


UNIVERSITY OF
Southampton

INSPECTING THE INSPECTORATE: NEW INSIGHTS INTO OFSTED INSPECTIONS

Press presentation
'Are some school inspectors more
lenient than others?'
26 January 2023
(embargoed until 2-Feb 00:01h)
Project funded by the Nuffield Foundation



1



Christian Bokhove John Jerrim Sam Sims

2

Press release

Press release

Under embargo until Thursday 2nd February 2023, 00:01 UK time.

Male school inspectors award more lenient grades than female inspectors

Male school inspectors are more likely to award higher Ofsted grades than female inspectors when inspecting similar primary schools, according to new research by the University of Southampton and UCL.

Released today as an academic working paper, the study explores the Ofsted grades awarded by 1,376 different inspectors across 35,751 school inspections conducted between 2012 and 2019. It is the first piece of independent research into how school inspection outcomes are linked to characteristics of the lead inspector, either in England or worldwide.

The study shows that male lead inspectors awarded a low grade ('Requires Improvement' or 'Inadequate') to around a third (33.1%) of primary schools. By contrast, female lead inspectors awarded a low grade to 36.4% of primary schools.

<http://is.gd/nuffieldofsted> (embargoed until 2/2/23)

3

Working paper

Are some school inspectors more lenient than others?

Christian Bokhove (University of Southampton)
John Jerrim (UCL Social Research Institute)
Sam Sims (UCL Centre for Education Policy and Equalising Opportunities)
Author list alphabetical. All joint first authors.

February 2023

- Working paper
- Five blogs
- FAQ

A few findings from the
paper/blogs in the next
few slides.

School inspections are a common feature of education systems across the world. These involve trained professionals visiting schools and reaching a high-stakes judgement about the quality of education they provide. By their nature, school inspections rely upon professional judgement, with different inspectors potentially putting more emphasis on certain areas than others. Yet there is currently little academic evidence investigating the consistency of school inspections, including how judgements vary across inspectors with different characteristics. We present new empirical evidence on this matter, drawing upon data from more than 30,000 school inspections conducted in England between 2011 and 2019. Male inspectors are found to award slightly more lenient judgements to primary schools than their female counterparts, while permanent Ofsted employees (Her Majesty's Inspectors) are found to be harsher than those who inspect schools on a freelance basis (Ofsted Inspectors).

Key Words: Ofsted, school inspection, consistency, accountability.

Contact details: Christian Bokhove, University of Southampton, C.Bokhove@soton.ac.uk
<https://doi.org/10.5258/SOTON/P1106>

4

Data

Inspections between Sep 2011 – August 2019.

Primary and secondary schools (not other remits)

Names of inspectors extracted from inspection reports:
By us.
By "Watchsted" (who gave us access to their website)
<https://www.watchsted.com/>

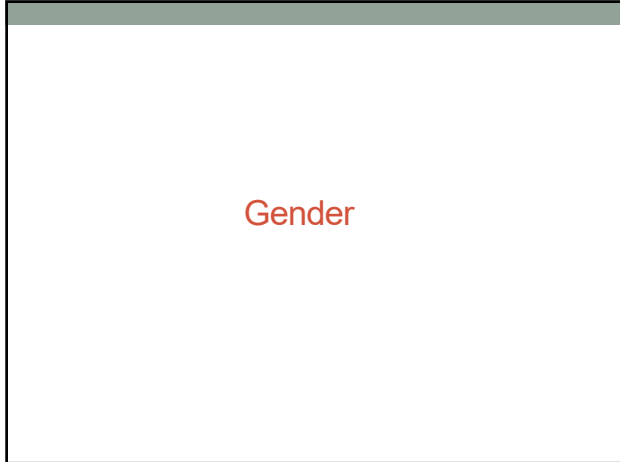
Have quality assured as far as possible
E.g. Manually checked for a sample inspector names correct

5

Three main findings

1. Evidence of a (modest) gender difference in primary inspection outcomes.
2. Evidence of a difference in inspection outcomes between HMIs and OIs.
3. Differences by inspection team size

6



7

Inspector gender. Primary

Table 1. Difference in Ofsted overall effectiveness grades awarded by male and female inspectors. Primary schools.

| | Female | Male | Difference |
|-------------------------|---------------|---------------|------------|
| Outstanding | 7.8% | 8.2% | 0.4% |
| Good | 55.9% | 58.7% | 2.9% |
| Requires improvement | 30.5% | 28.6% | -1.9% |
| Inadequate | 5.9% | 4.5% | -1.4% |
| # of inspections | 11,056 | 11,698 | |

Source: Bokhove, Jerrim and Sims (2022: Table 5).

Female = 36.4% RI / Inadequate
 Male = 33.1% RI / Inadequate

- Very unlikely to be due to "selection" into different inspection tasks
- Not due to small sample size.

8

Inspector gender. Secondary

Table 3. Difference in Ofsted overall effectiveness grades awarded by male and female inspectors. Secondary schools.

| | Female | Male | Difference |
|-------------------------|--------------|--------------|------------|
| Outstanding | 10.9% | 10.1% | -0.9% |
| Good | 45.4% | 44.9% | -0.5% |
| Requires improvement | 34.6% | 34.6% | -0.1% |
| Inadequate | 9.1% | 10.5% | 1.4% |
| # of inspections | 2,188 | 2,813 | |

Source: Bokhove, Jerrim and Sims (2022: Table 5).

Evidence of gender gap inconclusive.

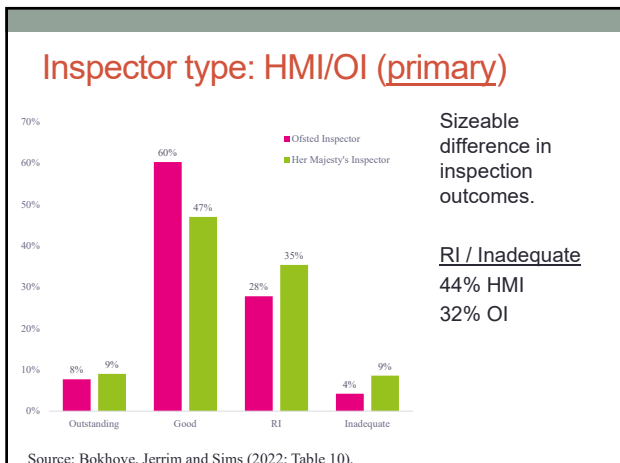
Sample size (# inspections) not large enough to say whether there is a gender difference or not

9

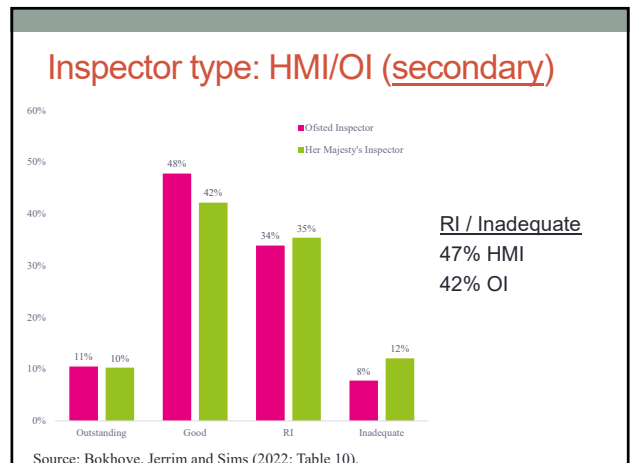
HMI / OI

HMI = Civil servants. Ofsted is only job.
 OI = Freelancers. Many hold jobs elsewhere (e.g. headteachers)

10



11



12

HMI/OI differences partly driven by differences in inspection tasks assigned

Table 2. The prior inspection ratings and Key Stage 2 maths quintiles of schools inspected by HMIs and OIs

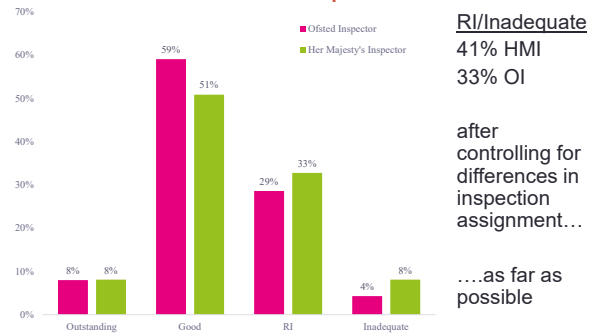
| | OI | HMI |
|-----------------------------------|-----|-----|
| Prior inspection rating | | |
| Outstanding | 7% | 13% |
| Good | 42% | 37% |
| RI | 47% | 32% |
| Inadequate | 2% | 14% |
| Key Stage 2 maths quintile | | |
| Q1 (low achievement) | 22% | 31% |
| Q2 | 20% | 18% |
| Q3 | 19% | 20% |
| Q4 | 16% | 12% |
| Q5 (high achievement) | 13% | 10% |
| Missing | 10% | 10% |

HMIs more likely assigned to inspect low-performing schools

Source: Bokhove, Jerrim and Sims (2022: Table 6).

13

...but this can only explain some of the difference in HMI/OI inspection outcomes



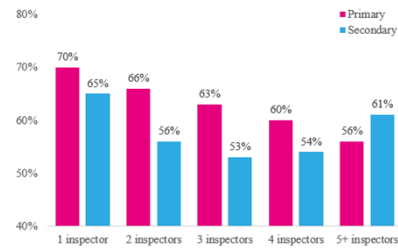
14

Size of inspection team

15

Inspection team size

Figure 2. The percent of schools rated Outstanding or Good by inspection team size. Unconditional results.



Source: Bokhove, Jerrim and Sims (2022: Table 5). Sample excludes short inspections.

16

Joint effects of these together

Note: Looking at the extremes!
Most "lenient" versus "harshes"t inspector

17

Joint effects. The two extremes.

Table 1. Variation in Overall Effectiveness judgements awarded to primary schools by two different hypothetical inspectors.

| | Inspector A | Inspector B | Risk ratio (A/B) |
|---|--------------|-------------|------------------|
| Overall effectiveness | | | |
| Outstanding | 6% | 8% | 0.77 |
| Good | 45% | 60% | 0.76 |
| Requires improvement | 35% | 29% | 1.23 |
| Inadequate | 13% | 3% | 3.94 |
| Short inspection | | | |
| Conversion with downgrade or S5 next due to concerns. (Jan18 - Aug19) | 16% | 10% | 1.60 |
| Inspector characteristics | | | |
| Team size | 2 inspectors | 1 inspector | |
| Contract status | HMI | OI | |
| Gender | Female | Male | |

Source: Bokhove, Jerrim and Sims (2022: Table 29). Estimates based upon a multinomial logistic regression model controlling for percent of pupils eligible for FSM, region, previous Ofsted inspection outcome, inspection type, Key Stage 2 maths and English scores, school absences, percent of pupils with English as an additional language and whether the inspection was conducted after 2018.

18

Joint effects. All combinations.

| Contract type | Gender | Team size | % RI / inadequate |
|---------------|--------|--------------|-------------------|
| OI | Male | 1 inspector | 32% |
| OI | Female | 1 inspector | 35% |
| OI | Male | 2 inspectors | 36% |
| OI | Female | 2 inspectors | 39% |
| HMI | Male | 1 inspector | 40% |
| HMI | Female | 1 inspector | 44% |
| HMI | Male | 2 inspectors | 44% |
| HMI | Female | 2 inspectors | 49% |

19

Where we didn't find difference

- Inspector experience (though tough to measure)
 - Though this tough to measure in our data
- Inspecting inside / outside of "home region"
- By primary / secondary specialism
 - E.g. An inspector who main inspects secondary schools doing a primary school inspection.

20

Caveats

- Data pre-EIF (2012 – August 2019)
- HMI vs OI differences.
 - Can't control for some likely differences in allocation of inspection tasks (e.g. due to safeguarding concerns).
 - Think likely to explain some – but not all – if the HMI/OI gap.....
 - Still see HMI/OI difference for short inspections
- Sample for secondary schools much smaller
 - Hence more uncertainty
- Joint effects = extreme example. (Harshes vs most lenient)
- Many potentially interesting differences we can't look at (as don't have data)

21

Timeline

- Press 26 January 2023
- Launch 2 February 2023
- Feb submission to journal
- March – webinar of findings (anyone like to host us!?)

22

Next steps

SEQUENCE ANALYSIS

TEXT ANALYSIS OF INSPECTION REPORTS

23

THANK YOU

QUESTIONS?

24