics and English. Earlier studies had also documented SSE/I's statistically-significant positive effect on lower-performing schools. In their analysis of lower-performing English schools, Shaw and co-authors calculated that Ofsted inspections resulted in 2% to 4% increases in students' national exam scores.[69] Luginbuhl, Webbink, and de Wolf similarly found that inspection in the Netherlands accounted for small test score increases (2% to 3%) that persisted for two to four years for lower-performing schools.[70]

However, SSE/I's potential impact of student outcomes seems likely to be dependent at least in part upon schools' capacity for change and instructional improvement. There is modest empirical evidence in support of the relationship between the quality of internal processes and practices and improved student learning in lower-achieving schools, but SSE/I alone is unlikely to ensure improvement. In a study using a regression discontinuity design, Allen and Burgess found that two years after a failed inspection, there were small positive test score gains only for schools judged to have adequate leadership capacity.[71]

Summary

Overall, findings about SSE/I efficacy are mixed in relation to its ability to support improvement in policies and practices critical to equity as well as its potential for improving student outcomes; it is also unclear exactly how SSE/I generates valued educational outcomes. There is evidence, for example, that SSE/I contributes to improving school processes and practices by validating school goals rather than by guiding school improvement efforts in new directions. The majority of schools reported that inspections facilitated collaboration amongst staff in completing their school self-evaluation, which was widely described as a productive process for planning and monitoring improvement efforts. Further, inspection reports provided stakeholders with useful information about a school's quality in multiple areas, including not simply standardized test scores but also results of teaching observations and analyses of school safety. Although inspectors' feedback varied in quality, school personnel generally agreed that inspection results

provided helpful external validation about areas that needed improvement and served as leverage to obtain necessary resources.

That said, there is only modest evidence to date of SSE/I's relatively greater impact on lower-performing schools in England and the Netherlands. Some initial evidence suggests that SSE/I has potential for influencing student outcomes improvement in lower-performing schools—but it's not clear what mechanisms may be essential to leveraging such improvement. Inspection focuses on processes that are purported to influence student achievement, like teaching and leadership; however, clear links between improvements and specific processes have not been determined. Furthermore, construct validity critiques and other assessment issues are also relevant. Conceptualizing and developing adequate measures of complex constructs like student engagement, student achievement, and critical thinking continue to pose psychometric and practical challenges.

Redesigning Accountability in the U.S.?

Without a broader set of educational and social policy initiatives aimed at improving learning opportunities for student groups, neither test-based accountability nor school evaluation and inspection will generate substantial gains in student learning. Nonetheless, particular principles and practices underpinning SSE/I may provide useful possibilities as policymakers redesign school accountability in the U.S. SSE/I is a policy instrument designed to safeguard quality assurance standards and the efficient use of resources. Under some circumstances, whether standards of quality—school safety, for example—are being met can be determined only when an inspection or external review team visits the school.[72] In addition, the inspection report provides stakeholders a range of relevant information about school quality, including information about specific strengths and shortcomings that school personnel can use to improve the school's performance. There are, however, some important issues and challenges involved in considering how SSE/I purposes and practices might be used in a U.S. quality improvement system. Below we briefly note a few potential issues.

- **Purpose and Scope**. Given the current, sanctions-based NCLB system, it seems prudent to adopt the notion of external *review*, rather than *inspection,* in considering UK-style school inspections for the U.S. context. Development-oriented, these risk-based, proportionate external reviews would focus on providing guidance and support for improving district and school quality to improve student learning.

- **External Reviewers/Inspectors**. In England, school inspectors are certified and regulated by Ofsted, either directly as Her Majesty's Inspector (HMI) or through contractual arrangements. The qualifications and experience required and standards to meet are prescribed. Robust training is compulsory initially and required on an on-going basis. In the U.S., currently, there is no such equivalent institutional and administrative structure. A critical issue to be resolved is who

should conduct the reviews and what qualifications and training should be required.

- **Conceptualization and Assessment of School Quality**. It will be a challenge to develop accurate, reliable, fair, and instructionally useful measures of school quality for such complex factors as teaching quality. Constructs and their measurement can be defined too narrowly or broader than desired. When measurements are too narrow—as in total reliance on standardized test scores—they under-represent other valued aspects of quality teaching and learning, such as student engagement and critical thinking in content areas.

- **Process and Model Development**. One of the important features of SSE/I is the inclusion of administrators', teachers', students' and parents' perspectives in assessing school quality and developing improvement plans. It will be critical to incorporate the views of a variety of key education stakeholders, including teachers as well as administrators, on what external review or inspection might entail in the U.S.

- **Funding.** Most states are working with severely-constricted resources. With few new fiscal resources available, policymakers need to consider how existing state resources might be recalibrated and revitalized for such efforts. For example, the Illinois Regional Offices of Education (ROE) system, which already provides systematic support and services to its districts and schools, might be modified to enable an SSE/I accountability system.[73] The same is true for the 42 states that have a regional delivery system similar to that of Illinois.

With other scholars and policymakers, we concur that devising and investigating richer and more robust accountability and improvement systems are important next steps.[74] Specifically, if schools are being held accountable for improving teaching and student learning, policymakers at all levels of the educational system, regional and state levels as well as the national level, should also be expected to support the capacity required to produce improved teaching and learning.[75] Research on student outcomes associated with new quality improvement strategies will be crucial.[76]

We suggest that practitioners, policymakers, and researchers work in partnership to conduct research on improvement and accountability strategies that will most likely lead to improved student learning in the U.S. Given the important sociopolitical differences among countries, efforts to develop some version of SSE/I for U.S. schools should utilize the underlying principles and purposes, rather than any particular practices of actual systems from other countries.

Recommendations

In light of the above, we recommend that policymakers consider the following as starting points for incorporating SSE/I notions in the U.S. context:

1. Instead of sanctions-based *inspections* for making summative judgments about schools, employ ***external reviews*** **that focus on providing guidance and support** for school development and improvement. Proportionate reviews targeting struggling schools might provide external guidance and support where they are needed most, and they might also allow for the most efficient resource allocation.

2. Employ as external reviewers **qualified experts who meet a prescribed standard for qualifications and required experience**. Robust training should initially be compulsory and then required on an on-going basis.

3. Incorporate a **broad set of school quality criteria that goes beyond standardized test scores** to adequately represent other valued aspects of quality teaching and learning. In particular, we recommend teaching quality be defined in relationship to a specific set of criteria.

4. Include the perspectives of **multiple stakeholders** (administrators, teachers, students, parents, community leaders, and researchers) in design efforts, allowing all key stakeholder groups to participate directly.

5. **Recalibrate and revitalize existing state resources** (e.g., regional offices of education) to support external review initiatives.

Notes and References

1 National Academy of Education. (2009). *Education policy white paper on standards, assessments, and accountability*. Washington, DC: Author.

2 Dee, T. & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30* (3), 418-446.

Figlio D., & Loeb, S. (2011). School accountability. In E.A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbooks of the Economics of Education (Volume 3)*. San Diego, CA: North-Holland, 383-421.

Linn, R. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies, 40* (6), 699-711.

National Academy of Education, 2009 (see note 1).

3 See the Nation's Report Card website, http://nationsreportcard.gov/about.asp, for more details.

4 National Center for Education Statistics. (2011). *Mathematics achievement of fourth- and eighth- graders in 2011. Trends in international mathematics and science study*. Washington, DC: Author. Retrieved March 15, 2013, from http://nces.ed.gov/timss/results11_math11.asp.

5 Carnoy, M., & Rothsten, R. (2013). What do international tests really show about U.S. student performance? (EPI Report). Washington, DC: Economic Policy Institute. Retrieved March 17, 2013, from http://www.epi.org/publication/us-student-performance-testing/.

Organisation for Economic Co-operation and Development. (2010). *PISA 2009 results: What students know and can do – student performance in Reading, Mathematics and Science (Volume 1)*. Paris: OECD Publishing.

Tucker, M.S. (Ed.). (2011). *Surpassing Shanghai: An agenda for American education built on the world's learning systems*. Cambridge, MA: Harvard University Press.

6 Elmore, R.F. (2009). The problem of capacity in the (re)design of educational accountability systems. In M.A. Rebell & J.R. Wolff (Eds.), *NCLB at the crossroads: Reexamining the federal effort to close the achievement gap*. New York: Teachers College Press, 230-261.

Kim, J.S. & Sunderman, G.L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equality. *Educational Researcher, 34*(8), 3-13.

7 U.S. Department of Education. (2009). *Race to the Top Program:* Executive summary. Washington, DC: Author.

U.S. Department of Education. (2012). *ESEA Flexibility*. Retrieved June 12, 2012, from http://www.ed.gov/esea/flexibility.

8 While not the purpose of this policy brief, we acknowledge and value the robust arguments for dismantling accountability in its current form while engaging in discussions about what ought to be done and why—the means and ends of education and educational accountability.

Ryan, K.E. & Feller I. (2009). Evaluation, accountability and performance measurement in national education systems - trends, methods, and issues. In K.E. Ryan, & J.B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation*. Los Angeles, CA: Sage, 171-189.

9 Ladd, H.F. (2012). School Accountability: To what ends and with what effects? Keynote address for Conference on Improving Education through Accountability and Evaluation: Lessons from Around the World, Rome, Italy. Retrieved March 15, 2013, from http://fds.duke.edu/db/attachment/2050.

10 Darling-Hammond, L. (2006). *Standards, assessments, and educational policy: In pursuit of genuine accountability* (updated) (Report No. PIC-ANG8). The eighth annual William H. Angoff Memorial Lecture. Princeton, NJ: Educational Testing Service.

Ladd, H.F. (2012). *School Accountability: To what ends and with what effects?* Keynote address for Conference on Improving Education through Accountability and Evaluation: Lessons from Around the World, Rome, Italy. Retrieved March 15, 2013, from http://fds.duke.edu/db/attachment/2050.

11 Organisation for Economic Co-operation and Development. (2011). *Education at a glance 2011: OECD indicators*. Paris: OECD Publishing.

Whitby, K. (2010). *School inspection: Recent experiences in high performing education systems*. Reading, United Kingdom: CfBT Education Trust.

12 Whitby, K. (2010). *School inspection: Recent experiences in high performing education systems*. Reading, United Kingdom: CfBT Education Trust.

13 Ladd, H.F. (2010). Education inspectorate systems in New Zealand and the Netherlands. *Education, Finance, and Policy, 5*(3), 378-392.

14 Ozga, J. (2009).Governing education through data in England: From regulation to self-evaluation. *Journal of Education Policy, 24(2)*, 149-162.

Perryman, J. (2006). Panoptic performativity and school inspection regimes: Disciplinary mechanisms and life under special measures. *Journal of Education Policy, 21*(2), 147-161.

Plowright, D. (2007). Self-evaluation and OFSTED inspection: Developing an integrative model of school improvement. *Educational Management Administration and Leadership*, (35)3, 373-393.

De Wolf, I.F., & Janssens, F.J.G. (2007). Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education, 33,* 379-396.

15 McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). *Evaluation of the impact of section 5 inspections – Strand 3: Final report for Ofsted*. Slough: National Foundation for Educational Research. Retrieved March 17, 2013, from http://www.nfer.ac.uk/nfer/publications/SFO01/SFO01.pdf.

16 Cowie, M., & Croxford, L. (2007, June). *Intelligent accountability: Sound-bite or sea-change?* (Report No. 43). Edinburgh: Centre for Educational Sociology, University of Edinburgh. Retrieved March 15, 2013, from http://www.ces.ed.ac.uk/PDF%20Files/Brief043.pdf.

17 Scotland very recently did away with top-down school inspections in favor of a more locally-based school governance model, thus excluded from the category of UK-style SSE/I models.

18 Gross, S.J. (2012). *Review of "On Her Majesty's School Inspection Service."* Boulder, CO: National Education Policy Center. Retrieved March 15, 2013, from http://nepc.colorado.edu/thinktank/review-on-her-majestys/.

Jerald, C.D. (2012). *On Her Majesty's School Inspection Service*. Washington, DC: Education Sector. Retrieved March 15, 2013, from http://www.educationsector.org/sites/default/files/publications/UKInspections-RELEASED.pdf.

Ladd, 2010. Education inspectorate systems in New Zealand and the Netherlands. *Education, Finance, and Policy, 5*(3), 378-392.

McDonnell, L.A. (2008). The politics of educational accountability: Can the clock be turned back? In K.E. Ryan & L.A. Shepard (Eds.), *The future of test-based accountability*. Mahwah, NJ: Lawrence Erlbaum, 47-67.

Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: getting accountability right*. Washington, DC: Economic Policy Institute.

19 Organisation for Economic Co-operation and Development. (2011). *Education at a glance 2011: OECD indicators*. Paris: OECD Publishing. p. 435.

20 Organisation for Economic Co-operation and Development. (2011). *Education at a glance 2011: OECD indicators*. Paris: OECD Publishing. p.434.

21 Organisation for Economic Co-operation and Development. (2011). *Education at a glance 2011: OECD indicators*. Paris: OECD Publishing.

22 Penzer, G. (2011). *School inspections: What happens next?* Reading, United Kingdom: CfBT Education Trust.

23 Reynolds, D., Sammons, P., De Fraine, B., & Townsend, T., Van Damme, J. (2011). *Educational effectiveness research (EER): a state of the art review*. Paper presented at the annual meeting of the International Congress for School Effectiveness and Improvement, Cyprus.

24 Bryk, A., Sebring, P.B., Allensworth, E., Luppescu, S., & Easton, J. (2009). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press

Reynolds, D., Hopkins, D., & Stoll, L. (1993). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement, 4*(1), 37-58.

Ritchie, R. (2007). School self-evaluation. In S. Kushner & N. Norris (Eds.), *Dilemmas of engagement: Evaluation development under new public management and the new politics. Chevy* Chase, MD: Elsevier, 85-101.

Sammons, P**.** (2008). Zero tolerance of failure and new labour approaches to school improvement in England. *Oxford Review of Education*, *34*, 651-664.

25 Cowie, M., & Croxford, L. (2007, June). *Intelligent accountability: Sound-bite or sea-change*? (Report No. 43). Edinburgh: Centre for Educational Sociology, University of Edinburgh. Retrieved March 15, 2013, from http://www.ces.ed.ac.uk/PDF%20Files/Brief043.pdf

Plowright, D. (2007). Self-evaluation and OFSTED inspection: Developing an integrative model of school improvement. *Educational Management Administration and Leadership*, (35)3, 373-393.

Watson, A.E., & Destler, K. (2008, April). *Improving school quality through collaborative governance: Lessons from abroad*. Paper presented at the Consortium on Collaborative Governance Mini-Conference, Santa Monica, CA.

26 National Academy of Education. (2009). *Education policy white paper on standards, assessments, and accountability*. Washington, DC: Author.

27 Broader Bolder Approach to Education Campaign. (2009, June 25*). School accountability: a broader bolder approach. Report of the accountability committee of the broader bolder approach to education campaign*. Washington, DC: Economic Policy Institute. Retrieved March 15, 2013, from http://www.epi.org/files/2011/20090625-bba-accountability-2.pdf.

Ladd, H. F. (2010). Education inspectorate systems in New Zealand and the Netherlands. *Education, Finance, and Policy, 5*(3), 378-392.

Linn, R.L. (2012). *Test-based accountability*. Princeton, NJ: The Gordon Commission. Retrieved March 15, 2013, from http://www.gordoncommission.org/rsc/pdf/linn_test_based_accountability.pdf.

McDonnell, L.A. (2008). The politics of educational accountability: Can the clock be turned back? In K.E. Ryan & L.A. Shepard (Eds.), *The future of test-based accountability*. Mahwah, NJ: Lawrence Erlbaum, 47-67.

Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: getting accountability right*. Washington, DC: Economic Policy Institute.

28 Jerald, C.D. (2012). *On Her Majesty's School Inspection Service*. Washington, DC: Education Sector. Retrieved March 15, 2013, from http://www.educationsector.org/sites/default/files/publications/UKInspections-RELEASED.pdf.

29 Gross, S.J. (2012). *Review of "On Her Majesty's School Inspection Service*." Boulder, CO: National Education Policy Center. Retrieved March 15, 2013, from http://nepc.colorado.edu/thinktank/review-on-her-majestys.

30 Scheerens, J., Ehren, M., Sleegers, P. & de Leeuw, R. (2012). Country background report for The Netherlands. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Paris: OECD Publishing. Retrieved March 15, 2013, from http://www.oecd.org/edu/school/NLD_CBR_Evaluation_and_Assessment.pdf.

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

MacBeath, J. (2005). Leadership as distributed: A matter of practice. *School Leadership and Management, 25*(4), 349-366.

32 Ehren, M.C.M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools—describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability, 25*(1), 1-41.

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

De Grauwe, A. (2008). School supervision: A tool for standardization or for equity. *International Institute for Educational Planning, 3,* 19.

33 Ofsted (2013). *School inspection handbook: Handbook for inspecting schools in England under section 5 of the Education Act 2005 (as amended).* Manchester, United Kingdom: Author. Retrieved March 15, 2013, from http://www.ofsted.gov.uk/sites/default/files/documents/inspection--forms-and-guides/s/School%20inspection%20handbook.pdf.

34 The Dutch Inspectorate of Education (2010). *Risk-based inspection as of 2009. Primary and secondary education*. Utrecht, the Netherlands: Inspectorate of Education. Retrieved March 15, 2013, from http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2010/Risk-based+Inspection+as+of+2009.pdf.

35 Ofsted (2013). *School inspection handbook: Handbook for inspecting schools in England under section 5 of the Education Act 2005 (as amended).* Manchester, United Kingdom: Author. Retrieved March 15, 2013, from http://www.ofsted.gov.uk/sites/default/files/documents/inspection--forms-and-guides/s/School%20inspection%20handbook.pdf.

36 MacBeath, J. (2005). Leadership as distributed: A matter of practice. *School Leadership and Management, 25*(4), 349-366.

Sanders, J.R., & Davidson, E.J. (2003). A model for school evaluation. In T. Kellegan & D.L. Stufflebeam (Eds.), *International handbook of educational evaluation*. Dordrecht, The Netherlands: Kluwer, 807-826.

37 Elmore, R.F. (2009). The problem of capacity in the (re)design of educational accountability systems. In M.A. Rebell & J.R. Wolff (Eds.), *NCLB at the crossroads: Reexamining the federal effort to close the achievement gap*. New York: Teachers College Press, 230-261.

38 MacBeath, J. (2010). Self-evaluation for school improvement. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second international handbook of educational change (Volume 2)*. New York: Springer, 901-912.

39 Cowie, M., & Croxford, L. (2007, June). *Intelligent accountability: Sound-bite or sea-change?* (Report No. 43). Edinburgh: Centre for Educational Sociology, University of Edinburgh. Retrieved March 15, 2013, from http://www.ces.ed.ac.uk/PDF%20Files/Brief043.pdf.

Ozga, J. (2009).Governing education through data in England: From regulation to self-evaluation. *Journal of Education Policy, 24*(2), 149-162.

Plowright, D. (2007). Self-evaluation and OFSTED inspection: Developing an integrative model of school improvement. *Educational Management Administration and Leadership, 35* (3), 373-393.

40 MacBeath, J. (2010). Self-evaluation for school improvement. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second international handbook of educational change (Volume 2)*. New York: Springer, 901-912.

41 United Kingdom, Parliament, House of Commons, Education Committee (2011). *The role and performance of Ofsted: Second report of session 2010-11* (HC570-1). London: Stationery Office. Retrieved March 15, 2013, from http://www.publications.parliament.uk/pa/cm201011/cmselect/cmeduc/570/570i.pdf.

42 See other examples of newspaper articles and opinion pieces debating the merits of SSE/I at http://www.free-school-from-government-control.com/Ofsted.html.

43 Whitby, K. (2010). *School inspection: Recent experiences in high performing education systems*. Reading, United Kingdom: CfBT Education Trust.

44 We intended to include research about Ofsted inspections conducted since 2011 when substantial changes were made to the inspection framework. In addition, an English law was passed to exempt outstanding schools from routine inspections. However, no empirical study fitting this criterion was found.

45 We intended to include Dutch inspection research conducted since 2011 when the Dutch Supervision Act was amended to emphasize early identification of potentially failing schools and risk-based inspections. However, no empirical study fitting this criterion was found.

46 Perryman, J. (2006). Panoptic performativity and school inspection regimes: Disciplinary mechanisms and life under special measures. *Journal of Education Policy, 21*(2) 147-161.

Thrupp, M. (1998). The art of the possible: Organizing and managing high and low socioeconomic schools. *Journal of Education Policy, 13*(2), 197-219.

47 Learmonth, J. (2000). *Inspection: What's in it for schools?* London: Routledge.

48 De Wolf, I.F., & Janssens, F.J.G. (2007). Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education, 33,* 379-396.

49 Matthews, P., & Sammons, P. (2004, July). *Improvement through inspection: An evaluation of the impact of Ofsted's work*. London: Institute of Education, University of London. Retrieved March 15, 2013, from http://217.35.77.12/archive/england/papers/education/pdfs/3696.pdf.

50 McCrone, T., Rudd, P., Blenkinsop, S., & Wade, P. (2006). *Impact of section 5 inspections: maintained schools in England*. Slough, United Kingdom: National Foundation for Educational Research.

51 McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). *Evaluation of the impact of section 5 inspections – Strand 3: Final report for Ofsted*. Slough: National Foundation for Educational Research. Retrieved March 17, 2013, from http://www.nfer.ac.uk/nfer/publications/SFO01/SFO01.pdf.

52 McCrone, T., Rudd, P., Blenkinsop, S., & Wade, P. (2006). *Impact of section 5 inspections: maintained schools in England*. Slough, United Kingdom: National Foundation for Educational Research.

53 McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). *Evaluation of the impact of section 5 inspections – Strand 3: Final report for Ofsted*. Slough: National Foundation for Educational Research. Retrieved March 17, 2013, from http://www.nfer.ac.uk/nfer/publications/SFO01/SFO01.pdf.

54 Plowright, D. (2007). Self-evaluation and OFSTED inspection: Developing an integrative model of school improvement. *Educational Management Administration and Leadership, 35* (3), 373-393.

55 Ehren, M.C.M., & Visscher, A.J. (2008). The relationships between school inspections, school characteristics and school improvement. *British Journal of Educational Studies, 56*(2), 205-227.

56 Brimblecombe, N., Shaw, M., & Ormston, M. (1996). Teachers' intention to change practice as a result of Ofsted school inspections. *Educational Management Administration & Leadership, 24*(4), 339-354.

Ouston, J., Fidler, B., & Earley, P. (1997). What do schools do after OFSTED school inspections-or before? *School Leadership & Management, 17*(1), 95-104.

57 McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). *Evaluation of the impact of section 5 inspections – Strand 3: Final report for Ofsted*. Slough: National Foundation for Educational Research. Retrieved March 17, 2013, from http://www.nfer.ac.uk/nfer/publications/SFO01/SFO01.pdf.

58 Learmonth, J. (2000). *Inspection: What's in it for schools?* London: Routledge.

Ouston, J., Fidler, B., & Earley, P. (1997). What do schools do after OFSTED school inspections-or before? *School Leadership & Management, 17*(1), 95-104.

59 Brimblecombe, N., Shaw, M., & Ormston, M. (1996). Teachers' intention to change practice as a result of Ofsted school inspections. *Educational Management Administration & Leadership, 24*(4), 339-354.

Chapman, C. (2001). Changing classrooms through inspections. *School Leadership and Management, 21*(1), 59–73.

60 Matthews, P., & Sammons, P. (2004, July). *Improvement through inspection: An evaluation of the impact of Ofsted's work*. *London*: Institute of Education, University of London. Retrieved March 15, 2013, from http://217.35.77.12/archive/england/papers/education/pdfs/3696.pdf.

61 McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). *Evaluation of the impact of section 5 inspections – Strand 3: Final report for Ofsted*. Slough: National Foundation for Educational Research. Retrieved March 17, 2013, from http://www.nfer.ac.uk/nfer/publications/SFO01/SFO01.pdf.

McCrone, T., Rudd, P., Blenkinsop, S., & Wade, P. (2006). *Impact of section 5 inspections: maintained schools in England*. Slough, United Kingdom: National Foundation for Educational Research.

62 Ehren, M.C.M., & Visscher, A.J. (2008). The relationships between school inspections, school characteristics and school improvement. *British Journal of Educational Studies, 56*(2), 205-227.

McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). *Evaluation of the impact of section 5 inspections – Strand 3: Final report for Ofsted*. Slough: National Foundation for Educational Research. Retrieved March 17, 2013, from http://www.nfer.ac.uk/nfer/publications/SFO01/SFO01.pdf.

63 United Kingdom, Parliament, House of Commons, Children, School and Families Committee. (2010). *School accountability: First report of session 2009-10. Volume 1, report, together with formal minutes*. London: Stationery Office, 63. Retrieved March 15, 2013, from http://www.publications.parliament.uk/pa/cm200910/cmselect/cmchilsch/88/88i.pdf.

64 James, C., Brammer, S., Connolly, M., Fertig, M., James, J., & Jones, J. (2011). School governing bodies in England under pressure: The effects of socio-economic context and school performance. *Educational Management Administration & Leadership, 39* (4), 414-433.

Matthews, P., & Sammons, P. (2004, July). *Improvement through inspection: An evaluation of the impact of Ofsted's work*. London: Institute of Education, University of London. Retrieved March 15, 2013, from http://217.35.77.12/archive/england/papers/education/pdfs/3696.pdf.

65 Sammons, P. (2008). Zero tolerance of failure and new labour approaches to school improvement in England. *Oxford Review of Education*, *34*, 651-664.

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

Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review, 23,* 143-151.

Shaw, I., Newton, D.P., Aitkin, M., & Darnell, R. (2003). Do OFSTED inspections of secondary schools make a difference to GCSE results? *British Educational Research Journal, 29,* 63–75.

67 Hussain, I. (2012). *Subjective performance evaluation in the public sector: Evidence from school inspections* (CEE DP 135). London: Center for the Economics of Education, London School of Economics. Retrieved March 15, 2013, from http://cee.lse.ac.uk/ceedps/ceedp135.pdf.

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

Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review, 23,* 143-151.

Shaw, I., Newton, D.P., Aitkin, M., & Darnell, R. (2003). Do OFSTED inspections of secondary schools make a difference to GCSE results? *British Educational Research Journal, 29,* 63–75.

68 Hussain, I. (2012). *Subjective performance evaluation in the public sector: Evidence from school inspections* (CEE DP 135). London: Center for the Economics of Education, London School of Economics. Retrieved March 15, 2013, from http://cee.lse.ac.uk/ceedps/ceedp135.pdf.

69 Shaw, I., Newton, D.P., Aitkin, M., & Darnell, R. (2003). Do OFSTED inspections of secondary schools make a difference to GCSE results? *British Educational Research Journal, 29,* 63–75.

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

71 Allen, R. & Burgess, S. (2012). *How should we treat underperforming schools? A regression discontinuity analysis of school inspections in England* (DoQSS Working Paper no. 12/02). London: Institute of Education, University of London.

72 Ladd, H.F. (2012). *School Accountability: To what ends and with what effects?* Keynote address for Conference on Improving Education through Accountability and Evaluation: Lessons from Around the World, Rome, Italy. Retrieved March 15, 2013, from http://fds.duke.edu/db/attachment/2050.

73 Illinois Association of Regional Superintendents of Schools. (2011). *Statewide ROE/ISC services and statistics*. Retrieved July 30, 2012, from http://www.roe21.k12.il.us/roe_stuff/Statewide%20Stats.pdf.

74 Hout, M., Elliott, S., & Frueh, S. (2012). Do high-stakes tests improve learning? *Issues in Science and Technology, 29* (1), 33.

Ladd, H.F. (2012). *School Accountability: To what ends and with what effects?* Keynote address for Conference on Improving Education through Accountability and Evaluation: Lessons from Around the World, Rome, Italy. Retrieved March 15, 2013, from http://fds.duke.edu/db/attachment/2050.

75 Elmore, R.F. (2009). The problem of capacity in the (re)design of educational accountability systems. In M.A. Rebell & J.R. Wolff (Eds.), *NCLB at the crossroads: Reexamining the federal effort to close the achievement gap*. New York: Teachers College Press, 230-261.

Porter, A.C., Chester, M.D., & Schlesinger, M.D. (2004). Framework for an effective assessment and accountability program: The Philadelphia example. *Teachers College Record, 106* (6), 1358-1400.

76 Ladd, H.F. (2012). *School Accountability: To what ends and with what effects?* Keynote address for Conference on Improving Education through Accountability and Evaluation: Lessons from Around the World, Rome, Italy. Retrieved March 15, 2013, from http://fds.duke.edu/db/attachment/2050.