

A response to Ofsted's work with schools inquiry

Highlights

- There is evidence that inspectors with different characteristics reach different judgements, especially in primary schools for contract type (HMI versus OI) and gender.
- However, not enough research on Ofsted is done. There can be more transparency and data made available by Ofsted to facilitate more research.
- The argument that inspection reports are useful for school choice needs to be revised.
- There is a tension between two competing functions of an inspectorate: accountability and school improvement.
- There is a trade-off between consistency in judgements and the stakes: the higher the stakes, the more problematic unreliable judgements are.

We, as three academics from esteemed institutions, the University of Southampton (Christian Bokhove) and UCL (John Jerrim and Sam Sims), are experts in the field of education research, particularly specialising in quantitative education research. Over the past year, we have been conducting a research study into Ofsted inspections, including an extensive project supported by the Nuffield Foundation, focused on investigating the consistency of such inspections. In addition to our oral contribution to the Education Committee, the purpose in submitting this written evidence is to emphasise the significance of basing government decisions concerning school inspections on well-grounded research. We firmly believe that this research should be conducted by impartial and genuinely independent entities. Recognising the sensitivity surrounding Ofsted, we are aware that it is a subject that attracts substantial criticism within the education community, including among researchers in the field. Consequently, we hold reservations regarding potential bias that may infiltrate the evidence base pertaining to Ofsted. It is plausible that certain members of the education research community might be inclined to produce negative evidence about the organization in order to substantiate their own viewpoints.

Contrary to such practices, we firmly oppose any form of bias and remain committed to providing impartial and independent evidence. Our objective is to address some of the key criticisms levelled against Ofsted and its inspections, ensuring that our findings contribute to a balanced and comprehensive understanding. Furthermore, as we will highlight below as well, there is very little research on the consistency and reliability of Ofsted inspections. We pick up some of the specific points made in the call for evidence. In doing so, we restrict our focus to what the latest evidence we have produced has to say about these issues.

- The usefulness of Ofsted inspections and inspection reports, and whether inspections are carried out in sufficient depth to meet the expectations of schools, governors and parents.

The usefulness of Ofsted inspections for parents when choosing a school

A commonly asserted claim about Ofsted inspections is that they assist parents in making informed choices when selecting schools for their children. Ofsted themselves have put forth arguments supporting this notion, and surveys conducted by Ofsted indicate that parents do utilise these inspections, with some considering them to be "decisive."

However, the issue of the usefulness of Ofsted inspection reports for parents in school selection is thoroughly examined in a comprehensive study by Bokhove, Jerrim, and Sims (2023a). The study highlights several significant problems with the argument that Ofsted inspection reports are beneficial to parents during the school selection process. These problems include:

1. Time lags: There are often substantial delays between the time an inspection is conducted and when a parent's child actually attends the school. This issue has been exacerbated over the past decade due to the exemption of Outstanding schools from routine inspection.
2. Sibling enrolment: Parents typically select a school for their older child and tend to send their younger child to the same school to avoid the inconvenience of managing multiple school runs. This exacerbates the time lag problem mentioned in the point above.
3. Incomparable data: The data presented in inspection reports are unlikely to be directly comparable across different schools. This is because inspections may have been conducted under different frameworks, in different forms (e.g., graded versus ungraded inspections), and by different inspection teams.
4. Limited choice set: When parents are choosing a school, they often have a realistic choice between only two or three options. These schools may not significantly differ in terms of their Ofsted judgments.
5. Availability of alternative information: Parents have access to other sources of information that are likely to be more up-to-date, such as performance in national examinations, absence rates, and student demographics. The added value of Ofsted judgments beyond this information is not clear.

The study further demonstrates that the likelihood of a change in headteachers between the time a parent selects a school and the majority of their child's attendance is high. Moreover, it illustrates that the Ofsted judgment available to parents during the school selection phase has only a weak correlation with the school-level outcomes when their child actually attends the school.

Overall, the study by Bokhove, Jerrim, and Sims (2023a) raises significant concerns about the usefulness of Ofsted inspection reports for parents in the context of choosing a school. The study's findings suggest that parents may benefit more from considering additional factors when making decisions about their children's education.

At the same time, we feel it is important to note that the fact Ofsted judgements may not be that useful in informing school choice may not be a problem per se. We make this point explicitly in the conclusion to our paper, where we stress the following (<https://www.tandfonline.com/doi/full/10.1080/15582159.2023.2169813>):

"In our view, they do not illustrate problems with Ofsted inspections per se. In line with what many would expect from a school inspectorate, empirical evidence suggests that Ofsted inspections resulting in an "inadequate" judgment improve the outcomes of such schools (Hussain, 2017). Given this, we should not expect or even want Ofsted judgments to predict what such a school will be like in the medium term. Thus, the fact that once inadequate and requiring improvement schools become largely indistinguishable from their good counterparts may be a sign of success. Moreover, Ofsted judgments are likely to be of interest to parents whose children are currently attending a school, providing them with important information about what it is (and is not) currently doing well. Rather, our results provide a cautionary tale about using Ofsted judgments to inform school choice. Our advice to parents is to not place too much emphasis on them. While they may act as a catalyst for thinking about differences between schools, and perhaps some insight into a school's ethos, they are not going to provide much information about the academic environment and the outcomes of pupils during the period when their children will be going there."

This comes back to the point of what the role of Ofsted really is. As we also state in our paper *"there is a trade-off between inspections stimulating school improvement and providing useful information to parents when they are selecting schools – they cannot easily do both.* It is important to distinguish the sometimes contradictory functions of accountability and school improvement. When it comes to accountability it also seems relevant to distinguish safeguarding from schools' learning and teaching functions.

How useful Ofsted inspections and inspection reports are is closely linked to their consistency and reliability

The assessment of the usefulness of Ofsted inspections and inspection reports heavily relies on their consistency and reliability. However, the available evidence on this matter is limited, highlighting the need for further investigation. We highlight our recent independent study conducted by Bokhove, Jerrim, and Sims (2023b) on this topic, distinct from studies conducted by Ofsted themselves (a peer-reviewed version is now in press).

The aforementioned study sheds light on a few noteworthy findings. Firstly, it identifies small differences in Ofsted judgments between male and female inspectors (males being more lenient than females), although these differences were only detectable in primary school inspections, possibly due to the limited sample size for secondary schools. These discrepancies were observed in both graded and ungraded inspections, with the most notable variation found in Inadequate judgments. Additionally, the research highlights variations in judgments between Her Majesty's Inspectors (HMIs) and Ofsted inspectors (OIs). However, the study also underscores that HMIs and OIs are typically assigned to different inspection tasks, potentially influencing their respective judgments. The study found no differences in the inspection grades awarded by inspectors with more or less experience, working inside or outside of home region, or when inspecting schools within their primary/secondary school specialism. Late 2023 the Nuffield project will also explore sequence effects of inspections (for example, do inspectors become harsher or more lenient after a series of inspections) and text analysis of the inspection reports, which builds on methodological work with Ofsted reports.

While the study by Bokhove, Jerrim, and Sims (2023b) offers some insight into the matter, it should be viewed as merely the initial step in understanding the issue. The authors relied on data extracted directly from inspection reports using natural language processing techniques. It is reasonable to assume that Ofsted holds more comprehensive data, enabling a more thorough investigation of these concerns. Moreover, Ofsted possesses more recent data, considering that the study by Bokhove, Jerrim, and Sims utilised data up until 2019, preceding the implementation of the Education Inspection Framework (EIF).

Consequently, obtaining a detailed understanding of the usefulness of school inspections necessitates a comprehensive examination of their consistency and reliability. Regrettably, the current body of evidence in this area remains remarkably thin, emphasising the urgency to expand and deepen the available evidence base on this crucial aspect. We would also note that although 1-to-1 policy 'transplant' from different international contexts is not advisable, we can of course learn from other systems and structures. As mentioned before, issues of consistency also relate to the function of inspections, accountability and/or school improvement, and as such the length of inspections will almost certainly lead to a trade-off between all these different factors.

Whether inspections are carried out in sufficient depth to meet the expectations of schools, governors and parents.

It is important to address whether inspections are conducted with sufficient depth to meet the expectations of schools, governors, and parents. In response to this inquiry, the following points can be made.

Firstly, it is crucial to acknowledge that schools, teachers, and governors may have unrealistic expectations regarding the precision of inspection outcomes. Inspections are inherently a human process, which means that a degree of subjectivity and personal judgment will always be involved. It is unrealistic to expect complete consistency among different inspectors, even within the same inspector on different days.

The key question, then, becomes whether inspection judgments are consistently and reliably applied. This directly depends on how these judgments are utilised and the consequences associated with them. For instance, the distinction between an "Inadequate" and "Requires Improvement" rating carries significant implications. A headteacher is likely to face job loss if their school receives an "Inadequate" grade. Therefore, it is of utmost importance that these judgments are consistently and reliably applied to ensure fairness.

On the other hand, one could argue that a certain level of unreliability or inconsistency may be more tolerable when it comes to distinguishing between "Outstanding" and "Good" ratings. While there are no immediate direct consequences for schools in this distinction, it should be recognised that such variations could impact future pupil numbers and school choices made by parents who do use Ofsted ratings (despite the issues with doing so, as we note above). Additionally, inconsistent judgments may have an impact on staff morale within the school.

In summary, while it is unrealistic to expect complete uniformity in inspection outcomes due to the human nature of the process, the focus should be on ensuring that inspection

judgments are as consistent and reliable enough for the purpose of their use. It will hence be particularly important for reliability and consistency to be very high when the stakes are also high (e.g. the Inadequate judgement).

The effectiveness of Ofsted's complaints procedure and the extent to which Ofsted is accountable and transparent in its work.

In our submission, we would like to address the issue of transparency regarding the consistency and reliability of Ofsted inspections. It is our belief that Ofsted should take further steps to enable independent scrutiny of the reliability and consistency of its inspection outcomes.

Currently, there is a dearth of evidence on this matter, both within England and in other countries. It is imperative that Ofsted invest more effort in this area, including making the data it possesses available for independent scrutiny. For example, Ofsted possesses valuable and comprehensive data on the allocation of inspectors to inspections. This data includes crucial information, such as the inspection experience of the lead inspector and their team, which has been explored in studies by Bokhove, Jerrim, and Sims (2023b). Investigating these characteristics, alongside others like the inspectors' own backgrounds (such as whether they have ever led a challenging school themselves), holds immense potential for future examination.

To facilitate an in-depth analysis of inspection outcomes under the Education Inspection Framework (EIF), Ofsted should release this data to independent researchers who can produce evidence based on their findings. Furthermore, it is essential for Ofsted to provide greater clarity on the process of assigning inspectors to inspections and inspection teams. By doing so, they can enhance transparency and ensure a thorough understanding of the inspection process.

References

Bokhove, C., Jerrim, J., & Sims, S. (2023a) How Useful are Ofsted Inspection judgements for Informing Secondary School Choice?, *Journal of School Choice*, 17(1), 35-61, <https://doi.org/10.1080/15582159.2023.2169813>

Bokhove, C., Jerrim, J., & Sims, S. (2023b). *Are some school inspectors more lenient than others?*, CEPEO Working Paper Series 23-03, UCL Centre for Education Policy and Equalising Opportunities.

Bokhove, C., Jerrim, J., & Sims, S. (in press). Are some school inspectors more lenient than others? *School Effectiveness and School Improvement*