September 2017

Would School Inspections Work in the United States?¹

Ashley Berner, Deputy Director

Introduction

The passage of ESSA (2015) brought new protocols for school improvement² and has led to a new interest amongst United States’ policymakers in the school inspection model. On-site inspections are widely used in Europe to evaluate schools. The appeal of inspections lies in their capacious understanding of school quality that includes not only students’ test scores but also leadership capacity, instructional strengths and weaknesses, historical performance, and demographics.³ What do such programs look like? Are they effective in raising student achievement? And what have been the challenges faced by early adopters in the United States?

Description of the English model

The most commonly known and widely cited school inspection regime is England’s. Her Majesty’s Inspectorate, as it is called, came into existence in the mid-19th century when Parliament first allocated funds for primary schools. The inspectorate has undergone numerous changes since 1837, most recently in 1992, when Parliament created the Office for Standards in Education, Children’s Services and Skills (Ofsted) as a non-governmental body to undertake detailed analyses of the nation’s schools. Ofsted reports directly to Parliament, and all inspection reports are public.

Distinguishing features:

Human Capital. Ofsted has a high bar for entry into the inspectorate. Its central office employs several hundred inspectors and holds contracts with three providers, who between them employ approximately 2,000 assistant inspectors to conduct site visits. Hiring “can take up to a year,” due to the experience necessary and to the extensive vetting process. One inspector reported in 2012,

¹ Elements of this commentary were developed with the partnership of Chiefs for Change. The Institute is grateful to Chiefs for Change for permission to use this material.
“Ofsted recruits individuals who have both breadth and depth of experience...For example, many have experience as head teachers - sometimes having led several schools, each of them successful - and also offer a specialization beyond school leadership.”

The new hires undergo training that may “include up to seven days of face-to-face sessions plus several days spent shadowing a team during a live inspection.” As they enter the field, inspectors receive in-depth feedback from their mentors and formal professional development.

**Site Visits.** Ofsted conducts visits of “good” schools every five years and “satisfactory” schools every three years. “Outstanding” schools are exempt from further inspection. Those in the lowest tier, “unsatisfactory,” receive full-blown site visits every one to two years and intermittent monitoring visits in between.

Ofsted prepares inspectors for each school visit with academic data that includes test scores in major subjects and at specified grade levels (known in England as “key stages”). The data include not only achievement but also student academic growth. Inspectors are charged with placing test scores in the cultural context of the school (student demographics, historical performance), thus enabling human judgment to engage with statistics and standards (about this, see below).

During site visits, inspectors place the highest premium upon visiting classrooms, but they also interview staff, review parent and student satisfaction surveys, and examine school records for attendance and grades. They are required to de-brief with any teacher whose class they observe for more than 20 minutes. The format of classroom observations varies. In some cases, the inspectors will visit most of the school’s classrooms for 20-30 minutes each; in others, they may “conduct whole-period observations to investigate certain issues more deeply, or they might shadow a group of students for all or part of the day.”

**Reports.** Following site visits, the inspection team prepares reports that grade schools on 27 indicators of performance that result in four tiers: outstanding, good, satisfactory, and inadequate. The distribution of tiers between September 2006 and July 2009: 13% ranked “outstanding,” 48% “good,” 33% “satisfactory,” and 6% “inadequate.” This distribution is highly correlated to other measures of school quality, such as student perceptions of teacher practices and parent satisfaction from the Longitudinal Study of Young People in England.

The 27 measures stretch across the school’s key domains. A table published by Ofsted shows the number of elementary schools that received these grades during school year 2010-11.

---

5 Ibid.
6 Ibid., 36.
Inspectors are trained to write clearly, giving specific directions but without generalizations or jargon. Schools are expected to discuss the Ofsted reports with the larger school community, including parents.

Monitoring at-risk schools. Schools in the lowest tier, “inadequate,” are placed in one of two categories: “notice to improve” or “special measures.” In schools under special measures, education officials
have the authority to change the staff and/or governing board or to close the school. In both cases, the local district has to submit a plan of improvement, and subsequent inspections monitor whether the school’s status has met improvement targets.9

Updates
Ofsted has come under criticism for creating unnecessary bureaucratic burdens upon teachers and schools. The Department of Education responded in February 2015 by studying teacher workload, which resulted in a slimmed-down inspection regime (shorter inspections for schools deemed “good,” exemption for schools deemed “outstanding”), a reduction in the artifacts that teachers needed to present to inspectors, and a pledge from Ofsted not to alter protocols during a school year.10

Inspections in Other Countries

The external school-inspection model is widely used across OECD countries for both improvement and accountability purposes:

In the majority of OECD countries schools are regularly inspected: this is the case in Australia, the Flemish Community of Belgium, the Czech Republic, United Kingdom, France, Germany, Greece, Iceland, Ireland, Korea, Luxembourg, Mexico, the Netherlands, Portugal, Spain, Sweden, Switzerland, and Turkey.11

All school inspection programs share the following characteristics:

• Regularized inspections for all schools, not merely those deemed at risk;
• Use of student data (whether school-based or national/provincial tests);
• Use of qualitative measures (such as interviews, surveys, and student absenteeism) to create performance indices; and

---

9 Consider Peterhouse, a Norfolk County school judged inadequate and placed into special measures in March 2010. By the time of its next full inspection, in May 2011, students’ academic attainment was still below average and earned a grade of “low” from inspectors. However, the headteacher, Martin Scott, had moved decisively to address the problems with classroom teaching that were spelled out in the previous inspection report, by adopting new ways to assess student progress and plan instruction, providing focused professional development, and tapping a new deputy headmaster to run an enhanced system of lesson observations. By May 2011, inspectors judged the quality of teaching at Peterhouse to have become not merely satisfactory but good.” Ibid., p. 10.


• Extensive training for inspectors.  

The consequences of school inspections vary, however. Some countries (among them Austria, Iceland, and Denmark) do not place sanctions upon low-performing schools and do not make reports public. In New Zealand, the Netherlands, and England, by contrast, reports are made public, “inspection is intensified at institutions with problems,” and the funding bodies can withdraw support (the Netherlands) or even close schools (the UK) when schools fail to reach targets in a timely fashion.  

Do school inspections boost student achievement?

It is difficult to evaluate the effect of universal policies that, by their nature, render appropriate control groups impossible. As the OECD notes, “When reforms are implemented nationwide, it is by definition difficult to assess the impact of a specific evaluation scheme and to establish a counterfactual.” Indeed, one of this field’s prominent researchers – Dr. M. Ehren - wrote in 2013, “Surprisingly, there is little research knowledge about how school inspections drive the improvement of schools and which types of approaches are most effective and cause the least unintended consequences.”

Analyses of the effect of inspections upon student performance are few and ambiguous. A study of the effects of the re-authorized Ofsted inspection process between 1992-1997 found mixed results: among selective state-run schools, inspections had a slight, positive effect of 0.5% upon students’ GCSE (General Certificate of Secondary Education) test results in the same year; among non-selective state-run schools, inspections had a slight negative effect upon the same. A second study during the same period also found a slight reduction in non-selective, state-run students’ GCSE test scores during the year of inspection. By contrast, Hussain’s 2013 examination of the effect of Ofsted inspections on English primary schools that it deemed “unsatisfactory,” found a rise in student test scores in post-inspection years. The OECD’s review of literature on the empirical

---

12 Ibid., 14-19.
13 Ibid., 25.
14 Ibid.
16 National, subject-specific exams administered to 15-year-olds.
19 He found that failure rating “raises standardized test scores in a school by 0.12 standard deviations in math and by 0.07 to 0.09 standard deviations in English. These gains, which roughly equate to between one-third and one-half a year of typical instruction,” are achieved in the eight months between reviews and testing. Iftikhar Hussain, “Subjective Performance Evaluation in the Public Sector Evidence from School Inspections,” Journal of Human Resources 50, no. 1 (January 1, 2015): 189–221, doi:10.3368/jhr.50.1.189.
An emerging body of evidence provides indications that the threat of sanctions on low-performing schools can raise student test scores in the short run. However, the extent to which these test score improvements are due to schools’ gaming of the accountability system is unclear.

The majority of studies investigate the influence of inspections upon the expectations and experiences of principals and teachers. Such studies do indicate that there can be positive results from clear messaging about the high standards required for a “good education”: “Inspection [in the Netherlands and the UK] does result in significant impact on schools where clear expectations, norms, and standards are set and where stakeholders are knowledgeable about and engaged with the process. School improvement through school inspections seems to take place through indirect developmental processes rather than through more direct coercive methods (emphasis added).”

This indirect effect is confirmed by a literature review undertaken in 2012 by a member of the Dutch Inspectorate, who concluded:

Overlooking the overall results of the systematic review, no evidence has been found that school inspections automatically lead to the improvement of the educational quality. Also it cannot be concluded that characteristics of school inspections themselves lead to the improvement of educational quality. Instead, research shows that in practice there is a complex interaction between different characteristics of school inspections and the inspector on the one hand, and the school with its pupils, teachers and management on the other hand.

This indirect effect is similar to what Hanushek and Raymond concluded about the United States’ accountability protocols in general (not inspections per se): accountability measures do advance performance outcomes for students, but this impact is driven largely by the informational aspects of accountability rather than from any explicit negative consequences of failure.

---

20 And of high-stakes accountability in general.
23 Ehren et al., “Impact of School Inspections on Improvement of Schools—describing Assumptions on Causal Mechanisms in Six European Countries.”
Inspections in the United States

The United States’ closest analog to Ofsted is charter-school evaluation and re-authorization protocols. For example, New York State requires site visits of Regents-authorized charter schools, and its protocols mirror many of Ofsted’s. The Regents’ representatives undertake two-day site visits during which they visit classrooms, inspect financial documents, survey important stakeholders (principals, teachers, board of governors), and conduct focus groups. The subsequent report becomes a key text in the Regents’ decision to re-authorize or not.26

Several states and districts in the United States have put non-charter inspections in place. Some (Vermont, New York City) are universal, whereas others (such as Ohio and New York State) focus exclusively on low-performing schools and/or on schools that receive federal funds such as Title I or School Improvement Grants.27 The Institute could not locate independent evaluations of program effects, with the exception of a 2016 analysis of NYC’s School Quality Reviews (see below). A few examples:

**New York City (universal).** In 2006, NYC hired a British company to design and lead on-site School Quality Reviews (SQR), a version of which is still in place.28 SQRs are universal and include two-day site visits by experienced educators, who assess the school along ten indicators of quality.29 The reports are made publicly available, and schools that receive low marks are subject to a series of sanctions from the NYC Department of Education.30 Under Mayor Bill de Blasio, the SQRs began to collect different data and ceased to include a letter grade. According to a recent study by Marcus Winters, removing the letter grade had a negative effect upon school improvement.31

**New York State (low-performing schools).** In addition to its on-site reviews for charter schools (see above), New York State Education Department (NYSED) undertakes on-site reviews of schools in “focus districts,” which comprise their lowest-performing schools and districts.32 NYSED’s description:

---

28 Jerald, On Her Majesty’s School Inspection Service. Education Sector Reports.,
NYSED prepares its focus districts and schools to diagnose problems at a granular level and equips its inspectors to evaluate progress against district and school improvement plans. NYSED had put a similar process in place for school turnarounds under the federal government’s School Improvement Grants, 2011-2015. Two challenges in NYS have been, first, the need for SEA staff to train inspectors at the district level – a time-consuming task with uncertain quality control - and second, the perceived needs for follow-up visits, for which funding has often been inadequate.

Ohio (low-performing schools). In 2008, Ohio instituted School Improvement Diagnostic Reviews (SIDRs) for schools identified as academic underperformers. Inspectors, who have undergone “extensive training,” conduct two-day visits to collect evidence from teachers, classrooms, and school leaders. Their reports include recommendations for improvement. The Ohio Department of Education describes the purpose as follows:

---

35 Conveyed in conversation by David Steiner, Executive Director of the Johns Hopkins Institute for Education Policy and former New York State Commissioner of Education, on June 22, 2017.
36 Jerald, On Her Majesty’s School Inspection Service. Education Sector Reports.
The Department’s explanation of process (from identification to intervention) references evidence of best practices: “Over a period of nearly 10 years, ODE has recognized its high-risk/high-performing schools, publicized their achievements and studied their success stories through case studies and analysis of common themes. These findings have informed the development of Indicators of Effective Practice that provide the underlying framework for the SIDR.”

Vermont (in pilot stage but soon to be universal). Vermont is currently piloting a school-inspection program with the intention of scaling up across all schools. As described in EducationWeek, Vermont sends a team of reviewers to a local district, with three or four reviewers visiting each school. The inspectors interview teachers, observe classes, and examine school data including test reports. Each school receives a visit from two different teams. The resulting “integrated field reviews” are not (yet?) part of the state’s accountability protocols and as such carry no negative consequences. They are, rather, seen as information that is valuable to teachers, principals, and school board members.

The above examples are illustrative rather than exhaustive. The majority of this country’s school-inspection programs are geared to assist low-performing schools rather than to evaluate all schools. As in the case of the inspections regimes in Europe, we know little about the mechanisms by which they achieve success (or do not).

Inspections in the USA

---


As school systems in the United States adapt to the principles and practices of charter authorization and private-school scholarship programs, is there a role for an inspectorate?

The most immediate appeal is this: a rigorous program of school inspection provides multiple measures of school quality and offers considerable nuance beyond merely a snapshot view of student test scores. For example, Ofsted’s use of contextual data reassures school leaders that the school will not be penalized for offering places to students from disadvantaged backgrounds, and its interviews with teachers and principals can shed light on underlying impediments to school improvement (for example, leadership issues) that need to be rectified. Moreover, education leaders from across the spectrum could get behind a well-rounded examination of school culture and student achievement.

Second, school inspections – when made public - can help parents make informed enrollment decisions. This will be particularly important as systems become more flexible and responsive to parents’ wishes and students’ needs. Again, however, public reporting should not focus on test scores alone but, rather, provide a wider picture of a school’s values and results. England offers a model here, as well: school websites must make plain their ethos and values, curriculum, and special education policies along side assessment results.40

Finally, and perhaps most importantly, school inspections can change the cultural expectations around education. Inspectors in England and in the Netherlands, for instance, operate with a clear understanding of what a “good education” looks like and, as such, reinforce expected norms that exert an indirect effect upon the culture of schooling – as numerous surveys of principals and teachers attest.41

Policymakers will also want to be mindful of operational challenges and potential pitfalls as they craft an inspection regime. The first is the possibility of levying excessive bureaucratic burden. RAND’s 2011 report on expanded measures for school success notes that, “Those who develop or mandate new measures should consider not only the monetary costs but the burdens that additional measurement could impose on educators’ and students’ time. For example, school quality reviews, such as are done in Charlotte-Mecklenburg, are resource-intensive because they require substantial preparation by school staff and a trained team that observes the school for at least two days and prepares a report with its findings.”42 Indeed, this unintended consequence of school inspections was a causal factor in the 2015 reduction in Ofsted’s requirements placed on schools, as discussed above.

Cost is another issue. Craig Jerald’s (2012) estimate of the annual cost for English-style inspections in the United States ranged from a lower bound of $645 million to an upper bound of $1.13 billion.43

---

43 Jerald, On Her Majesty’s School Inspection Service. Education Sector Reports.
- an amount substantially below Richard Rothstein’s 2008 estimate of $2.5 billion annually.\textsuperscript{44} Charlotte-Mecklenburg’s foray into school inspections, called “school quality reviews,” was short-lived (2009-2011) because of the high cost to field and maintain the site visits.\textsuperscript{45} It is not clear how to cost out such a program in advance so as to ensure its sustainability.

Third, a well-run inspectorate requires high-quality expertise – and lots of it. The process by which Ofsted hires, prepares, and evaluates inspectors, described above, is extensive and assumes a pool of highly qualified educators who have already assumed significant school leadership. Quantity also matters, particularly when teams need to visit a school multiple times after a negative site visit. In the case of New York State, follow-up visits have not been possible – and thus, the efficacy of site visits is blunted.\textsuperscript{46} Some states, such as Louisiana,\textsuperscript{47} have made substantial headway in training teacher leaders and thus possess a natural advantage, as do school systems that work with high-octane programs such as Leading Educators.

A final tension is cultural habit about judgment versus compliance. The English inspectorate places a high premium on human judgment - a point to which Craig Jerald returns again and again in his 2012 description of the English system. School systems in the United States, however, often favor bureaucratic processes that remove human judgment as much as possible. As Chubb and Moe highlighted in 1990, detailed administrative procedures and verifications diminish - indeed, are designed to diminish - the role of individual judgment.\textsuperscript{48} Embracing (well-trained) judgment would be a boon to teachers and students alike, but getting there would require sustained leadership and patience.

Citations:


\textsuperscript{46} Conveyed in conversation by David Steiner, Executive Director of the Johns Hopkins Institute for Education Policy and former New York State Commissioner of Education, on June 22, 2017.


