

## FAQs for Are some school inspectors more lenient than others?

### 1. What can we confidently say based upon the results?

In primary schools, inspections led by women have received slightly less generous judgements than inspections led by a man. 36.4% of primary inspections led by a woman led to a “requires improvement” or “inadequate”, compared to 33.1% of primary inspections led by a man. There was no detected gender difference for secondary schools.

The other major difference is with respect to the employment relationship inspectors have with Ofsted. In both primary schools and secondary schools, inspections conducted by “Her Majesty’s Inspectors” (HMI) received lower judgements than those conducted by Ofsted Inspectors (OIs). For instance, 44% of primary school inspections led by an HMI were awarded a “requires improvement” or “inadequate” grade, compared to 32% led by an OI. See FAQ2 below further details on the difference between HMIs and OIs.

Inspection team size was also found to be linked to inspection outcomes.

### 2. What is the difference between an “Ofsted Inspector” (OI) and a “Her Majesty’s Inspector” (HMI)?

In simple terms, HMIs are civil servants who hold a permanent, salaried position with Ofsted, most of whom are employed on a full-time basis. Ofsted are their sole employer.

OIs on the other hand work for Ofsted on a freelance basis and are paid per day(s) worked. Many OIs also hold other jobs within the education sector, including as senior school leaders.

### 3. What characteristics of inspectors were not associated with inspection outcomes?

No link was found with inspector experience and whether the inspector was conducting an inspection outside of their home region. The association with whether the inspector was a primary or secondary specialist was relatively weak.

### 4. What should not be concluded from the research?

Although some may be tempted to interpret our results as illustrating “bias” between different inspectors, we do not believe that this conclusion can be drawn from the evidence that we present. Ofsted inspectors are instructed to use their professional judgement when reaching their judgements, and it could be this difference in professional opinions that is driving our results. Moreover, for some our findings, we cannot rule out the possibility that Ofsted deploys particular inspectors (or particular types of inspectors) to what they consider to be the most challenging inspections. See FAQ 5 below for further discussion.

5. Could the findings just be due to Ofsted assigning particular inspectors to certain inspections?

The statistical methods we have used are unable to completely rule this possibility out. We can offer our views, however, as to whether we feel this is a credible explanation for our results.

In terms of gender differences in primary inspections, we do not believe differences in inspector deployment can explain our findings. Very similar patterns – of inspections led by women leading to slightly worse outcome than those led by men – emerged across inspection type, academic years and whether they are employed as an HMI or OI. The association is also unchanged after we have attempted to control for differences in the background characteristics of the schools being inspected.

Inspector deployment may, however, at least partially explain the difference in outcomes we observe between HMIs and OIs. As Table 6 of the paper illustrates, HMIs are more likely to be assigned to tougher inspections (e.g. to those schools that were graded as inadequate previously and/or with worse outcomes in national examinations). We do still continue to see sizeable differences between HMIs and OIs even when we try to take into account such differences in inspector deployment within our analysis. But we have probably only partially been able to control for such differences; there are likely to be decisions Ofsted makes about inspector deployment that we cannot adequately control for in our analysis (e.g. where there are safeguarding concerns or parental complaints). Our view is hence that there is probably a genuine difference in outcomes between HMIs and OIs that is not simply due to the former being assigned more challenging inspections than the latter, though it's probably a bit smaller than our estimates suggest.

The same likely holds true with respect to the difference we find for inspection team size.

6. What do you think Ofsted should do about the results?

We have recommended Ofsted do three things.

First, given that we find the biggest differences in awarding the inadequate judgement – which also has the greatest consequences for schools – Ofsted should publish further details on the quality assurance that is undertaken when such a judgement is reached. Currently, section 13 of the Education Act 2005 (<https://www.legislation.gov.uk/ukpga/2005/18/section/13>) notes that the report will be sent to the governing body or proprietor of the school, with Ofsted then considering “*any comments on the draft .....within the prescribed period*”. Further clarity on what this means in terms of quality assurance of the findings would help further reassure the sector about how many stages of detailed scrutiny inspection evidence goes through before such a judgement is reached.

Second, we recommend that Ofsted publishes more details about how inspectors are deployed to different inspections. This will further help with the interpretation of our results, and help schools to understand how conflicts of interests are managed, and the reason(s) why a particular inspector may be assigned.

Finally, it's clearly important that more independent work is done in this area. The big barrier to it happening is data access. We therefore suggest that Ofsted should deposit in the Office for National Statistics Secure Research Service an inspection-inspector linked dataset, facilitating further independent research into school inspections.

7. Why do you find stronger evidence (and bigger effects) for primary school inspections than for secondary school inspections?

We are not entirely sure but can offer some suggestions.

First, secondary school inspections tend to involve more inspectors (larger team sizes). This could drown the influence of the lead inspector out.

Second, much more background data about school performance is available for secondary schools than primary schools. During the time period our data are from, Ofsted was criticised for being too focused upon data/metrics, and these were having an undue influence on inspection outcomes. This may have been particularly true for secondary schools – leaving less room for there to be variation in outcomes according to professional judgement (and thus a weaker relationship with lead inspector characteristics).

Third, it may be a sample size issue. Statistically, we cannot rule out there being the same differences in secondary schools as we observe for primary schools. It is just that the number of secondary school inspections within our data (around 5,000) is so much smaller than the number of secondary inspections (around 23,000) that we do not have the ability to detect the same effects.

Finally, of course, it may just be that lead inspector characteristics matter when making judgements about primary schools than secondary schools.

8. Your research focused upon primary and secondary schools. What about Ofsted's other inspection remits (e.g. early years, further education, social care, initial teacher education).

We had to limit our scope somehow – bringing together all the data was a huge amount of work for just these two remits – and so chose to focus upon just schools to begin with.

However, having completed our analysis, we believe there would be challenges with conducting the same analysis for some of Ofsted's other remits. Given what we say about sample size issues for secondary schools in FAQ7 above, an analysis of further education providers may end up being underpowered. Likewise, there might be similar challenges in looking at variation in outcomes of social care providers and initial teacher education by inspector characteristics. We thus believe that Early Years might be the most fruitful additional remit to investigate, were this piece of work to be extended.

9. Where did you get your data from?

We used two data sources. First, there is a website called "Watchsted" that includes all the details on the inspections completed by each Ofsted inspector since 2011 (<https://perspective.angelsolutions.co.uk/Perspective/LiteUsers/Ofsted/InspectorSearch.aspx>). They have very kindly agreed to share their data with us to help us with this work.

Second, we have downloaded all inspection reports published since 2011 from the Ofsted website, converted these from .pdf to .txt format, and then written some Python code to strip out from the reports the name of the inspector(s) who conducted the inspections.

These data have then been merged with information Ofsted has published on all inspections that took place between 2005 and 2019.

10. How have you quality assured the data?

Appendix C of the paper goes into detail about how we have quality assured the data we have used. In brief, we have manually checked the names of inspectors in 150 short inspections and 150 full inspections, to see how these compare to the names we have recorded in our database.

For both full and short inspections, we find that the name of the lead inspector matches between our manual checks and our database on 97% of occasions.

11. What was the sample size?

For primary schools, our headline analysis is based upon 22,754 inspections conducted by 983 inspectors between September 2011 and August 2019.

For secondary schools, our headline analysis is based upon 5,024 inspections conducted by 586 inspectors between September 2011 and August 2019.

12. The data you used were based upon inspections conducted between 2011 and 2019. This was before Ofsted changed its inspection framework in September 2019. How do you think this might impact your results?

With COVID hitting the UK in March 2020, the EIF had only been in place for six months. Over the following two years, inspections were either halted, conducted remotely or were subject to a large number of deferrals.

Given this context, we decided to focus our efforts upon the pre-COVID, pre-EIF period where we felt we would have most confidence in the results.

It is of course difficult to say for certain how our findings might differ under a different inspection framework.

However, with the EIF increasingly moving away from using data/metrics to judge schools and more towards professional judgements being made about curriculum design, sequencing and implementation, the role of trusting the professional judgement of inspectors has arguably increased. Our feeling is that the variation we observe across inspector characteristics is therefore unlikely to have diminished with the introduction of the EIF. But this is of course an empirical question which, currently, we do not believe there is sufficient data to robustly answer.