

## The strengths and limitations of using quantitative data to inform school inspections

John Jerrim & Alex Jones

**To cite this article:** John Jerrim & Alex Jones (05 Apr 2024): The strengths and limitations of using quantitative data to inform school inspections, School Effectiveness and School Improvement, DOI: [10.1080/09243453.2024.2337640](https://doi.org/10.1080/09243453.2024.2337640)

**To link to this article:** <https://doi.org/10.1080/09243453.2024.2337640>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 05 Apr 2024.



Submit your article to this journal [↗](#)



Article views: 851



View related articles [↗](#)



View Crossmark data [↗](#)

# The strengths and limitations of using quantitative data to inform school inspections

John Jerrim<sup>a</sup> and Alex Jones<sup>b</sup>

<sup>a</sup>UCL Social Research Institute, University College London, London, UK; <sup>b</sup>Insights and Research, Ofsted, London, UK

## ABSTRACT

School inspections are a common feature of many education systems. These may be informed by quantitative background data about schools. It is recognised that there are pros and cons of using such quantitative information as part of the inspection process, though these have rarely been succinctly set out. This paper seeks to fill this gap by presenting arguments both for and against the use of quantitative data in informing school inspections. We argue that while quantitative data provide objective information about important outcomes, their usefulness is limited somewhat by a range of factors including missing data, small sample sizes, the creation of perverse incentives, and the fact that most readily available measures capture aspects other than school quality. We conclude by discussing how the Office for Standards in Education, Children's Services and Skills (Ofsted) – the school inspectorate in England – currently makes the trade-off between these pros and cons.

## ARTICLE HISTORY

Received 10 July 2023  
Accepted 27 March 2024


## KEYWORDS

Ofsted; school inspection;  
quantitative data

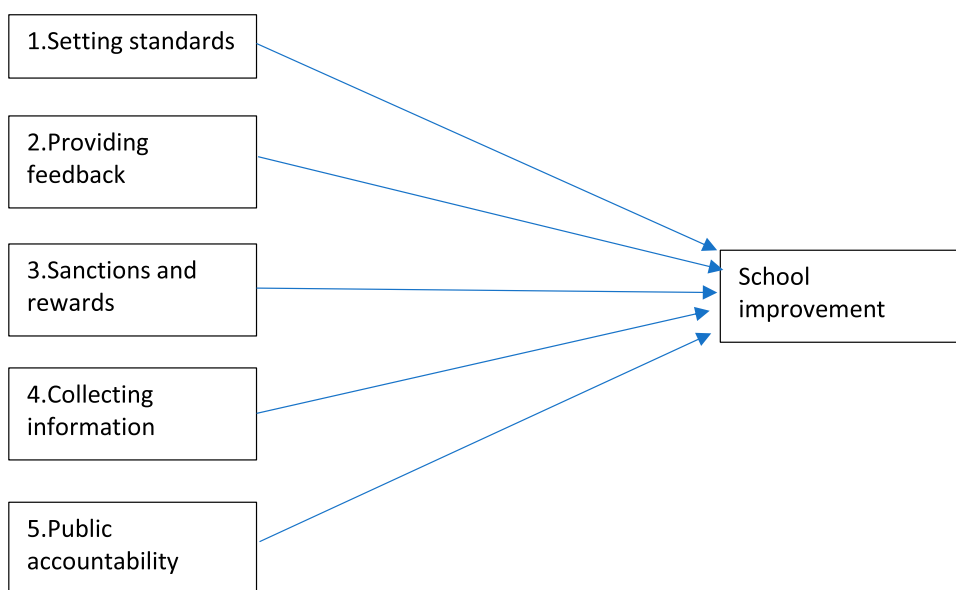
## Introduction

Inspections are a key part of the school accountability architecture in many countries. Such inspections typically involve a team of experienced professionals observing how the school functions, witnessing aspects of its provision, and ultimately judging the school's overall effectiveness. In doing so, it is hoped that schools – and education standards more generally – will improve as a result. [Figure 1](#) provides a broad framework for understanding how inspections may help to foster school improvement, building upon the logic model presented by Jones and Tymms (2014).

Within this framework, five key mechanisms are at play. First, school inspectorates help to set educational standards in a country. In England, these standards are made clear within the education inspection framework (EIF; Office for Standards in Education, Children's Services and Skills [Ofsted], 2023a). Schools then have the responsibility to ensure these standards are met. Second, inspectors – who are experienced education professionals – provide feedback on areas where a school could improve. This may include

**CONTACT** John Jerrim  jjerrim@ucl.ac.uk

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.



**Figure 1.** Jones and Tymms' framework for understanding how school inspections promote school improvement.

Note: Adapted from Jones and Tymms (2014).

inspectors offering their views of a school's own self-evaluation, or advising on the areas they should prioritise for development in the future. Third, inspection judgements may lead to sanctions or rewards. These encourage failing schools to improve to avoid the negative consequences, while schools that are already doing reasonably well are incentivised to develop further. Examples from England include enforced changes to management and governance structures upon receipt of an Inadequate judgement (a sanction), while until recently schools judged Outstanding were exempt from further routine inspection (a reward). Fourth, school inspectors gather a huge amount of data which can feed into system-level policy decisions and help to disseminate best practice across the school sector. Finally, in some countries, inspection judgements form a key part of public accountability. The fact that inspection findings are made publicly available may lead to parents putting pressure on schools for changes to be made, with the leadership team compelled to act to keep admission numbers high. Different inspectorates will, of course, pull upon these levers to different extents. For instance, while public accountability is a key feature of school inspections in England, the same is not necessarily true elsewhere.

Given the prevalence of school inspections in many countries, it is perhaps unsurprising to find they have been the subject of a great deal of research. For instance, in a cross-national analysis of seven countries, Altrichter and Kemethofer (2015) found that headteachers who feel more pressure from school inspections are more attentive to the quality standards set and more actively pursue school improvement activities. They also noted how school leaders in some countries feel more pressured by inspection than others. Gaertner et al. (2014) investigated how inspected and uninspected schools in Germany compared in terms of their school improvement activities. They

found that teachers' and leaders' perceptions of school quality were largely the same regardless of whether the school was inspected or not. Ehren et al. (2015) investigated the impact of school inspections across six countries. Their analysis suggests that inspections are most likely to lead to improvement when they evaluate both the educational practices and outcomes of schools, with the findings then made publicly available. Swiss school inspections use a traffic light system, where six aspects of school quality are allocated a green, yellow, or red rating. Quesel et al. (2021) evaluated this approach, finding that it generally had high acceptance rates amongst staff. Studying the Dutch inspection system, Ehren et al. (2015) drew upon teacher surveys to suggest that inspections which set clear expectations are the most likely to lead to staff taking the required action to improve. On the basis of survey data from over 2,000 primary and secondary school teachers in Flanders, Penninckx et al. (2016) argued that, for school inspection to facilitate improvement, teachers need to believe in the quality of the inspection process. Results from the 2022 round of the Programme for International Student Assessment (PISA) also illustrate how lesson observations by inspectors are correlated with mathematics performance, especially in systems where schools have high autonomy over curriculum such as in England. This encompasses the validity and reliability of the inspections, and the transparency of how decisions are reached. Relatedly, Quintelier et al. (2020) found that – for teachers to accept the feedback they receive from inspections – they need to perceive the process as fair, and the feedback received as relevant.

Several studies have directly focused on Ofsted – England's school inspectorate – which is also the focal setting of this paper. Parts of this literature have questioned whether these inspections actually help schools to improve. For instance, Munoz-Chereau et al. (2022) found that receipt of a poor Ofsted judgement can act as a barrier to school improvement, initiating a downward cycle where a low rating leads to a more disadvantaged pupil intake and higher levels of teacher turnover. von Stumm et al. (2022) argued that Ofsted judgements are only weak predictors of student outcomes, and that they do not provide a robust measure of a school's quality. In contrast, Allen and Burgess (2012) found that schools narrowly "failing" their Ofsted inspection (defined as receipt of an Inadequate grade) improve pupil's test scores by slightly more than schools that narrowly "pass" (schools that just receive a Satisfactory grade). Using administrative data, Hutchinson (2016) raised questions regarding the equity of Ofsted inspections, noting that schools with more disadvantaged pupils are disproportionately likely to receive worse inspection ratings. Another strand of this literature has investigated issues related to inspection consistency and reliability. For instance, Bokhove et al. (2023a) noted how female inspectors award slightly harsher judgements to primary schools than their male counterparts, with a more sizeable difference emerging between permanent and contracted members of Ofsted staff. Similarly, Bokhove et al. (2023b) found that schools with an Ofsted inspector on their payroll tend to achieve better inspection grades than schools that do not.

Yet, despite this attention, certain issues surrounding school inspection remain somewhat underexplored, or less frequently discussed. One such example is the role that quantitative data should have in the inspection process; this is despite quantitative data informing many important decisions that school inspectorates make. For instance, quantitative data play a key role in Flanders' "customized" system of school inspections,

with it helping to determine the content and frequency of inspections, their length, and the number of inspectors involved (Onderwijs Inspectie, n.d). Similarly, the 2021 inspection framework from the Netherlands noted how data are used to identify risks early and to gain insight into the functioning of school governing boards. This includes analysis of “financial data, data on staff, safety at schools, pupils’ results and how quickly those were achieved”, at least once per year (Inspectie van het Onderwijs, 2023, Section 7.3.3, p. 48). In Sweden, data – including financial information, academic performance, and the views of students, parents, and teachers as captured through biannual questionnaires – have a role in determining when a school gets inspected (Skolinspektionen, 2024). Data on student outcomes, attendance, and injuries occurring at school – along with parent and student questionnaires – are also used in the Czech Republic, forming part of the evaluation of schools (Czech School Inspectorate, 2015). On the other hand, data play a higher level role in countries such as Norway, helping the inspectorate to select municipalities (rather than individual schools) for inspection and the themes that the inspections cover (Norwegian Directorate for Education and Training, 2015).

England – the empirical setting of this paper – is a particularly data-rich country, with a wide variety of sources providing important contextual information about primary and secondary school. Ofsted has previously made extensive use of these data to inform its inspection judgements. This, however, led to criticism that inspection outcomes were too closely tied to performance in national examinations (Nye & Rollett, 2017). The introduction of the EIF in September 2019 changed this, with inspections becoming more focused on the quality and sequencing of a school’s curriculum (Ofsted, 2023a). While data continue to play a role – including in selecting how schools are inspected and providing contextual information to inspectors before an inspection takes place – they are now a less direct determinant of the inspection outcomes than previously.

These recent changes to the role of data in England’s school inspection lead one to ponder several important questions. What exactly are the pros and cons of using quantitative data to inform school inspections? How do these trade-offs vary across different parts of the inspection process (e.g., selecting when and how schools are inspected; to gather evidence from students, teachers, and parents; directly determining inspection outcomes)? And, ultimately, what is the optimal role for data to play in the inspection of a country’s schools?

The contribution this paper makes to the existing literature is to discuss such matters in detail, and succinctly set out (in a non-technical manner) the pros and cons of using quantitative data in school inspections. Our aim in doing so is to help inform readers’ own views of the role that quantitative data should play in the inspection of schools.

### **What quantitative data are available to inform school inspections in England?**

A host of information about England’s schools is available from administrative records. One important data set is the National Pupil Database (NPD). This encompasses a set of linked data sources including achievement in national examinations throughout school, school/Early Years/Alternative Provision censuses, Individual Learner Records, Children in Need information, and higher education records. Such information can

be aggregated from individual into year group and/or school-level averages to inform inspections. Together, this provides information in the inspection data summary report (IDSR) inspectors receive prior to an inspection taking place (along with interpretation of what these data show – see below for further details). This includes information such as:

- a measure of progress made between key stages (e.g., between Key Stage 2 and Key Stage 4 for secondary schools), which for primary schools encompasses English and mathematics, and for secondary schools the subset of curriculum subjects that pupils are still studying at the end of Year 11;
- attainment in national examinations (including 3-year trends);
- the pattern of subject entries in national examinations (which may, e.g., indicate whether a school may be disproportionately encouraging pupils to take qualifications which they deem to be easier);
- pupil movement between schools;
- absence rates;
- suspensions and permanent exclusions;
- Key Stage 5 qualification types and retention;
- destinations (i.e., the percentage of secondary school pupils that are in employment, education, or completing an apprenticeship after graduation);
- the percentage of pupils in receipt of free school meals (FSM);
- ethnic composition of the school;
- the percentage of pupils whose first language is not English;
- the percentage of pupils with special educational needs;
- the number of pupils on roll.

These data are then complemented with other information about the school. For instance, the IDSR will include information about the trust or local authority a school is part of, such as the number of schools and their Overall Effectiveness grades. Information is also available from the School Workforce Census, including details about staff vacancy rates, absences, and turnover. Information about the local area includes its level of deprivation. Similarly, financial information is provided to inspectors prior to their inspections, including spend per pupil, financial reserves, and the level of grant funding the school receives. Together, the above demonstrates the extent of background data available about schools, encompassing information about its workforce, finances, local area, and governance structures, in addition to pupil attainment, absences, suspensions, and outcomes.

Inspectors also conduct some additional quantitative data collection as part of the inspection process via surveys with pupils, parents, and staff. The pupil survey asks about their enjoyment of school, their experiences there, whether there are any behaviour/bullying issues, and the extent that the school encourages them to look after their physical health and wider wellbeing. A similar range of topics are included in the parent survey, including about their children's happiness and safety at the school, behaviour and bullying, whether the school contributes to their child's broader personal development, and their overall satisfaction with the school. Finally, the teacher survey asks staff about their views on behaviour in the school, the general school environment, the school's leadership, and their enjoyment of working there.

## **The advantages of using quantitative data to inform school inspections**

In this section we offer six benefits of using quantitative data to inform school inspections. Although this list may not be exhaustive, we believe it encompasses the key reasons why drawing on quantitative data may be useful.

### ***Data capture some key outcomes known to affect young people's future***

One important piece of quantitative information available about schools is how its pupils perform in national examinations, along with the academic progress they have made (e.g., between the end of primary school and the end of secondary school). General Certificate of Secondary Education (GCSE) grades and Key Stage 2 test scores are known to predict future life outcomes (Hodge et al., 2021). If many young people are leaving school without the intended knowledge and skills, then schools are failing to achieve a key part of their remit. Quantitative data – when used appropriately – can provide inspectors with some tangible evidence on this matter. Critically, it can help inspectors identify where some of the key functions of schools – such as ensuring young people learn to read and write to a sufficient standard – are not being achieved.

### ***Data may provide warning signs about declining quality of provision***

One may reasonably expect inspectorates to prioritise inspections of schools where challenges are starting to emerge. In this case, quantitative data may provide some early warning signs that something about a school may be amiss, or at least warrants further investigation. Of course, as examination data refer only to the year group that most recently left the school, they will capture the cumulative impact of a school over several years. Such data will hence suffer from time lags in terms of providing information about school “quality”.<sup>1</sup> However, other quantitative data – such as a notable increase in pupil or staff absence at a school, or unusually high levels of staff turnover – may be pointing towards an increasingly difficult situation, suggesting it may be prudent for an inspection to take place to find out more. This of course includes what can be done to help the school address any challenges that it is facing.

### ***Data can help inspectors to understand the context of a school***

Schools have different student bodies, each bringing their own unique challenges. Some will have a high proportion of pupils from disadvantaged socioeconomic backgrounds who may lack access to the resources they need outside of school to fully support their progress. Others may serve communities where English is often spoken as an additional language, meaning staff may have to navigate complex linguistic and cultural barriers. Yet another school may be in a leafy middle-class suburb, where affluent but time-pressured parents create a rather different challenge. It is important for inspectors to be aware of such circumstances, to help them understand the school environment and to contextualise what they go on to observe. Access to quantitative data about the local area, composition of the student body, and school environment can provide such contextual information in an efficient way, early in the inspection process.

### ***Data have high levels of cross-school comparability***

Using quantitative data to judge schools may help ensure they are evaluated using the same metrics, providing clear benchmarks that their performance can be compared against. Although interpretation of the quantitative data may still differ across inspectors, the fact that decisions are being made on the basis of the same information should lead to a greater level of consistency than if purely qualitative judgements were made alone. In other words, when inspections draw on objective quantitative data, there should, at least in theory, be less room for subjectivity and bias to play a role. The use of quantitative data may hence lead to greater levels of between-inspector consistency.

### ***Data provide objective information about a school and its performance***

Relatedly, a key benefit of an inspection regime grounded in quantitative data is objectivity; inspectors would draw on a set of key indicators and use these to evaluate each school. In the extreme situation where only background quantitative data are used to inform inspection judgements, there would be little in the way of subjective, professional judgement involved. Such reliance on objective quantitative data would likely be beneficial for ensuring consistency across different inspectors. It may, however, also have some negative consequences as well (see the next section for further details).

### ***Surveys provide an opportunity for all key stakeholders to offer their view***

With limited inspection time available, inspectors are unable to speak to all stakeholders in a school. Through primary quantitative data collection – in the form of pupil, parent, and teacher surveys – there is an opportunity for everyone to offer their view. This should help inspectors triangulate the in-depth conversations they have with a smaller number of individuals with views held across the broader school community. In other words, these additional data should help inspectors better understand variation in views about the school, how the bigger picture looks across a wider body of individuals, and the generalisability of the smaller number of qualitative perspectives that they hear.

## **The limitations of using quantitative data to inform school inspections**

### ***Increases the stakes attached to examination results and league tables***

In a world where inspection outcomes are heavily influenced by data, it significantly raises the stakes attached to examination results and school performance measures; that is, poor performance on national assessments could lead a data-driven inspectorate to reach an Inadequate or Requires Improvement judgement. This may then exacerbate the negative consequences often associated with too much emphasis placed upon national examination results. This includes:

- schools feeling pressure to “teach to the test”, focusing time and resources on the subset of areas that will inform the inspection judgement (e.g., mathematics and English at Key Stage 2);



- schools removing intended content from the curriculum in order to devote more time to other parts (e.g., teaching Key Stage 4 content in Year 9 by removing elements from the Key Stage 3 curriculum);
- focusing teaching on the parts of the curriculum most likely to be tested (e.g., teaching certain aspects of mathematics that schools deem most likely to appear on the Key Stage 2 tests);
- teaching pupils' "examination technique" (e.g., how to answer test questions) rather than developing their knowledge and skills in a given subject;
- entering pupils into certain subjects to maximise "league table" position, rather than because it is in the pupils' best interest. A widely cited example is the European Computer Driving License (Nye, 2018).

Due to each of the above, the curriculum pupils receive may get narrowed. It may also lead to teachers and school leaders feeling increasingly stressed by these assessments, having a negative effect on their wellbeing and retention in the teaching profession. Likewise, it increases incentives for perverse behaviour such as "off-rolling" the academically weakest pupils in the school. There is also evidence to suggest that such pressures can lead to maladministration; a Teacher Tapp poll has found that one in five teachers involved in the administration of Key Stage 2 Statutory Assessment Tests (SATs) have been encouraged to engage in questionable practices, such as pointing out incorrect answers (Teacher Tapp, 2019); see Koretz (2008) for further discussion of the unintended consequences that can result from overreliance on educational test data.

### ***Which measure(s) to use?***

If data are to be used to inform inspection outcomes, then which measure(s) should we draw upon? After all, we often have multiple pieces of information available, with no single measure always being unequivocally better than others. Rather, each measure will have different strengths and weaknesses, with a decision having to be made about which one – or what combination – is the most appropriate.

Take, for instance, attainment at Key Stage 4 in England. One may argue that a data-informed inspection regime should focus on the percent of pupils failing to achieve a certain threshold (e.g., achieving at least five standard GCSE passes including English language and mathematics) given that these reflect the knowledge and skills we expect all young people to have by the time they finish school. Thus, if secondary schools are not reaching this minimum standard, an inspectorate might step in and award a Requires Improvement or Inadequate judgement.

However, secondary schools do not all have the same intakes, with some having pupils who are (on average) further behind than others upon entry. One may hence argue that schools should not all be judged according to the same end point, given they can face markedly different starting positions. A prime example here are selective grammar schools – where pupils are amongst the highest achievers in an area on entry and thus disproportionately likely to achieve any minimum threshold set. It is such an argument that led the Department for Education in England to turn to Progress 8 (a metric which captures the academic progress a child makes between the end of primary and secondary school) as its headline measure at secondary school. But, as noted by Leckie and Goldstein

(2017), what government – and a very data-informed school inspectorate – would really want is a measure of the academic progress pupils make purely due to *school quality*. Progress 8 does not capture this; rather it partially reflects effects that are realistically outside of a school’s control (see next subsection for further details on this point). It thus only partly overcomes the problem of different schools being attended by pupils with different characteristics, with the advantage enjoyed by some school types in achieving strong examination score metrics (such as selective grammar schools) only partially controlled (see Prior et al., 2021, for further discussion of the strengths and limitations of Progress 8 as a measure).

Given this, a case can be made for a *contextualised* value-added measure, where different rates of progress are expected for schools with different pupil intakes (e.g., pupils in secondary schools with a mainly socioeconomically disadvantaged intake would be expected to make less progress than their peers in schools with few disadvantaged pupils). In many ways, this would be the most appropriate measure for a purely data-driven school inspectorate to use in its consideration of examination performance. Yet this would clearly lead to the difficult situation of explaining why the school inspectorate has lower expectations for the academic progress of disadvantaged pupils, and that by doing so, it is encouraging educational inequalities to be maintained. In other words, it could lead to undesirable “backwash” in the education system, which would work against the interests of disadvantaged socioeconomic groups. Others have also argued that such contextualised measures of progress can be unstable (Gorard, 2010). It is for such reasons that the Department for Education abandoned contextual value added as a headline accountability measure more than a decade ago (Leckie & Goldstein, 2017).

What this discussion hopefully illustrates is that any system of inspection that is too reliant on data faces the challenge of which measure(s) to use – or at least which measures to focus on. Different measures have different pros and cons, as well as requiring different levels of statistical literacy to interpret appropriately and understand the caveats involved. They thus also put different burdens on the statistical understanding of inspectors (see below for further discussion). While multiple measures could be used, this heightens issues surrounding interpretation and how to balance the various sources of information. And, of course, the measure(s) that a data-driven inspectorate chooses will have a disproportionate impact on a school’s inspection results (i.e., if different measure[s] were chosen, inspection judgements would likely change).

### ***The attribution of examination results to schools***

Building on the above, one of the key issues with using examination and test score data is that achieving good or poor results may not be attributable to the school. Rather, other factors may be at play. Take the following hypothetical example. A school with a largely middle-class intake has a weak leadership team, behaviour in the school is poor, and the overall quality of teaching is low. Parents are likely to recognise this. Rather than sit back and do nothing, many may use their own resources – such as paying for private tuition – to minimise the impact this has on their child’s education. Thus, despite the clear problems with this school, it will continue to look good in the published “league tables”.

Two key points from this example thus emerge. First, one should not automatically attribute strong (or weak) examination results to the actions of a school. This may or may not be the case. Second, such issues will lead to what – if taken at face value – appear to be discrepancies in “school quality” as measured by inspection versus “school quality” measured by examination performance data. While they are likely to be positively correlated to some extent, there will also always be some discrepancies.

### ***The value added of inspection***

If inspection outcomes are heavily influenced by data, one may question the value of sending inspectors into a school. After all, the judgement they reach will not be about what they observe, but will rather be based on background data drawn from other sources (e.g., school performance tables).<sup>2</sup> Of course, the real value added of school inspections is the rich, qualitative evidence that education experts – that is, inspectors – gather and evaluate about what life is like “on the ground”. In other words, inspection activities should add new information about a school’s education offering, rather than simply regurgitating what is already known from background data held.

### ***Inspections should draw on inspectors’ expertise***

Inspectors are highly trained education professionals, many of whom have spent years teaching in and leading schools. They are not – and should not be expected to be – expert statisticians. Data are often complex, with their interpretation and appropriate caveats often misunderstood (even by those with advanced quantitative training). Given inspectors’ specialist knowledge and skills as *education professionals*, it is these skills that they should be expected to use. Given their wealth of experience, inspectors are uniquely well placed to use their professional judgement to establish the quality of education that a school provides. But, if inspection outcomes are too closely tied to data and metrics, then inspectorates are unlikely to be making best use of the specialist knowledge and skills of its workforce.

### ***Many routinely available measures will only partially capture the effects of schools***

The discussion above focused on measures of school examination performance. But what about measures that, at face value, are perhaps easier to interpret, such as pupil absence rates? Unfortunately, these share many of the same limitations. For instance, pupil absence rates are only partially in the control of schools, with families and the home environment having a significant impact as well. If a data-driven inspectorate uses raw absence figures when making inspection judgements, it will be basing its decisions on something that the school has limited influence over, and thus not really capturing the effect of schools.<sup>3</sup> Moreover, as disadvantaged and low-achieving pupils tend to have higher absence rates (Department for Education, 2023) – again likely due to having more challenging home environments – such a measure would penalise schools serving disadvantaged communities. Yet the alternative – a contextualised absence measure that accounts for differences in pupil intakes – becomes (a) harder to interpret and (b) arguably engrains lower expectations of lower achieving, disadvantaged pupils of attending school. In other

words, the problem with most quantitative background measures that a data-driven inspectorate (and its inspectors) has access to do not purely capture the effects of schools.

### ***Small cohort sizes and stability over time***

A further practical challenge with using data in school inspections is that cohort sizes in a single year can be small, particularly in primary providers with one form entries. For some schools this will lead to their data being suppressed within published records (school performance tables), which – for a data-driven inspection regime – would make it difficult to judge such schools on a comparable basis with others. Yet even if data are available, limited cohort sizes – often 30 pupils or less in primary schools for a single year group – mean there can be quite a great deal of uncertainty surrounding the results. Small cohort problems can be compounded if one attempts to use more complex measures (e.g., contextual value added rather than raw achievement) or when attention turns to the performance of specific subgroups (e.g., the academic progress made by disadvantaged pupils). This can lead to overinterpretation of data, with a danger that inspectors focus on outlying metrics that could simply be driven by statistical noise. Similarly, limited school cohort sizes can lead to instability in key measures (e.g., school performance outcomes), which – if interpreted too literally – could lead to erroneous conclusions being drawn.

### ***The relationship between statistical significance and cohort (school) size***

A related challenge is that there is a mechanical relationship between “statistical significance” and school size. Put simply, in bigger schools there are more data, meaning that there is more power to detect statistically significant differences than in smaller schools. For instance, take two schools – one big and one small. They both have the same Progress 8 measure (e.g.,  $-0.25$ ), meaning that pupils made the same academic progress during secondary school that year. Yet the Department for Education’s Progress 8 bandings may put the large school into the “below average” category, while the small school is put into the “average” category (Department for Education, 2020). Why? Because bigger schools have more observations, leading to narrower confidence intervals that have less chance of crossing zero than smaller schools. Hence, attempts to simplify the interpretation of quantitative evidence for non-quantitative audiences can backfire. In the example above, it is likely to lead people to believe that the progress made by pupils in these two schools differs, when the progress they actually make is exactly the same.

### ***Many important things are not measured***

Although England is fortunate in the amount of quantitative data it holds about schools, there are still limits to the information available. Some important things are not measured, while others cannot be quantitatively measured at all (at least at scale). A prime example is pupils’ academic achievements at the end of primary school. Despite the presence of Key Stage 2 tests, there is no external measurement of knowledge and skills in subjects such as science, music, and geography. An entirely data-driven inspectorate would thus be

forced to focus on the narrow set of areas with information available (i.e., mathematics and English) and thus judging providers only partially on what they do. As noted above, this also risks a narrowing of the curriculum, with schools focusing on the areas where data are available. Being less reliant on data – and instead drawing on qualitative evidence to take a more holistic view of a school – means inspection judgements are based upon provision across a wider array of academic areas.

### ***Data could lead to prior beliefs about schools being formed***

We previously noted that a potential advantage of using data to inform inspections is that they will help inspectors to understand the background context of the school. However, the converse is that the data may lead inspectors to start forming views about a school and the quality of its provision before they have set foot through the door. The social psychological literature has discussed the issue of such “halo” and “horn” effects at length (Law, 2009), including how positive/negative views formed about one area (such as performance in national examinations) can bias one’s views about other unrelated areas. It is thus important that any pre-inspection data do not unduly influence inspectors’ assessment of other aspects of the school (e.g., the quality of its curriculum) and the ultimate judgement that they form. This would obviously be a challenge for a strongly data-driven inspection regime. In contrast, less data-driven systems of inspection can emphasise to inspectors that the data should only be used as a starting point, that their role is to help inspectors understand the background context of a school, and that they should be treated as complementary information to the evidence they are seeing for themselves on the ground.

### ***Non-response to primary data collections within inspections***

Although some inspectorates survey parents, teachers, and pupils during inspections, these tend to have a relatively short window for response. This is because they are part of the inspection evidence inspectors gather to inform their final judgement. Consequently, not everyone will have time to respond. Indeed, response rates to some of these surveys are quite low. Take Parent View in England (a parental questionnaire), for example. The average response rate across schools is less than 20%. In other words, although parents are invited to share their views, many choose to not do so. Hence the quantitative data captured in such surveys are unlikely to be fully representative of the views held across the wider school community. Moreover, response patterns could vary by school – for example, in School A the 20% most satisfied parents might respond to the parental questionnaire, while in School B it could be the 20% least satisfied. This clearly has the potential to complicate the interpretation of these data, including how they should be used by inspectors in their judgement of schools.

### ***A case study of England. How does Ofsted currently use quantitative data in its inspections to manage these trade-offs?***

Having set out the strengths and limitations of using quantitative data to inform school inspections, we now provide a short overview of how Ofsted currently uses such information. In doing so, we discuss how it attempts to manage the trade-off between the

useful information that such data can bring, while simultaneously trying to avoid the potential negative consequences set out above.

### ***The role of quantitative data in Ofsted's risk assessment***

Most schools previously judged as good are subject to Ofsted's risk assessment process (Ofsted, 2023b). This is conducted in two stages. In the first stage, Ofsted analyses school-level data (e.g., national examination performance, school workforce statistics, amongst many others) to predict the likelihood that a school has declined and will receive a lower judgement at its next inspection. This is then followed by a short desk-based review, where a Senior His Majesty's Inspector looks through the information and complements it with any insights they have from parental complaints, warning notices, and any other local intelligence gathered. The outcome of this two-stage process helps determine – along with the time since the last inspection and the type of the previous inspection – what type<sup>4</sup> of inspection it will receive next (see Ofsted, 2023b, for further details).

Quantitative background data are thus used by Ofsted as a “warning signal” of a possible decline in effectiveness, as discussed above. They thus play a role in how the inspectorate manages its finite resources, contributing towards decisions about how and when certain schools get inspected. Even then, they only form part of this process – being followed by a qualitative human review. In doing so, Ofsted attempts to use the background data available intelligently, using them to help set its priorities for inspection, but without directly informing the final judgement made.

### ***Inspectors are provided with background data prior to inspections via the inspection data summary report***

The second way Ofsted uses quantitative data is to provide inspectors with data about the schools they inspect. It does so via the inspection data summary report (IDSR; see Ofsted, 2022b, for further details). The IDSR provides inspectors with useful information to help them understand the context of the school, including its recent performance in national examinations, absences and exclusions, pupil movements, and the local/school context. Returning to the discussion above, its purpose is to help inspectors understand a school's context and background. In particular, the IDSR is designed to (a) draw inspectors' attention to possible valid inferences from the available data and (b) discourage them from making invalid inferences by using statistical filters (see below for further details). Importantly Ofsted (2022b) explicitly notes however that “the IDSR can only provide a starting point”, with inspectors wanting “to see first-hand the quality of education as experienced by pupils and understand how well leaders know what it is like to be a pupil at the school”.

In providing the IDSR, Ofsted is conscious of limiting the impact of any of the possible negative side effects previously discussed. For instance, data are often presented for the last 3 years where possible – not just the most recent year – so that outlying data points can be identified. The IDSR does not just provide information on examination results but also includes indicators of possible off-rolling and some kinds of curriculum narrowing. Moreover, rather than just providing inspectors with raw data, the IDSR includes a

written interpretation of the quantitative evidence to help those without a statistical background to digest the information available. Certain parts of the IDSR are also “greyed out” – written in light grey font to distinguish it from the main text – where the school is not significantly different from national averages or other key benchmarks. This helps ensure that quantitative information based on small sample sizes does not get over-interpreted by inspectors. This helps inspectors focus on the most relevant background data about the school. Inspectors are also trained into how to use the data included in the IDSR appropriately, with it again emphasised that such information should be used only as a starting point.

The discussion presented in previous sections recognises that there are both pros and cons of providing inspectors with the IDSR. Yet much of its content is already freely available in the public domain from, for instance, the Department for Education’s School Performance Tables. In a counterfactual world where Ofsted did not provide the IDSR to inspectors, they could access much of the same information for themselves if they so wished. It is therefore arguably better for all inspectors to be provided with the same information in a standardised way, and to help ensure that appropriate inferences are drawn from it. The alternative would likely be for different inspectors to draw upon different quantities and quality of quantitative information prior to inspections, potentially interpreting different measures in different ways. The provision of the IDSR to inspectors prior to inspections is an attempt to resolve such issues. The overarching purpose of doing so is to avoid inappropriate inferences being drawn.

### ***The role of data in directly informing inspection outcomes***

The introduction of the EIF in September 2019 represented a shift in Ofsted’s inspection framework (Ofsted, 2023a). Less weight is now put directly on examination results, with more attention given to the quality of the curriculum, how it is sequenced, and its delivery. This does not mean that examination outcomes are not an important part of what schools deliver, just that there is less focus on them in inspections than previously. As noted above, it is of course the fundamental job of schools to equip young people with the knowledge and skills they will need in later life. If children are emerging without these, it is only right that the inspectorate takes note. Quantitative data – in the form of school performance metrics – should be part of the mix to judge whether this is indeed the case. Such data are after all perhaps the most robust indicator of whether a school’s pupils are achieving at least as well as they should. Having pupils who achieve well – or are making sufficient progress given differences in their starting points – should thus be seen as a necessary (if not sufficient) condition for what constitutes a Good or Outstanding school.

To reflect this, the quality of education criteria within the EIF framework note how “learners develop detailed knowledge and skills across the curriculum and, as a result, achieve well. Where relevant, this is reflected in results from national tests and examinations that meet government expectations, or in the qualifications obtained” (Ofsted, 2023a). The Ofsted (2022a) annual report hence illustrates how inspection outcomes continue to be correlated with school performance data. Yet there are also schools that despite strong performance data are not judged to be Good or Outstanding by Ofsted,

and vice versa. This illustrates how a Good school must rely on more than just strong data and metrics alone.

### ***Limiting attempts to game the system***

As noted previously, using data as part of inspection processes can have unintended and undesirable consequences. One is that schools may seek to artificially inflate their test scores in ways that are not in the best interest of their pupils. Examples include off-rolling pupils and disproportionately entering pupils into certain subjects and examinations that they perceive to be relatively easier than others. Ofsted recognises this issue, and uses the data it has available to explore unusual patterns of pupils leaving schools that may be suggestive of off-rolling, and unusual patterns of examination entries. Again, Ofsted uses these data as a starting point to help inform conversations between inspectors and schools.

### ***Surveys with parents, teachers, and schools***

Inspectors are tasked with gathering a wide array of evidence about schools. Although much of this is qualitative, some primary quantitative data are gathered as well. In particular, pupils, teachers, and parents are asked to complete a short questionnaire offering their views. This incorporates a wide variety of topics surrounding the school environment, including about pupil behaviour, teacher workload, the quality of information provision to parents, and their overall satisfaction levels. Together, these aim to satisfy the advantage of gaining a wide range of perspectives about a school. In doing so, it is of course recognised that selective non-response to such surveys is likely to limit their representativeness. They are, nevertheless, likely to provide a more generalisable picture of how key stakeholders view a school than if qualitative evidence were gathered alone.

### **Conclusions**

Inspections are a key part of school accountability in many countries, including England. These typically involve a team of experienced education professionals gathering qualitative evidence about a school to form a judgement about the quality of its provision. However, such inspections can – and do – draw on quantitative data as well. Although this mostly encompasses routinely collected administrative information – for example, pupils' performance in examinations, absence rates, and staff turnover – it also includes surveys conducted with pupils, parents, and teachers. This means that England is uniquely rich in terms of the range of quantitative data that school inspectors can potentially draw on.

Precisely because many quantitative data are available, great care must be taken in how these data are used and interpreted. School inspectorates must draw out the useful information in available data while appreciating – and trying to avoid – their limitations. The aim of this paper has been to discuss the pros and cons of using such quantitative information in school inspections, including the trade-offs that must be made.

We have noted that a strength of quantitative data – such as examination results – is that they provide important information about pupils' outcomes that will have



consequences for the rest of their lives. Ensuring that young people acquire knowledge and skills is the fundamental purpose of schools, and it is only right that a school inspectorate takes note where this function is not being performed sufficiently well. As quantitative data provide perhaps the most objective measures of a young person's achievements, it is important that this has some role in the inspection process. Moreover, drawing on objective quantitative data may help consistency in inspection outcomes between inspectors, while also providing inspectors with key background context about the school.

Yet an overreliance on quantitative data also carries dangers. An inspection regime too heavily tied to school performance measures is likely to increase the stakes of national assessments, potentially putting more pressure on schools and staff as a result. This can lead to unintended negative consequences for other parts of the education system – including a narrowing of the curriculum, with schools/teachers having the perverse incentive to “teach to the test”. Indeed, if inspection judgements are too closely tied to such data, the added value of sending inspectors into schools – having a team of experienced education professionals using their expert judgement to assess the quality of current curriculum and teaching and their impact as well as other aspects of school life – is greatly reduced. On the more technical side, the quantitative measures available will only partly reflect school “quality”, with other inputs (such as parents) influencing the metrics as well. It is consequently not even a straightforward decision about which basket of measures to use, particularly given the challenges in appropriate interpretation by non-quantitative audiences. This is compounded by issues such as small sample sizes, non-response/missing/suppressed information, and a degree of instability in the data over time. It may also lead to inspectors developing prior beliefs about the quality of a school, which may then not be straightforward to shift.

School inspectorates must manage these trade-offs. This is not an easy task. They should of course try to maximise the benefits of having such a rich array of quantitative data available, but without suffering the negative consequences that an overreliance on such data bring. The balance of these trade-offs differs according to the purpose of the quantitative data used (e.g., to select “high-risk” schools for inspection, to inform inspectors before the visit, as a basis for the verdict of the inspection). In England, Ofsted does so by trying to use the quantitative data available in a proportionate way. Most notably, quantitative data play a role in deciding when and how some schools are next inspected, but only following further qualitative input from regional inspection teams. Pupil outcomes are judged as part of Ofsted's quality of education criteria, but are just one element of it. Views from all parents, pupils, and teachers are sought by Ofsted via questionnaires, and are triangulated with the other information that inspectors gather. And, while quantitative contextual information is provided to inspectors as part of Ofsted's pre-inspection report, it is one of several inputs and strands of information to an inspection.

Hence the use of data in school inspections is clearly a balancing act. There is no single “right” or “wrong” answer to how data are used. Rather, it is a judgement call that needs to be made based on a detailed understanding of the trade-offs it involves. In documenting the strengths and limitations with using quantitative data for this purpose, it is hoped that this paper will facilitate further discussion about this important issue.

## Notes

1. Inspections, in contrast, suffer from a different type of time lag. At the time they are conducted, they provide a snapshot about the quality of provision across all year groups, and thus provide a very up-to-date view. However, as inspections are comparatively infrequent, the information they provide becomes dated over time.
2. The Ofsted 2021/2022 report notes that in 2021/22 the correlation between Ofsted judgements and Progress 8 scores was +0.46. The analogous correlation between Ofsted judgements and Key Stage 2 scores was +0.30 (see Ofsted, 2022a, Footnote 25).
3. The Education Endowment Foundation (2022) conducted a recent review of the evidence for attendance interventions. They found that the “overall quality of evidence is weak” (p. 3) and that many interventions do not have “sufficient evidence to reach a conclusion on effectiveness” (p. 4).
4. Whether it is a “full” Section 5 inspection or a “reduced-tariff” Section 8 inspection.

## Acknowledgements

John Jerrim has co-authored this paper during his part-time secondment to Ofsted from UCL.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

**John Jerrim** is a Professor of Education and Social Statistics at UCL. He has a long-standing interest in policy evaluations, having led a number of randomised controlled trials funded by the Education Endowment Foundation. He has previously led work comparing levels of financial literacy across countries, as well as comparing socioeconomic inequality in young people’s financial capability in the UK. John has sat on a number of expert and advisory groups, including with Ofqual, the Education Endowment Foundation, and the Department for Education. He is currently a special advisor to Ofsted focusing on their research programme. With almost 100 academic publications, John’s research has been disseminated widely, including coverage by national and international media, such as *The Economist*, BBC Breakfast, *The New York Times*, and Sky News.

**Alex Jones** was appointed as Ofsted’s Director, Insights and Research in March 2022. He was previously Head of Evidence and Transformation at 10 Downing Street from 2020 to 2022, following 10 years working at the Department for International Development in roles in the UK, Asia, and Africa. Before that Alex worked across a number of roles in central and local government, including for Devon Social Services, supporting children and young adults with disabilities to live independently.

## References

- Allen, R., & Burgess, S. (2012). *How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England* (CMPO Working Paper Series No. 12/287). Centre for Market and Public Organisation. <https://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/wp287.pdf>
- Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, 26(1), 32–56. <https://doi.org/10.1080/09243453.2014.927369>

- Bokhove, C., Jerrim, J., & Sims, S. (2023a). Are some school inspectors more lenient than others? *School Effectiveness and School Improvement*, 34(4), 419–441. <https://doi.org/10.1080/09243453.2023.2240318>
- Bokhove, C., Jerrim, J., & Sims, S. (2023b). *Do schools that employ an inspector get better inspection grades?* EdArXiv. <https://doi.org/10.35542/osf.io/2bj3q>
- Czech School Inspectorate. (2015). *The Czech School Inspectorate*. <https://www.sici-inspectorates.eu/getattachment/a68cf441-2be9-48c0-b0ed-49fac91d6719/CSI-final-profile-June-2015.pdf;jpg.aspx>
- Department for Education. (2020). *Progress scores for key stage 4: School and college performance tables*. [https://dera.ioe.ac.uk/id/eprint/34977/1/Key\\_stage\\_4\\_progress\\_banding\\_calculations\\_bandings\\_2019.pdf](https://dera.ioe.ac.uk/id/eprint/34977/1/Key_stage_4_progress_banding_calculations_bandings_2019.pdf)
- Department for Education. (2023). *Pupil absence in schools in England: Academic year 2021/22*. <https://explore-education-statistics.service.gov.uk/find-statistics/pupil-absence-in-schools-in-england/2021-22>
- Education Endowment Foundation. (2022). *Attendance interventions rapid evidence assessment*. <https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/attendance-interventions-rapid-evidence-assessment>
- Ehren, M. C. M., Gustafsson, J. E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, 51(3), 375–400. <https://doi.org/10.1080/03050068.2015.1045769>
- Gaertner, H., Wurster, S., & Pant, H. A. (2014). The effect of school inspections on school improvement. *School Effectiveness and School Improvement*, 25(4), 489–508. <https://doi.org/10.1080/09243453.2013.811089>
- Gorard, S. (2010). Serious doubts about school effectiveness. *British Educational Research Journal*, 36(5), 745–766. <https://doi.org/10.1080/01411920903144251>
- Hodge, L., Little, A., & Weldon, M. (2021). *GCSE attainment and lifetime earnings*. Department for Education. [https://assets.publishing.service.gov.uk/media/60c36f0cd3bf7f4bd11a2326/GCSE\\_Attainment\\_and\\_Lifetime\\_Earnings\\_PDF3A.pdf](https://assets.publishing.service.gov.uk/media/60c36f0cd3bf7f4bd11a2326/GCSE_Attainment_and_Lifetime_Earnings_PDF3A.pdf)
- Hutchinson, J. (2016). *School inspection in England: Is there room to improve?* Education Policy Institute. <https://epi.org.uk/wp-content/uploads/2018/01/school-inspection-in-england-web.pdf>
- Inspectie van het Onderwijs. (2023). *Inspection framework 2021 for the supervision of secondary education*. <https://acrobat.adobe.com/id/urn:aaid:sc:EU:ef00fac4-ec75-4d0f-80ed-8262e6b0ae64?viewer%21megaVerb=group-discover>
- Jones, K., & Tymms, P. (2014). Ofsted's role in promoting school improvement: The mechanisms of the school inspection system in England. *Oxford Review of Education*, 40(3), 315–330. <https://doi.org/10.1080/03054985.2014.911726>
- Koretz, D. (2008). *Measuring up. What educational testing really tells us*. Harvard University Press.
- Law, J. (2009). Horns and halo effect. In *A dictionary of business and management* (5th ed.). Oxford University Press. <https://www.oxfordreference.com/view/10.1093/acref/9780199234899.001.0001/acref-9780199234899-e-3082>
- Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'. *British Educational Research Journal*, 43(2), 193–212. <https://doi.org/10.1002/berj.3264>
- Munoz-Chereau, B., Hutchinson, J., & Ehren, M. (2022). 'Stuck' schools: Can below good Ofsted inspections prevent sustainable improvement? [https://epi.org.uk/wp-content/uploads/2022/06/Final\\_report\\_stuck\\_schools.pdf](https://epi.org.uk/wp-content/uploads/2022/06/Final_report_stuck_schools.pdf)
- Norwegian Directorate for Education and Training. (2015). *Inspection of kindergartens and education in Norway*. <https://www.sici-inspectorates.eu/getattachment/b1fe7b1a-0d84-4511-9694-31b0cef8925e/Inspection-of-kindergartens-and-education-in-Norway-04052015.pdf;jpg.aspx>
- Nye, P. (2018, November 16). *Some MATs look to have been hit hard by the withdrawal of ECDL*. education datalab. <https://ffteducationdatalab.org.uk/2018/11/some-mats-look-to-have-been-hit-hard-by-the-withdrawal-of-eecd/>
- Nye, P., & Rollett, S. (2017, October 11). *How do Ofsted ratings relate to Progress 8 scores?* education datalab. <https://ffteducationdatalab.org.uk/2017/10/how-do-ofsted-ratings-relate-to-progress-8-scores/>

- Office for Standards in Education, Children's Services and Skills. (2022a). *The annual report of His Majesty's Chief Inspector of Education, Children's Services and Skills 2021/22*. <https://www.gov.uk/government/publications/ofsted-annual-report-202122-education-childrens-services-and-skills/the-annual-report-of-his-majestys-chief-inspector-of-education-childrens-services-and-skills-202122#fn:25>
- Office for Standards in Education, Children's Services and Skills. (2022b). *School inspection data summary report (IDSr) guide*. <https://www.gov.uk/guidance/school-inspection-data-summary-report-idsr-guide#report-information>
- Office for Standards in Education, Children's Services and Skills. (2023a). *Education inspection framework*. <https://www.gov.uk/government/publications/education-inspection-framework/education-inspection-framework-for-september-2023>
- Office for Standards in Education, Children's Services and Skills. (2023b). *Risk assessment methodology for good state-funded schools*. <https://www.gov.uk/government/publications/risk-assessment-methodology-for-maintain-schools-and-academies/risk-assessment-methodology-for-good-state-funded-schools>
- Onderwijsinspectie. (n.d). *Inspectie 2.0 op maat* [Customized inspection 2.0]. Retrieved March 22, 2024, from <https://www.onderwijsinspectie.be/nl/doorlichten/inspectie-20-op-maat>
- Penninckx, M., Vanhoof, J., De Maeyer, S., & Van Petegem, P. (2016). Explaining effects and side effects of school inspections: A path analysis. *School Effectiveness and School Improvement*, 27(3), 333–347. <https://doi.org/10.1080/09243453.2015.1085421>
- Prior, L., Jerrim, J., Thomson, D., & Leckie, G. (2021). A review and evaluation of secondary school accountability in England: Statistical strengths, weaknesses and challenges for 'Progress 8' raised by COVID-19. *Review of Education*, 9(3), Article e3299. <https://doi.org/10.1002/rev3.3299>
- Quesel, C., Schweinberger, K., & Möser, G. (2021). Responses to positive and negative feedback on organizational aspects of school quality: Teachers' and leaders' views on a Swiss traffic light approach to school inspection. *School Effectiveness and School Improvement*, 32(3), 345–362. <https://doi.org/10.1080/09243453.2020.1856886>
- Quintelier, A., De Maeyer, S., & Vanhoof, J. (2020). Determinants of teachers' feedback acceptance during a school inspection visit. *School Effectiveness and School Improvement*, 31(4), 529–547. <https://doi.org/10.1080/09243453.2020.1750432>
- Skolinspektionen. (2024). *Anledning till inspektion* [Reason for inspection]. <https://www.skolinspektionen.se/inspektion/inspektion-steg-for-steg/infor-inspektion/urval-for-inspektion/#ReadMoreBlock-1279>
- Teacher Tapp. (2019, May 28). *SATs reform: 50 shades of playing the system (and how to reform it)*. <https://teachertapp.co.uk/articles/sats-reform-change-future-teacher-assessment-reliable/>
- von Stumm, S., Smith-Woolley, E., Cheesman, R., Pingault, J.-B., Asbury, K., Dale, P. S., Allen, R., Kovas, Y., & Plomin, R. (2021). School quality ratings are weak predictors of students' achievement and well-being. *Journal of Child Psychology and Psychiatry*, 62(3), 339–348. <https://doi.org/10.1111/jcpp.13276>