Are some school inspectors more lenient than others?

Christian Bokhove (Southampton Education School, University of Southampton, Southampton, UK)

John Jerrim (UCL Social Research Institute, UCL, London, UK)

Sam Sims (UCL Centre for Education Policy and Equalising Opportunities, UCL, London, UK)

Author list alphabetical. All joint first authors.

Abstract

School inspections are a common feature of education systems across the world. These involve trained professionals visiting schools and reaching judgements about the quality of education they provide. Yet there is currently little academic research 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 Office for Standards in Education, Children's Services and Skills (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.

Conflict of interest statement: John Jerrim is currently a part-time specialist advisor to Ofsted on academic research on secondment from UCL. This paper is part of his independent research conducted as an academic at UCL.

Author contribution statement: Bokhove, Jerrim and Sims are joint first authors. They made an equal contribution to the conceptualisation of the research and the methodology used. Jerrim and Bokhove have led on extracting inspector names and linking this with publicly available data on Ofsted inspection outcomes. Jerrim led the writing of the introduction and data sections. The methodology, analysis, results and concluding sections were co-produced jointly by the three authors.

Acknowledgements: The Nuffield Foundation is an independent charitable trust with a mission to advance social wellbeing. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org. We are grateful for their support. Helpful comments have been received on the draft from our project advisory group, whom we would like to thank for their input and support.

**Introduction**

School inspections involve a team of trained inspectors visiting schools and judging the quality of education that they provide. The outcomes are often high stakes for schools and their staff (Kemethofer et al., 2017), with judgements widely reported by local media. In the extreme, inspection judgements can lead to school closures or the removal of headteachers (Eyles & Machin, 2019). Data and reports from inspections also get widely used by a variety of stakeholders, including parents when choosing schools (Bokhove et al, 2023). Given the importance attached to inspection outcomes, it is vital they are as valid, consistent and reliable as possible. Inspectorates – such as the Office for Standards in Education, Children's Services and Skills (Ofsted) in England – therefore devote significant time and resource into developing inspection frameworks, and training inspectors in their use (Ofsted, 2022).

Despite these efforts, some have questioned the validity of Ofsted inspections (whether they accurately capture school quality) and the consistency of outcomes across different inspectors (whether the same judgements would be made if the inspection were conducted by different inspectors or on different days). Despite the effort of inspectorates to develop frameworks and provide training in their use, evidence from the public administration literature questions how much control central government bodies have over the actions of their front-line employees (Ingersoll, 1993). Moreover, the subjective nature of inspection means that a degree of human judgement will always be involved (Spielman, 2017). This is recognised within Ofsted’s inspection handbook, which states that inspectors should draw on all the evidence they have gathered and use their professional judgement (Ofsted, 2022). Yet this has led some to question the usefulness of school inspections as a mechanism for monitoring school standards and as a force for improvement (National Education Union, 2023). There is particular concern that inspection outcomes may be influenced – at least in part – by factors outside of a school’s control (Richmond, 2019). This includes, for instance, the characteristics of the inspector(s) they are assigned.

Relatively few studies have been conducted into the consistency of school inspections. In the late 1990s, Ofsted investigated whether different inspectors observing the same lesson awarded it the same grade (Matthews et al., 1998). Collecting data from 100 inspections, encompassing 173 pairs of inspectors, they found a strong correlation between the judgements (Pearson correlation = 0.81) but only moderate inter-rater reliability (Cohen’s Kappa = 0.53). They hence concluded “that OFSTED’s Framework and related advice provide an effective means by which such inspectors can judge teaching with considerable reliability” (Matthews et al., 1998, p. 186). More recently, Ofsted (2017) investigated the consistency of 24 short inspections. Each school was assigned two inspectors, with their judgements compared. The same inspection outcome was reached for 22 out of the 24 schools. This research has, however, been criticised on methodological grounds (Pearson, 2018).

Other research by Ofsted has focused on the inter-rater reliability of specific inspection tasks. One example is where nine of Her Majesty’s Inspectors (HMIs) undertook “workbook scrutiny”, with the same documents evaluated by two or three independent inspectors (Ofsted, 2019b). “Moderate” levels of inter-rater reliability were found (Cohen’s Kappa around 0.5). There have been similar investigations into lesson observations (Ofsted, 2019a), with moderate-to-substantial levels of reliability found for schools (Kappa around 0.6), but lower for colleges (Kappa around 0.3). The same study reported greater levels of consistency between two of Her Majesty’s Inspectors (HMIs) than between an HMI and a (freelance) ‘Ofsted inspector’.

Yet there remain notable limitations with the existing evidence base. The important work previously published by Ofsted is based on a small number of inspectors, and it is not clear to what extent the results generalise across the inspection workforce. In particular, existing work has focused on Her Majesty’s Inspectors (HMIs) - permanent Ofsted employees - with less focus on the freelancers (‘Ofsted Inspectors’) that the inspectorate employs. Moreover, HMIs that participated in previous work would have known they were involved in a research study, potentially influencing their behaviour. Less attention has been paid to variation in inspection outcomes across different inspectors when they are conducted in a more ‘natural’ setting (i.e., using data from live inspections). More generally, there is little existing work exploring how inspection outcomes vary depending on the inspector(s) that schools are assigned. Finally, much work has been conducted by inspectorates themselves, rather than by independent academics.

This paper begins to fill these gaps in the literature. Using data from more than 30,000 school inspections conducted in England between 2011 and 2019, we present novel evidence on how inspection outcomes vary across different (lead) inspectors, including how this is related to a set of observable inspector characteristics.

***Research questions***

We start by exploring differences according to gender. A wide body of evidence suggests important gender differences in decision making processes (Villanueva-Moya & Expósito, 2021) with it reported that *“*men decide faster, more lineal, whereas women gather information in a different way and are more aware of informal sources of information” (Gernreich & Exner, 2015, p. II). Research in criminology has also found female judges to impose harsher sentences than males (Steffensmeier & Hebert, 1999). In contrast, male and female assessors were found to provide roughly equal scores to candidates in the context of medical examinations ([McManus](https://pubmed.ncbi.nlm.nih.gov/16919156/) et al., 2006). Yet there is currently no analogous evidence with respect to gender differences in the judgements made by school inspectors. Our first research question is therefore:

RQ1. Do female inspectors make harsher or more lenient judgements about schools than their male counterparts?

Inspection outcomes may also differ between inspectors with different employment relationships with Ofsted. Broadly speaking, inspectors work for Ofsted on one of two bases. The first group are Her Majesty’s Inspectors (HMIs) – civil servants who are permanent Ofsted employees and work for Ofsted as their only job. The second group are Ofsted Inspectors (OIs) who conduct inspections for Ofsted on a freelance basis. This group typically hold other jobs in the education sector (e.g., as headteachers)[[1]](#footnote-1). Given their different employment circumstances, HMIs and OIs may differ in their views of what constitutes good practice, and in their understanding of young people’s educational and pastoral needs. For instance, OIs may be more ‘in-touch’ with the current challenges facing the teaching profession. Moreover, another key difference from HMIs is that OIs may have recently (or could soon be) subject to Ofsted inspections themselves (in their roles elsewhere in the education sector). Evidence from the management literature also suggests that employees with different contract types may differ in their motivation ([Grund](https://link.springer.com/article/10.1007/s41464-017-0033-z) & Thommes, 2017), work-related expectations and commitment (Süß & Kleiner, 2007). It is plausible that such factors may influence how OIs and HMIs go about their job, leading to a difference in the judgements they reach. Research question two is therefore:

RQ2. Do inspection judgements differ between OIs and HMIs?

Next, we turn to the link between inspection outcomes and the lead inspector’s experience. Evidence from elsewhere in the education literature (e.g., on teacher effectiveness) illustrates how experience is linked to staff productivity (Burroughs et al., 2019). Moreover, employees new to their roles tend to be less confident and more liable to make mistakes than senior staff (Grohnert et al., 2019). Indeed, experience in jobs is linked to competence development (Paloniemi, 2006). On the other hand, newly appointed inspectors may be concerned about making potentially controversial, high-stakes decisions when they are fresh into the role (e.g., awarding an Inadequate judgement or downgrading a school). Hence (in)experience could be a key source of inconsistency (and thus variation) in outcomes across inspectors. Our third question is therefore:

RQ3. How are inspection outcomes linked to inspection experience of lead inspectors?

We then consider where the inspection is taking place. Ofsted’s regional operating model means that inspectors usually conduct their inspections within one of Ofsted’s eight regions (their ‘home region’). Although all regions inspect to a common framework, with a certain degree of centralised guidance and training, regions also have autonomy over delivering and managing inspections. It is possible that, when an inspector works outside their home region, they come across certain practices and approaches they are not used to. There may also be regional differences in how schools operate that impact the judgements inspectors reach. We investigate this in research question four:

RQ4. Do inspectors judge schools more harshly when they are working outside of their home region?

School inspectors will have specialist knowledge, background and skills in particular areas. One is whether they have a background in primary or secondary education (and thus primary or secondary inspections). Yet England has many more primary than secondary schools. This inevitably means that some inspectors who have knowledge and inspection experience in one school phase (e.g., secondary) will sometimes lead inspections in another (e.g., primary). This could impact the judgements made. For instance, those with a specialism in secondary inspections may ‘play it safe’ when asked to inspect a primary school, given they have less experience in this area. They may thus shy away from issuing potentially high-stakes grades (e.g., Inadequate judgements). Alternatively, secondary schools in England tend to receive lower Ofsted grades than primary schools[[2]](#footnote-2). Inspectors who usually inspect secondary schools may hence also award lower grades to primary schools. This is investigated in our fifth research question:

RQ5. Do inspectors with a specialism in secondary school inspections judge primary schools more harshly than inspectors with a primary specialism?

Finally, some school inspections are carried out by a single inspector rather than by a team (Table 1 provides details). Yet previous research has noted how, when making decisions, “individuals are more likely to be influenced by biases, cognitive limitations, and social considerations” than groups (Charness & Sutter, 2012, p. 158). This is potentially due to the benefits of pooling information, discussing the evidence, overcoming unconscious biases and drawing on the wisdom of groups (Bang & Frith, 2017). Indeed, within the broader literature on inspection, research has found that “groups of inspectors produced more reliable assessments than individual inspectors” in the context of hospitals (Boyd et al., 2017, p. 36). Yet, in terms of optimal team size, the evidence remains inconclusive – although somewhere between 5 and 12 team members is often cited (Powell & Lorenz, 2019). Moreover, the potential advantages of larger teams may be dissipated if it leads to “groupthink”, a tendency to focus on only information that is available to all inspectors (shared information bias) or to individuals “free-riding” on the effort of others (Bang & Frith, 2017). Again, we know of little analogous evidence in the context of school inspections. We thus conclude by asking:

RQ6. Do school inspection outcomes vary by inspection team size? Do outcomes differ between teams versus individual inspections?

**Data**

Our data are mainly drawn from the ‘Watchsted’ database[[3]](#footnote-3). For each inspector, this includes details of all inspections they have conducted since September 2011, drawing on the lead inspector named in the published Ofsted reports. All secondary inspections and all primary inspections done by inspectors who have conducted at least five between September 2011 and August 2019 have been extracted. When cleaning these data, we merge together instances where a similar name is used (e.g., Ash Rahman and Ashfaq Rahman have been combined into a single record)[[4]](#footnote-4).

These data have then been merged with publicly available information on inspection outcomes[[5]](#footnote-5). This was done in three steps. First, for each inspection extracted from the Watchsted database, we take the start date and restrict the data on inspection outcomes to only those inspections conducted on that date (i.e., we force an exact match on inspection start date). Second, within this subset, we fuzzy match across the databases on school name. Finally, we check that the information on inspection outcomes – including sub-judgements – is consistent. Cases were dropped in the few instances where differences were found. This process was conducted separately for primary and secondary schools. The final dataset includes 35,751 inspections (29,850 primary and 5,901 secondary) conducted between September 2011 and August 2019 by a total of 1,376 inspectors. This represents 81% of all inspections conducted over this period (see Appendix 1). Appendix 2 discusses our data in further detail and also provides alternative estimates using a slightly larger sample (40,959 inspections - 93% of the total). This leads to little change to our substantive results.

Appendix 3 provides details about how we have checked the quality of our data. In brief, we randomly sampled 300 inspections, accessed the relevant inspection reports from the Ofsted website and manually recorded the relevant information (e.g., inspector name, whether an HMI led the inspection). This information was then cross-referenced against what is recorded in our dataset. The level of agreement is high, with the name of the lead inspector matching on more than 97% of occasions (confidence interval 94%-100%). This provides reassurance that measurement error in our data is likely to be low.

These data were then subsequently linked to the Department for Education’s (DfE) School Performance Tables. This includes background characteristics (e.g., admissions policy, religious denomination, school type), composition of the student body (e.g., percent of pupils eligible for Free School Meals, percent of pupils with English as an Additional Language) and national examination performance.

Our primary analysis is concerned with how the characteristics of inspectors are linked to the Overall Effectiveness judgements they make, using Ofsted’s four-point scale:

1. Outstanding
2. Good
3. Requires Improvement / Satisfactory[[6]](#footnote-6)
4. Inadequate

This measure is only available for ‘full’ inspections. Yet, since 2015, around half of all Ofsted inspections are short (‘section 8’) inspections. We hence also investigate how lead inspector characteristics are linked to short inspection outcomes between January 2018 and August 2019 [[7]](#footnote-7). Specifically, we create a binary measure coded one if the inspector decided the school should either receive a full inspection next due to concerns or immediately converted it to a full inspection with a subsequent downgrade, and zero otherwise.

The following information has been derived about individual inspectors:

* Whether an inspector is an HMI. For each inspector in the Watchsted database, there is a flag to indicate whether they are an HMI. Any inspector with such a flag is coded as an HMI, with all others assumed to be an Ofsted Inspector (OI).
* Gender. The python GenderGuesser package ([Perez](https://pypi.org/project/gender-guesser/), 2016) was used to predict the gender of each inspector, based on first name. A small amount of manual coding has also been conducted, where results were ambiguous.
* Primary/secondary specialism. Some inspectors conduct inspections in a single school phase (primary or secondary) while others work across both. We derive a variable, based on each inspector’s inspection history, identifying whether they have conducted primary inspections only, secondary inspections only, or a mix.
* Home region. Ofsted operates a regional operating model, with each inspector sitting within a regional team. It is however possible for inspectors to conduct inspections outside of their ‘home’ region. For each inspector who has conducted more than 10 inspections between September 2011 and August 2019, we define their ‘home region’ as the area where they have conducted most inspections[[8]](#footnote-8). A binary variable is then derived, identifying for each inspection whether the inspector was working in their home region.
* Experience. Total inspection experience is measured as the number of inspections an inspector has previously conducted (before their current inspection), with the count starting in September 2011.
* Inspection team size. This is measured as the number of inspectors named in the inspection report. This information is not available from the Watchsted database; we have extracted it via our own scraping of Ofsted reports (see Appendix 2 for further details).

The distribution of these variables across all inspections included in our analysis can be found in Table 1. HMIs are slightly more likely than OIs to lead short inspections (60%/40% split). For other inspection types, however, OIs are more likely to be the lead than HMIs – particularly in primary schools (80%/20% split). This is important given that – as noted in the introduction – most previous work into Ofsted inspections has not included OIs. Despite women being more likely to work in the teaching profession than men - particularly in primary schools (Jerrim & Sims, 2019) - the same does not hold true with respect to inspections, where the gender split is broadly even. Most primary inspections are conducted by primary inspection specialists, although around 10% are led by an inspector whose workload has included a significant proportion of secondary inspections. The analogous holds true with respect to secondary inspections. While short inspections are almost always conducted within an inspector’s home region, approximately one-in-seven (15%) of non-short inspections are conducted outside it. The average primary inspection is conducted by someone who has led around 30 inspections since 2011, though there is quite a lot of variability around this figure (standard deviation ~25). For secondary inspections, the average amount of prior lead experience is somewhat lower (an average of 19 prior inspections led). Finally, primary inspections are conducted by smaller teams. Almost two-thirds of primary inspections (that are not short inspections) are conducted by one or two inspectors (63%), compared to just 14% of secondary inspections. This partly reflects differences in school size.

**<< Table 1 >>**

**Methodology**

To begin, we present cross-tabulations describing how each inspector characteristic is related to inspection outcomes. Of course, these unconditional relationships may be confounded by other factors. For instance, Ofsted could be more likely to assign inspectors with certain characteristics to inspect certain types of school.

We consequently estimate a set of ordered logistic regression models to try and account for possible differential selection of lead inspectors to different inspection tasks. These control for a set of factors related to inspection outcomes and may be associated with lead inspector (and inspection team) assignment. All models will be estimated separately for primary and secondary schools and are of the form:

(1)

Where:

= The Overall Effectiveness judgement made by the inspector.

= The characteristic of the lead inspector under investigation.

= A vector of inspection-specific controls. These are either characteristics of the school being inspected (e.g., performance in national examinations) or the type of inspection being conducted.

= Other characteristics of the lead inspector (other than the characteristic under investigation).

i = Inspection i.

j = Inspector j.

k = A specific category on Ofsted’s four-point overall effectiveness scale.

The parameter of interest is . This captures the strength of the association between the characteristic under investigation (e.g., gender) and inspection outcomes. Estimates will be presented as odds-ratios, capturing the increase in the odds of receiving a *worse* inspection rating. For instance, an odds ratio of two will indicate that the odds of receiving an Outstanding versus a Good/RI/Inadequate rating are twice as large, conditional on the factors controlled in the model. A separate model is estimated for each characteristic under investigation. Analogous models to those presented in equation (1) are estimated via binary logistic regression for short inspection outcomes.

The headline results reported in the main body of the paper include controls for:

* Percent of pupils eligible for FSM
* School religion
* School gender
* Ofsted region
* Inspection type
* Prior Ofsted rating
* School performance data (e.g., prior Key Stage 2 scores for primary and prior Key Stage 4 scores for secondary)
* School absences
* Percent of pupils at the school with Special Educational Needs
* Percent of pupils who speak English as an Additional Language
* Other background inspector characteristics

Appendix 7, 8 and 9 provides additional results for each characteristic to illustrate the robustness of findings to several different model specifications.

To account for the nested structure of the data, standard errors are clustered at the inspector (j) level. In Appendix 4 we compare estimates to those from multi-level (random effects) models and find little substantive difference[[9]](#footnote-9). Appendix 5 presents a selection of sub-group estimates by gender and contract status. Likewise, in Appendix 6 we present alternative estimates based upon multinomial (rather than ordinal) logistic regression to investigate the sensitivity of our findings to relaxing the proportional odds assumption.

***Joint effect – looking at the impact of multiple characteristics together***

To investigate the combined effect of multiple inspector characteristics we estimate an ordinal logistic regression model including the two inspector characteristics that we have found to be clearly associated with inspection outcomes (gender, HMI/OI) along with inspection team size and school/inspection controls. We then consider differences in the predicted Overall Effectiveness distribution between two hypothetical inspectors:

* Inspector A. A female HMI who is accompanied by one other inspector.
* Inspector B. A male OI who is conducting the inspection alone.

Our focus here will be primary school inspections, given the much larger sample available. The results we report will be when these two hypothetical inspectors are inspecting schools with a similar proportion of disadvantaged pupils, within the same Ofsted region, have similar levels of performance in the Key Stage 2 tests, have the same previous Ofsted inspection judgement, have similar levels of school absence, similar proportions of pupils who speak English as an Additional Language and are undergoing the same type of inspection.

**Results**

***RQ1. Do female inspectors make harsher judgements than their male counterparts?***

Table 2 presents crosstabulations between Overall Effectiveness grades and lead inspector characteristics. Panel (a) refers to gender.

**<< Table 2 >>**

Starting with primary schools, evidence emerges of a modest gender difference. Female lead inspectors make slightly harsher judgements about primary schools than males. For instance, male lead inspectors judged 33.1% of primary schools as Require Improvement or Inadequate, compared to 36.4% of female leads. The difference in the Inadequate grade (5.9% versus 4.5%) is notable given the relative size of the gender difference and the high-stakes consequences attached. Male lead inspectors are, on the other hand, almost three percentage points more likely to judge schools to be Good than females. Yet there is little evidence of a gender gap for the Outstanding grade. Nevertheless, Table 2, panel (a) suggests primary inspection outcomes may differ slightly by the gender of the lead inspector.

The results for secondary schools – presented on the right-hand side of Table 2, panel (a) – are more ambiguous. The percentage of male and female lead inspectors awarding Good and Requires Improvement grades are very similar. There is perhaps more of a difference at the extremes of the grading scale, with male lead inspectors more likely to reach an Inadequate judgement (10.5% versus 9.1%) and female leads more likely to award Outstanding grades (10.9% versus 10.1%). Yet even these differences are quite small.

To what extent might these unconditional results be driven by selection? Are the hasher judgements made by female inspectors due to them being assigned more challenging primary schools to inspect? Two pieces of evidence are presented. First, Table 3 compares the distribution of observable characteristics of the primary schools inspected by male and female lead inspectors. If the gender difference in Overall Effectiveness grades for primary schools observed in Table 2, panel (a) is due to selection effects, one would expect to see female inspectors being disproportionately assigned to inspect lower ‘quality’ schools (e.g., those with lower prior inspection ratings, worse performance in national examinations, higher absence levels). Table 3 provides little indication that this is the case; the distribution of inspection is similar across male and female lead inspectors.

**<< Table 3 >>**

Second, Table 4 presents estimates from our ordinal regression models. The top row refers to those for gender, with values below one indicating that female lead inspectors make harsher judgements than their male counterparts (conditional on the controls).

**<< Table 4 >>**

The estimated odds-ratio is 0.84 and is statistically significant at the five percent level. This confirms that female lead inspectors tend to award lower inspection grades to primary schools than male inspectors, even after controlling for observable differences in the schools they are assigned to inspect (and other inspector-level characteristics)[[10]](#footnote-10). Moreover, in Appendix 9 (Table 9.3), we illustrate how the odds-ratio is very stable across multiple specifications – suggesting that any unobserved confounding would have to be generated by a factor that is strongly associated with inspection outcomes, but orthogonal to a school’s intake, performance in examinations, pupil absences and previous Ofsted grades. It is not clear what such a characteristic could be. Our interpretation is hence that the gender difference we observe in primary inspection outcomes is unlikely to be driven by inspector selection.

Table 5 turns to analogous modelling results for short primary school inspections, with the top row providing those for gender. Odds ratios below one indicate that short inspections led by females are more likely to result in a negative outcome (i.e., a full Section 5 inspection within the next year due to concerns) than those led by males.

**<< Table 5 >>**

Evidence again emerges that female lead inspectors reach slightly harsher verdicts than their male counterparts. The odds ratio sits around 0.75, indicating that the odds of a negative outcome from a short primary inspection are around 25% lower for males than females. In absolute terms, this represents a modest difference of between two and three percentage points (9.6% of short inspections with a male lead led to a negative outcome for the school compared to 12.1% of those with a female lead).

There is no evidence of such a gender difference with respect to secondary inspections. The ordinal logistic regression model presented in Table 4 illustrates that, for secondary schools, the odds ratio stands at 1.09 and is not statistically significant at conventional thresholds. Additional analyses we have performed for short secondary inspections have also proven inconclusive. Hence our finding of a small gender difference in inspection outcomes seems to only hold for primary schools.

***RQ2. Do Ofsted inspection judgements differ between OIs and HMIs?***

Table 2 panel (b) illustrates the distribution of inspection outcomes by contract status (HMI versus OI). Starting with primary schools, HMIs are around 13 percentage points less likely to judge a school to be Good than OIs (60% versus 47%). They are much more likely to award Requires Improvement (35% versus 28%) and Inadequate (9% versus 4%) grades instead. For secondary schools, HMIs judge fewer to be Good than OIs (42% versus 48%) and place more in the Inadequate category (12% versus 8%). Nevertheless, the difference between HMI and OI lead inspectors is greater for primary than secondary schools.

The right-hand columns in Table 3 provide one possible explanation of this result – Ofsted deploys OIs and HMIs to different inspection tasks. In particular, HMIs are more likely to be assigned to schools with lower performance in national examinations and those that were judged to be Inadequate during their last inspection. Thus, the ‘contract status’ row in Table 4 illustrates whether we continue to find a difference in Overall Effectiveness outcomes between HMI and OI after we have controlled – as far as possible – for differences in their inspection tasks. Odds-ratios greater than one indicate that HMIs tend to provide harsher inspection judgements than OIs.

There are two key points. First, the relationship between contract status and inspection outcomes is strong and statistically significant. Roughly speaking, the odds of a primary school being placed in a lower Overall Effectiveness category are around 50% higher if the lead inspector is an HMI rather than an OI. Second, as Appendix 9 illustrates (see Table 9.7), the inclusion of inspection, school and inspector controls only slightly weakens the relationship across model specifications with the estimated odds ratio consistently between 1.4 and 1.5). This suggests the result is not being solely driven by the selection of HMIs/OIs into different types of inspection, at least in terms of key observable characteristics such as examination performance and demographic composition. We cannot rule out the possibility, however, that HMIs and OIs are disproportionately chosen to conduct inspections based upon factors we cannot observe (and is not well-proxied by our controls). Our analysis of short primary inspections in Table 5 produces similar results; the odds of a negative outcome are around 1.6 times higher if conducted by an HMI rather than an OI (conditional on the controls). In absolute terms, this suggests that about 13% of short inspections led by HMI result in a full inspection next due to concerns, compared to 9% of those led by OIs.

Analogous results for secondary schools in Table 4 point towards a similar – although slightly weaker – relationship (odds ratio = 1.32), with the odds ratio fluctuating slightly (between 1.13 and 1.32) depending on the exact specification used (see Appendix Table I8). Moreover, alternative estimates based upon multinomial (rather than ordinal) logistic regression in Appendix 6 makes clear that for secondaries, the main point of difference between HMIs and OIs is with respect to Good and Inadequate judgements.

***RQ3. How are inspection outcomes linked to inspection experience of lead inspectors?***

The next set of estimates in Table 4 presents results for inspector experience. Those for both primary and secondary schools suggest there is no clear relationship with inspection outcomes. Moreover, in Appendix 9 (Appendix Table 9.12), we illustrate how this holds true regardless of the model specification used. Similar results also emerge for primary school short inspections in Table 5.

***RQ4. Do inspectors judge schools more harshly when they are working outside of their home region?***

Table 2 panel (c) presents a cross-tabulation between whether the inspection was conducted inside the inspector’s home region and Overall Effectiveness judgement. For primary schools the distribution is very similar whether the inspection was conducted within inspector’s home region or not. The regression model estimates presented in Table 4 suggest that, although statistically significant due to the large sample size, any association here is weak (odds ratio = 1.13). For primary school short inspections (Table 5), only 96 out of the 3,311 within our database have been conducted outside of the lead inspectors home region meaning one should not read too much into these estimates (the odds ratio is quite sizeable at 1.41 but not statistically significant).

For secondary schools, the cross-tabulations in Table 2, panel (c) suggest that those conducted outside the inspectors’ home region are more likely to be rated Outstanding (15 versus 9 percent) and less likely to rated Inadequate (7 versus 11 percent). However, the estimated odds ratio quickly approaches one in the ordinal logistic regression model results presented in Table 4. We thus conclude that there is no evidence that the inspection judgements secondary schools receive are related to whether the lead inspector was working in their home region or not.

***RQ5. Do inspectors with a specialism in secondary school inspections judge primary schools more harshly than inspectors with a primary specialism?***

Table 2, panel (d) presents a cross-tabulation between the percent of primary school inspections each inspector has conducted during their career and Overall Effectiveness judgements. No clear relationship is found for either primary or secondary schools. This continues to hold after controlling for a set of school, inspection and inspector characteristics within our ordinal regression models in Table 4. For short primary school inspections (Table 5), there is some suggestion of a difference between those who have only conducted primary inspections and those who have conducted a mix of primary and secondary inspections (odds ratio around 1.4), though these differences are only statistically significant at the ten percent level. Overall, it seems there is relatively little evidence of an association between whether inspectors have a specialism in the primary/secondary sector and inspection outcomes.

***RQ6. Do school inspection outcomes vary by inspection team size? Do outcomes differ between team versus individual inspections?***

Table 6 presents a cross-tabulation between inspection team size and Overall Effectiveness grades. For primary inspections, larger teams are less likely to reach a Good judgement and are more likely to rate schools as Inadequate or Requires Improvement. The difference between a single inspector and 2 to 3 inspectors remains statistically significant in our ordinal regression models, with the estimated odds ratio standing around 1.25. Appendix 9 illustrates that this result is robust to a wide set of alternative model specifications (see Table 9.23).Additional multinomial logistic regression estimates (see Appendix 6 – Table 6.7) point towards the most notable difference to occur with respect to the Inadequate grade. Specifically, the predicted probability of receiving an Inadequate grade is 3.4% when the primary inspection is conducted by a single inspector, versus around 6% when it is conducted by a team of two, three or four inspectors. Interestingly, however, we find no association between short primary inspection outcomes and team size.

**<< Table 6 >>**

For secondary schools, very small teams (one inspector) and large teams (five inspectors or more) seem to make slightly less harsh judgements than secondary inspections conducted by a team of four. The ordinal regression estimates in Table 4 are consistently statistically significant at the five percent level, with the estimated odds ratio around 0.4 with respect to a single inspector (relative to a team of four inspectors) and 0.8 for a team of five inspectors. There is hence some evidence that – for both primary and secondary inspections – inspection team size is independently associated with Overall Effectiveness judgements, over and above our school and inspection level controls.

***Joint effect – looking at the impact of multiple characteristics together***

To conclude, we examine the combined effect of multiple inspector (and inspection team) characteristics at the same time. Results can be found in Table 7. This part of our analysis focuses on primary schools only, given this is where we have found the most convincing evidence of difference in preceding subsections.

**<< Table 7 >>**

There is a clear, sizeable difference in inspection outcomes reached by our two hypothetical lead inspectors. Inspector A is around twice likely to award an Inadequate judgement than inspector B (9.0% versus 4.5%), while being around half as likely to judge a primary school to be Outstanding (4.5% versus 9.1%). Likewise, almost half of the primary schools inspected by inspector A will be judged to be Inadequate or Requires Improvement, compared to less than a third of those inspected by inspector B. The analogous difference for short primary inspection outcomes – with respect to recommending a full section 5 inspection next due to concerns - is 15.5% for Inspector A compared to 9.7% for Inspector B. Appendix 9 provides an alternative version of Table 7 using multi-nominal – rather than ordinal – logistic regression modelling (see Appendix Table 9.25). The clearest point of difference when using this alternative analytic approach is an increase in the difference between the two hypothetical inspectors awarding the Inadequate grade (13.3% versus 3.4%).

**Conclusions**

School inspections are a common feature of education systems across the globe. Although such inspections come in different shapes and sizes, in some countries – such as England – they are a key part of the accountability system. Ofsted – the school inspectorate in England – is one example where a team of inspectors make high-stakes judgements about schools. Yet relatively few studies have been conducted into variation in school inspection outcomes, with most existing work limited in scope and conducted by school inspectorates themselves (e.g., Matthews et al., 1998; Ofsted, 2017).

This paper has sought to address this gap in the literature. Using data from more than 30,000 school inspections conducted over an eight-year period, we have produced the first evidence on how school inspection outcomes are linked to characteristics of the lead inspector. Robust evidence emerges that male inspectors make more lenient judgements about primary schools than females. Although the magnitude of the gender differences is relatively small, it is most apparent at the high-stakes (Inadequate) grade. Much larger differences are observed between inspectors working under different contractual arrangements (HMIs versus OIs), with the former reaching harsher judgements than the latter. Likewise, inspection team size also appears to be independently associated with Overall Effectiveness grades. In contrast, there is little – or at best mixed - association between inspection outcomes and the lead inspector’s experience, primary/secondary specialism or whether the inspection was conducted outside their home region. Likewise, partly due to the smaller sample size – and potentially also the bigger average inspection team size – weaker and more uncertain evidence of variation by lead inspector characteristics has emerged for secondary schools (in comparison to primary schools).

Previous research published by Ofsted has found short inspections carried out by different inspection teams usually reach the same judgement (Ofsted, 2019a), and has claimed that (Ofsted, 2019b, p. 1) “Her Majesty’s Inspectors (HMI) can assess the quality of education by using workbook scrutiny indicators and they do so reliably”. At first glance, this appears at odds with our findings. However, it is important to note that the research Ofsted has published into workbook scrutiny and lesson observations actually shows there to be non-trivial differences in the opinions formed by different inspectors even when they are looking at the same piece of evidence – suggesting that the headline claim that HMIs can do such tasks “reliably” may be somewhat oversold. Moreover, the previous work conducted by Ofsted has only utilised a small number of HMIs. Our findings – particularly the sizeable gap in judgements made by HMIs and Ois – suggest that Ofsted’s previous work may lack external validity; that results from their studies cannot necessarily be generalised to the inspection workforce as a whole. Indeed, it is vital that any future research conducted into school inspection consistency and reliability involves a truly representative cross-section of inspectors, rather than being restricted to a small number of selected individuals. This in-turn motivates the need for future studies involving the different inspection teams making independent judgements about the same school on different days, and further analysis similar to ours that monitors what is happening across a wide array of live inspections on the ground.

We can only speculate as to why we observe the small but important gender differences in inspection judgements. One possibility is that the gender gap is being driven by differences in personality traits, with men being more likely to be overconfident in their knowledge and skills (\*blind for review\*), while women have higher levels of conscientiousness (Verbree et al., 2022). In job promotions, Hartman et al. (1991, p. 285) argued that it’s “predominantly the gender stereotype of the ratee’s personal characteristics rather than the ratee’s gender that influences the promotion process”. It is plausible that such personality traits are linked to school inspection outcomes, thus driving the gender difference that we observe. Alternatively, previous research has suggested that there are important gender differences in decision making processes when working as part of a team. For instance, Kennedy (2003) notes how women tend to be more altruistic in their decision-making and prefer reaching a universal solution, while men are more motivated by self-interest. In a similar vein, Friesdorf et al. (2015) note how men have a stronger preference for utilitarian judgements (i.e., consider the overall consequences of an action) over deontological judgments (i.e., consider the actions consistent with moral norms) compared with women when faced with moral dilemmas. This could lead men and women to make different (high stakes) decisions, such as the inspection judgement awarded to a school. Villanueva-Moya & Exposito (2021) highlight the relevance of psychosocial variables like stereotype threat and fear of negative evaluation, in women’s decision-making processes. Some evidence points towards effective interventions for stereotype threat (Liu et al., 2021), although some scholars argue that this depends on the form of stereotype threat (e.g., Shapiro et al., 2013). Finally, male and female inspectors may differ in their professional experiences, including their subject/phase specialisms and the leadership roles that they have held. Again, such factors may also be related to inspection outcomes, and thus are also potential explanations for the gender difference we observe. Ultimately, however, this is an empirical question – and one that do we not have the data to answer. An important direction for future research is hence to develop a better understanding of what exactly is driving the gender difference in primary school inspection outcomes.

These findings should be interpreted considering the limitations of our work. Three issues stand out. First, our estimates capture conditional associations only, rather than capturing cause and effect. Some of the differences we observe (e.g., between HMIs and OIs) may to some extent be driven by selection (different lead inspectors being assigned to different tasks). We have discussed this issue at length during our analyses and have attempted to control for such differences via estimation of regression models. Nevertheless, we recognise this may only partially overcome such issues. Second, we have only considered variation by a limited set of observable inspector characteristics. Arguably, there are likely to be more important sources of variation in inspection outcomes in terms of things we cannot observe, such as inspectors’ personalities and professional history (e.g., whether the inspector has previously led a challenging school). This should be a key line of inquiry in future research. Third, it has not been possible with the data currently available to understand what may be driving between-inspector variation in school inspection outcomes. Future research – both quantitative and qualitative - should seek to better understand the mechanisms behind the differences that we observe. Finally, a new inspection framework was introduced by Ofsted in September 2019, which puts less emphasis on performance in national examinations and more on the quality of the curriculum. Unfortunately, only six months of inspection data are available from this new framework before the COVID-19 pandemic disrupted school inspections for the following two years. Our analysis has thus been restricted to before the most recent framework change. However, given that our analytic sample covers an eight-year period during which multiple changes were made to how school inspections were conducted (including previous framework changes) we do not believe different findings would emerge now. Nevertheless, once data from further inspections are available under the new framework (outside of the pandemic era) we believe it is important that Ofsted publishes an update building on our work.

With these caveats in mind, the key question becomes how much should our results be cause for concern? After all, Ofsted inspection frameworks explicitly recognise that inspectors should use their professional judgements when interpreting the evidence collected, with the variation we observe perhaps just reflecting this. In other words, there will of course be some degree of variation in outcomes in any process that involves human judgement. The most pertinent question thus becomes how much variation in outcomes across different inspectors is too much? This is not an easy question to answer and is open to debate. That said, we note that one of the clearest points of difference across lead inspectors in our work is with respect to what is widely perceived to be the judgement with the highest stakes, the Inadequate grade. Given the consequences of receiving an Inadequate judgement, almost any variation across inspectors in reaching this decision might be considered an issue.

What then should be the next step for Ofsted and other school inspectorates? Given the dearth of evidence on this matter – across the UK and internationally – school inspectorates should publish more research into the reliability and consistency of inspections, including variation in inspection outcomes. It is only with such evidence that an open and informed debate can be had about such issues. Indeed, if governments are to have a school inspection regime with high stakes outcomes, then it is vital that they are proven to have a high degree of validity, reliability and consistency. Our findings suggest that, in the case of Ofsted, the high-stakes consequences attached to certain inspection outcomes may need to be adjusted downward. At the same time, it is equally important to ensure that inspections and inspectorates are appropriately resourced to deliver the level of reliability and consistency that government requires.

At the same time, open data sources should also be created by school inspectorates – such as depositing in the Office for National Statistics Secure Research Service an inspector-inspection linked database – to allow independent researchers to also explore such issues in a quicker, simpler way, than is currently possible. Likewise, more needs to be documented, investigated and discussed about inspector deployment – how exactly are inspectors assigned to different inspection tasks? Finally, Ofsted might consider publishing further details about its quality assurance processes, particular with respect to what happens when schools receive an Inadequate grade.

**References**

Bang, D., & Frith, C. (2017). Making better decisions in groups. *Royal Society Open Science*, *4*(8), 170193. <https://doi.org/10.1098/rsos.170193>

Bokhove, C., Jerrim, J. & Sims, S. (2023) How useful are Ofsted inspection judgements for informing secondary school choice? *Journal of School Choice* 17(1): 35-61. DOI: [10.1080/15582159.2023.2169813](https://doi.org/10.1080/15582159.2023.2169813)

Boyd, A., Addicott, R., Robertson, R., Ross, S., & Walshe, K. (2017). Are inspectors’ assessments reliable? Ratings of NHS acute hospital trust services in England. *Journal of Health Services Research & Policy*, *22*(1), 28–36. <https://doi.org/10.1177/1355819616669736>

Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., & Schmidt, W. (2019). A Review of the Literature on Teacher Effectiveness and Student Outcomes. In: *Teaching for Excellence and Equity*. IEA Research for Education (A Series of In-depth Analyses Based on Data of the International Association for the Evaluation of Educational Achievement (IEA)), Vol 6. Springer. <https://doi.org/10.1007/978-3-030-16151-4_2>

Charness, G., & Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives*, *26*(3), 157–76. <https://doi.org/10.1257/jep.26.3.157>

Eyles, A., & Machin, S. (2019). The introduction of academy schools to England’s education. *Journal of the European Economic Association*, *17*(4), 1107–1146. <https://doi.org/10.1093/jeea/jvy021>

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender Differences in Responses to Moral Dilemmas: A Process Dissociation Analysis. *Personality and Social Psychology Bulletin*, *41*(5), 696–713. <https://doi.org/10.1177/0146167215575731>

Gernreich, C., & Exner, C. (2015). A comparison of the influence of gender on managerial decision making. Retrieved from <https://www.researchgate.net/publication/278679030_A_Comparison_of_the_Influence_of_Gender_on_Managerial_Decision_Making>

Grohnert, T., Meuwissen, R., & Gijselaers, W. (2019). Enabling young professionals to learn from errors – the role of a supportive learning climate in crossing help network boundaries. *Vocations and Learning*, *12*, 217–243. <https://doi.org/10.1007/s12186-018-9206-2>

Grund, C., & Thommes, K. (2017). The role of contract types for employees’ public service motivation. *Schmalenbach Business Review*, *18*, 377–398. <https://doi.org/10.1007/s41464-017-0033-z>

Hartman, S. J., Griffeth, R. W., Crino, M. D., & Harris, O. J. (1991). Gender-based influences: The promotion recommendation. *Sex Roles*, *25*(5), 285–300. <https://doi.org/10.1007/BF00289757>

Ingersoll, R. (1993). Loosely coupled organizations revisited. *Research in the Sociology of Organizations*, *11*, 81–112.

Kemethofer, D., Gustafsson, J-E., & Altrichter, H. (2017). Comparing effects of school inspections in Sweden and Austria. *Educational Assessment, Evaluation and Accountability*, *29*, 319–337. <https://doi.org/10.1007/s11092-017-9265-1>

Liu, S., Liu, P., Wang, M., & Zhang, B. (2021). Effectiveness of stereotype threat interventions: A meta-analytic review. *Journal of Applied Psychology*, *106*(6), 921–949. <https://doi.org/10.1037/apl0000770>

Jerrim, J., & Sims, S. (2019). Teachers in primary and secondary schools: TALIS 2018. Department for Education research report. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/919064/TALIS_2018_research.pdf>.

Kennedy, C. (2003). Gender Differences in Committee Decision-Making. *Women & Politics* 25(3): 27-45. DOI: [10.1300/J014v25n03\_02](https://doi.org/10.1300/J014v25n03_02)

Matthews, P., Holmes, J. R., Vickers, P., & Corporaal, B. (1998). Aspects of the reliability and validity of school inspection judgements of teaching quality. *Educational Research and Evaluation*, *4*(2), 167–188. <https://doi.org/10.1076/edre.4.2.167.6959>

McManus, I., Thompson, M., & Mollon, J. (2006). Assessment of examiner leniency and stringency (‘hawk-dove effect’) in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Medical Education*, *6*, 42. <https://doi.org/10.1186/1472-6920-6-42>

National Education Union. (2021). *Replace Ofsted*. Retrieved from <https://neu.org.uk/campaigns/replace-ofsted>

Office for Standards in Education, Children's Services and Skills. (2012). Ofsted scraps ‘satisfactory’ judgement to help improve education. *Press release*. Retrieved from https://www.gov.uk/government/news/ofsted-scraps-satisfactory-judgement-to-help-improve-education

Office for Standards in Education, Children's Services and Skills. (2017). Do two inspectors inspecting the same school make consistent decisions? A study of the reliability of Ofsted’s new short inspections. *Ofsted research report 170004*. Retrieved from <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/596708/Reliability_study_-_final.pdf>.

Office for Standards in Education, Children's Services and Skills. (2019a). How valid and reliable is the use of lesson observation in supporting judgements on the quality of education. *Ofsted research report 190029*. Retrieved from <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/936246/Inspecting_education_quality_Lesson_observation_report.pdf>.

Office for Standards in Education, Children's Services and Skills. (2019b). Workbook scrutiny. Ensuring validity and reliability in inspections. *Ofsted research report 190028*. Retrieved from <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/936240/Inspecting_education_quality_workbook_scrutiny_report.pdf>.

Office for Standards in Education, Children's Services and Skills. (2020). Ofsted inspections illustrate high proportion of good or outstanding schools. Retrieved from <https://educationhub.blog.gov.uk/2020/10/30/ofsted-inspections-illustrate-high-proportion-of-good-or-outstanding-schools/>

Office for Standards in Education, Children's Services and Skills. (2022). School inspection handbook. Retrieved from <https://www.gov.uk/government/publications/school-inspection-handbook-eif/school-inspection-handbook>

Paloniemi, S. (2006). Experience, competence and workplace learning. *Journal of Workplace Learning*, *18*(7/8), 439–450. <https://doi.org/10.1108/13665620610693006>

Pearson, T. (2018). *A review of Ofsted’s test of the reliability of short inspections*. Retrieved from [https://www.researchgate.net/publication/327894743\_A\_review\_of\_Ofsted’s\_test\_of\_the\_reliability\_of\_short\_inspections](https://www.researchgate.net/publication/327894743_A_review_of_Ofsted%E2%80%99s_test_of_the_reliability_of_short_inspections)

Perez, I. (2016). *Gender-guesser*. Retrieved from <https://pypi.org/project/gender-guesser/>

Powell, D., & Lorenz, R. (2019). The effect of team size on the performance of continuous improvement teams: Is seven really the magic number? In F. Ameri, K. E. Stecke, G. von Cieminski, & D. Kiritsis (Eds.), *IFIP International Conference on Advances in Production Management Systems* (pp. 69-76). Springer, Cham. <https://doi.org/10.1007/978-3-030-30000-5_9>

Richardson, H. (2015, June 19). Ofsted purges 1,200 ‘not good enough’ inspectors. *BBC News*. Retrieved from <https://www.bbc.co.uk/news/education-33198707>

Richmond, T. (2019). Requires improvement. A new role for Ofsted and school inspections. *EDSK report*. Retrieved from <https://www.edsk.org/wp-content/uploads/2019/04/Requires-Improvement.pdf>

Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology*, *104*(2), 277–288. <https://doi.org/10.1037/a0030461>

Spielman, A. (2017). *HMCI’s commentary: New research into short school inspections*. GOV.UK. Retrieved from <https://www.gov.uk/government/speeches/hmcis-monthly-commentary-march-2017>

Steffensmeier, D., & Hebert, C. (1999). Women and men policymakers: Does the judge’s gender affect the sentencing of criminal defendants? *Social Forces*, *77*(3), 1163–1196. <https://doi.org/10.2307/3005975>

Süß, S., & Kleiner, M. (2007). The psychological relationship between companies and freelancers: An empirical study of the commitment and the work-related expectations of freelancers. *Management Revue*, *18*(3), 251–270. <https://doi.org/10.5771/0935-9915-2007-3-251>

## Verbree, A. R.; Hornstra, L.; Maas, L. *&* Wijngaards-de Meij, L. (2022)*.* Conscientiousness as a Predictor of the Gender Gap in Academic Achievement. *Research in Higher Education*. <https://doi.org/10.1007/s11162-022-09716-5>

Villanueva-Moya, L., & Expósito, F. (2021). Gender differences in decision-making: The effects of gender stereotype threat moderated by sensitivity to punishment and fear of negative evaluation. *Journal of Behavioral Decision Making*, *34*(5), 706–717. <https://doi.org/10.1002/bdm.2239>

**Notes on contributors (author biograph)**

Christian Bokhove is a Professor in Mathematics Education at the University of Southampton. He is a specialist on international comparisons in mathematics education, the use of technology, and innovative methodologies.

John Jerrim is a Professor of Education and Social Statistics at University College London. He specialises in international comparisons of educational achievement and applied quantitative education research.

Sam Sims is an assistant professor at University College London. He specialises in research about teachers and applied quantitative education research.

**Table 1.** Descriptive statistics for the distribution of inspector characteristics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Primary** | | **Secondary** | |
|  | **Short** | **Not short** | **Short** | **Not short** |
| **Lead inspector contract** |  |  |  |  |
| Her Majesty’s Inspector | 59% | 20% | 60% | 45% |
| Ofsted Inspector | 41% | 80% | 40% | 55% |
| **Lead inspector gender** |  |  |  |  |
| Female | 54% | 48% | 44% | 43% |
| Male | 45% | 51% | 55% | 56% |
| unknown | 0% | 0% | 0% | 0% |
| **Primary / secondary specialism** |  |  |  |  |
| Primary inspections only | 65% | 73% | 0% | 0% |
| 70-99% primary | 19% | 17% | 7% | 13% |
| 30-69% primary | 14% | 9% | 38% | 40% |
| Secondary inspections only | 1% | 1% | 55% | 47% |
| **Inspection outside home region** |  |  |  |  |
| Yes | 2% | 16% | 3% | 15% |
| No | 90% | 67% | 83% | 62% |
| Not available | 7% | 17% | 14% | 23% |
| **Academic year** |  |  |  |  |
| 2011/12 | 0% | 19% | 0% | 15% |
| 2012/13 | 0% | 23% | 0% | 21% |
| 2013/14 | 0% | 19% | 0% | 17% |
| 2014/15 | 0% | 15% | 0% | 14% |
| 2015/16 | 12% | 6% | 18% | 7% |
| 2016/17 | 33% | 5% | 36% | 7% |
| 2017/18 | 34% | 6% | 32% | 9% |
| 2018/19 | 21% | 7% | 15% | 9% |
| **Previous inspections led** |  |  |  |  |
| mean | 33 | 29 | 19 | 19 |
| Standard deviation | 30 | 28 | 17 | 20 |
| minimum | 1 | 1 | 1 | 1 |
| 25th percentile | 11 | 8 | 6 | 5 |
| 50th percentile | 25 | 20 | 14 | 12 |
| 75th percentile | 43 | 42 | 26 | 26 |
| maximum | 186 | 182 | 103 | 161 |
| **Team size** |  |  |  |  |
| 1 inspector | 83% | 28% | 12% | 7% |
| 2 inspectors | 6% | 35% | 56% | 7% |
| 3 inspectors | 6% | 33% | 8% | 26% |
| 4 inspectors | 4% | 4% | 9% | 44% |
| 5 | 1% | 0% | 15% | 15% |
| ***n*** | **8,329** | **21,521** | **1,199** | **4,747** |

**Table 2.** Crosstabulations between characteristics of the lead inspector and overall effectiveness judgements.

1. Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | **Female** | **Male** | **Difference** | **Female** | **Male** | **Difference** |
| Outstanding | 7.8 | 8.2 | 0.4 | 10.9 | 10.1 | -0.9 |
| Good | 55.9 | 58.7 | 2.9 | 45.4 | 44.9 | -0.5 |
| Requires improvement | 30.5 | 28.6 | -1.9 | 34.6 | 34.6 | -0.1 |
| Inadequate | 5.9 | 4.5 | -1.4 | 9.1 | 10.5 | 1.4 |
| ***n (inspections)*** | **11,056** | **11,698** |  | **2,188** | **2,813** |  |

1. Contract status

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | **OI** | **HMI** | **Difference** | **OI** | **HMI** | **Difference** |
| Outstanding | 7.7 | 9.0 | -1.3 | 10.5 | 10.3 | -0.2 |
| Good | 60.3 | 47.0 | -13.3 | 47.8 | 42.2 | -5.6 |
| Requires improvement | 27.8 | 35.4 | 7.7 | 33.9 | 35.4 | 1.6 |
| Inadequate | 4.2 | 8.6 | 4.4 | 7.8 | 12.1 | 4.3 |
| ***n (inspections)*** | **17,622** | **5,139** |  | **2,654** | **2,370** |  |

1. Inspection outside of home region

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | **No** | **Yes** | **Difference** | **No** | **Yes** | **Difference** |
| Outstanding | 8.4 | 7.4 | -1.0 | 9.1 | 14.7 | 5.6 |
| Good | 56.8 | 58.9 | 2.1 | 44.6 | 44.5 | -0.1 |
| Requires improvement | 29.5 | 28.5 | -1.0 | 35.7 | 33.6 | -2.1 |
| Inadequate | 5.2 | 5.2 | 0.0 | 10.6 | 7.2 | -3.4 |
| ***n (inspections)*** | **15,925** | **3,347** |  | **3,161** | **735** |  |

1. Primary / secondary specialism

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | **Split** | **Primary only** |  | **Split** | **Second only** |  |
| Outstanding | 11.0 | 7.2 | -3.8 | 10.2 | 11.1 | 0.9 |
| Good | 52.9 | 58.2 | 5.3 | 43.9 | 46.1 | 2.2 |
| Requires improvement | 30.1 | 29.5 | -0.6 | 35.3 | 33.1 | -2.2 |
| Inadequate | 6.0 | 5.0 | -1.0 | 10.6 | 9.8 | -0.8 |
| ***n (inspections)*** | **1,912** | **15,871** |  | **1,976** | **2,308** |  |

Notes: OI = Ofsted inspector. HMI = Her Majesty’s Inspector.

**Table 3.** Differences in inspection assignments by gender and contract status of the lead inspector (primary schools).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Gender** | | **Her Majesty’s Inspector** | |
|  | **Female** | **Male** | **No** | **Yes** |
| **Inspection type** |  |  |  |  |
| Section 5 | 68% | 70% | 74% | 49% |
| Requires Improvement reinspection | 20% | 18% | 18% | 22% |
| Academy first Section 5 | 5% | 4% | 3% | 10% |
| Section 8 deemed Section 5 | 4% | 4% | 4% | 7% |
| Serious weakness inspection | 1% | 1% | 1% | 4% |
| Exempt school inspection | 2% | 2% | 0% | 8% |
| Section 8 no formal designation | 0% | 0% | 0% | 1% |
| Missing | 0% | 0% | 0% | 0% |
| **Prior inspection rating** |  |  |  |  |
| Outstanding | 8% | 8% | 7% | 13% |
| Good | 41% | 40% | 42% | 37% |
| Requires improvement | 43% | 45% | 47% | 32% |
| Inadequate | 4% | 4% | 2% | 14% |
| Missing | 3% | 3% | 3% | 5% |
| **FSM quintile** |  |  |  |  |
| Quintile 1 (Low FSM) | 16% | 17% | 17% | 15% |
| Quintile 2 | 19% | 20% | 20% | 18% |
| Quintile 3 | 21% | 21% | 21% | 20% |
| Quintile 4 | 22% | 22% | 21% | 24% |
| Quintile 5 (High FSM) | 22% | 20% | 20% | 23% |
| Missing | 0% | 0% | 0% | 0% |
| **School absence quintile** |  |  |  |  |
| Quintile 1 (low absences) | 20% | 21% | 21% | 19% |
| Quintile 2 | 23% | 22% | 23% | 21% |
| Quintile 3 | 23% | 23% | 23% | 24% |
| Quintile 4 | 21% | 21% | 21% | 22% |
| Quintile 5 (high absences) | 13% | 13% | 13% | 14% |
| Missing | 0% | 0% | 0% | 0% |
| **Key Stage 2 English quintile** |  |  |  |  |
| Quintile 1 (low achievement) | 24% | 24% | 23% | 31% |
| Quintile 2 | 21% | 20% | 21% | 20% |
| Quintile 3 | 17% | 17% | 18% | 14% |
| Quintile 4 | 17% | 16% | 17% | 14% |
| Quintile 5 (high achievement) | 12% | 13% | 13% | 9% |
| Missing | 9% | 10% | 10% | 10% |
| **Key Stage 2 maths quintile** |  |  |  |  |
| Quintile 1 (low achievement) | 24% | 23% | 22% | 31% |
| Quintile 2 | 19% | 20% | 20% | 18% |
| Quintile 3 | 20% | 18% | 19% | 20% |
| Quintile 4 | 15% | 15% | 16% | 12% |
| Quintile 5 (high achievement) | 12% | 13% | 13% | 10% |
| Missing | 9% | 10% | 10% | 10% |

**Table 4.** Ordinal logistic regression model results of the association between lead inspector characteristics and Overall Effectiveness grades.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Primary** | | **Secondary** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Gender (Ref: Male)** |  |  |  |  |
| Female | 0.84\* | -3.18 | 1.09 | 1.16 |
| **Contract status (Ref: Ofsted Inspector)** |  |  |  |  |
| Her Majesty’s Inspector | 1.45\* | 6.21 | 1.32\* | 3.49 |
| **Experience (Ref: Bottom quintile)** |  |  |  |  |
| Quintile 2 | 0.94 | -0.6 | 0.72 | -1.95 |
| Quintile 3 | 1.00 | 0.03 | 0.71\* | -2.13 |
| Quintile 4 | 1.07 | 0.69 | 0.82 | -1.08 |
| Quintile 5 (most experienced) | 0.93 | -0.64 | 0.89 | -0.53 |
| **Inspection outside home region (Ref: No)** |  |  |  |  |
| Yes | 1.13\* | 2.50 | 0.99 | -0.14 |
| **Phase specialism (ref: primary/secondary only)** |  |  |  |  |
| 30-69% primary | 0.86 | -1.91 | 0.92 | -1.06 |
| 70-99% primary | 0.95 | -0.83 | 1.03 | 0.21 |
| **Team size** |  |  |  |  |
| 1 inspector | Reference | | 0.43\* | -4.92 |
| 2 inspectors | 1.25\* | 5.29 | 0.90 | -0.93 |
| 3 inspectors | 1.26\* | 5.23 | 1.08 | 0.93 |
| 4 inspectors | 1.05 | 0.56 | Reference | |
| 5+ inspectors | 0.68 | -1.60 | 0.83\* | -2.10 |

Notes: Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Estimates based on 22,754 inspections conducted by 983 inspectors. Standard errors have been clustered at the inspector level. Models control for percent of pupils eligible for FSM, school religion, school gender, region, inspection type, prior Ofsted rating, school performance data, school absences, percent of pupils at the school with Special Educational Needs, percent of pupils who speak English as an Additional Language and other inspector characteristics.

\**p* < .05.

**Table 5.** Logistic regression model estimates of the association between lead inspector characteristics and primary school short inspection outcomes.

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***n*** | ***OR*** | ***t*** |
| **Gender (Ref: Male)** |  |  |  |
| Female | 3,605 | 0.75\* | 2.17 |
| **Contract (Ref: Ofsted inspector)** |  |  |  |
| Her Majesty’s Inspector | 3,627 | 1.63\* | 3.56 |
| **Experience (Ref: Bottom quintile)** |  |  |  |
| Quintile 2 | 3,627 | 0.65 | -1.85 |
| Quintile 3 | 0.84 | -0.73 |
| Quintile 4 | 1.05 | 0.22 |
| Quintile 5 (most experienced) | 1.22 | 0.92 |
| **Inspection outside home region (Ref: No)** |  |  |  |
| Yes | 3,311 | 1.41 | 1.08 |
| **Phase specialism (ref: primary only)** |  |  |  |
| 30-69% primary | 3,594 | 1.41 | 1.64 |
| 70-99% primary | 1.36 | 1.77 |
| **Team size (Ref: 1 inspector)** |  |  |  |
| 2 inspectors | 3,511 | 0.99 | -0.05 |

Notes: Sample restricted to short primary school inspections conducted between January 2018 and August 2019. Dependent variable coded 1 if the short inspection resulted in the outcome “S5 next due to concerns” or was immediately converted to a full inspection resulting in a judgement of Inadequate or Requires Improvement, and zero otherwise. Models have been estimated separately for each characteristic. Models control for percent of pupils eligible for FSM, school religion, school gender, region, inspection type, prior Ofsted rating, school performance data, school absences, percent of pupils at the school with Special Educational Needs, percent of pupils who speak English as an Additional Language and other inspector characteristics.

\**p* < .05.

**Table 6.** Cross-tabulation between inspection team size and inspection outcomes.

1. Primary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Team size** | | | | |
|  | **1** | **2** | **3** | **4** | **5+** |
| **Overall effectiveness** |  |  |  |  |  |
| Outstanding | 9 | 7 | 8 | 14 | 26 |
| Good | 61 | 59 | 55 | 46 | 30 |
| Requires improvement | 27 | 29 | 31 | 33 | 41 |
| Inadequate | 3 | 5 | 7 | 8 | 3 |
| ***n*** | **5,546** | **7,184** | **7,158** | **1,093** | **150** |

1. Secondary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Team size** | | | | |
|  | **1** | **2** | **3** | **4** | **5+** |
| **Overall effectiveness** |  |  |  |  |  |
| Outstanding | 22 | 9 | 7 | 10 | 15 |
| Good | 43 | 47 | 46 | 44 | 46 |
| Requires improvement | 30 | 36 | 35 | 35 | 32 |
| Inadequate | 4 | 9 | 12 | 11 | 7 |
| ***n*** | **233** | **273** | **1,148** | **2,072** | **889** |

Notes: Figures refer to column percentages.

**Table 7.** Predicted distribution of primary school inspection outcomes for two hypothetical inspectors. Ordinal logistic regression model estimates.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Inspector A** | **Inspector B** | **Risk ratio (A/B)** |
| **Overall effectiveness** |  |  |  |
| Outstanding | 4.5% | 9.0% | 0.50 |
| Good | 48.0% | 59.3% | 0.81 |
| Requires improvement | 38.4% | 27.2% | 1.41 |
| Inadequate | 9.1% | 4.5% | 2.03 |
| **Short inspection** |  |  |  |
| Conversion with downgrade or S5 next due to concerns. (Jan18 - Aug19) | 15.5% | 9.7% | 1.60 |
| **Inspector characteristics** |  |  |  |
| Team size | 2 inspectors | 1 inspector |  |
| Contract status | HMI | OI |  |
| Gender | Female | Male |  |

Notes: Model controls 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, whether the inspection was conducted after 2018, school religion, school gender composition, Key Stage 1 scores and percent of pupils with special educational needs.

**Appendix 1. Inspection type of academic year. 2011/12 – 2018/19.**

**Appendix Table 1.1.** Cross-tabulation between inspection type by academic year. Primary inspections.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Inspection type** | **2011/12** | **2012/13** | **2013/14** | **2014/15** | **2015/16** | **2016/17** | **2017/18** | **2018/19** |
| Academy First Section 5 | 0 | 4 | 164 | 238 | 18 | 250 | 209 | 152 |
| Exempt School Inspection | 0 | 0 | 0 | 0 | 5 | 96 | 68 | 280 |
| Maintained Academy and School Short inspection | 0 | 0 | 0 | 0 | 1,091 | 2,903 | 3,034 | 1,906 |
| Notice to Improve Section 5 Reinspection | 101 | 131 | 0 | 0 | 0 | 0 | 0 | 0 |
| Requires Improvement Section 5 Reinspection Visit 1 | 0 | 0 | 0 | 0 | 1,139 | 535 | 177 | 365 |
| Requires Improvement Section 5 Reinspection Visit 2 | 0 | 0 | 0 | 0 | 96 | 217 | 122 | 156 |
| Requires Improvement Section 5 Reinspection Visit 3 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 14 |
| Requires Improvement monitoring Visit 1 | 0 | 0 | 0 | 0 | 275 | 259 | 234 | 83 |
| Requires Improvement monitoring Visit 2 | 0 | 0 | 0 | 0 | 49 | 22 | 6 | 8 |
| Requires Improvement monitoring Visit 3 | 0 | 0 | 0 | 0 | 2 | 3 | 0 | 0 |
| Section 5 Inspection | 4,370 | 5,588 | 3,407 | 1,924 | 9 | 97 | 888 | 855 |
| Section 5 Requires Improvement 1st Re-Inspection | 0 | 0 | 567 | 1,079 | 1 | 0 | 0 | 0 |
| Section 5 Requires Improvement 2nd Re-Inspection | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Section 5 Serious Weaknesses Re-Inspection | 0 | 0 | 29 | 37 | 0 | 0 | 0 | 0 |
| Section 8 Deemed Section 5 | 147 | 115 | 649 | 360 | 0 | 0 | 0 | 0 |
| Section 8 No Formal Designation Visit | 0 | 0 | 0 | 0 | 48 | 40 | 51 | 33 |
| Schools into Special Measures Visit 1 | 0 | 0 | 0 | 0 | 50 | 43 | 61 | 52 |
| Schools into Special Measures Visit 2 | 0 | 0 | 0 | 0 | 55 | 14 | 45 | 15 |
| Schools into Special Measures Visit 3 | 0 | 0 | 0 | 0 | 58 | 7 | 28 | 16 |
| Schools into Special Measures Visit 4 | 0 | 0 | 0 | 0 | 54 | 4 | 8 | 16 |
| Schools into Special Measures Visit 5 | 0 | 0 | 0 | 0 | 33 | 5 | 5 | 5 |
| Schools with Serious Weaknesses Visit 1 | 0 | 0 | 0 | 0 | 7 | 5 | 22 | 30 |
| Schools with Serious Weaknesses Visit 2 | 0 | 0 | 0 | 0 | 15 | 2 | 4 | 3 |
| Schools with Serious Weaknesses Visit 3 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 |
| Section 8 Inspection due to Parental Complaint | 0 | 0 | 0 | 0 | 22 | 14 | 10 | 4 |
| Serious Weaknesses Section 5 Reinspection | 0 | 0 | 0 | 0 | 19 | 3 | 3 | 13 |
| Special Measures Section 5 Reinspection | 18 | 8 | 10 | 17 | 14 | 6 | 7 | 22 |
| ***n*** | **4,636** | **5,846** | **4,826** | **3,656** | **3,064** | **4,525** | **4,985** | **4,028** |

**Appendix Table 1.2.** Cross-tabulation between inspection type by academic year. Secondary inspections.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Inspection type** | **2011/12** | **2012/13** | **2013/14** | **2014/15** | **2015/16** | **2016/17** | **2017/18** | **2018/19** |
| Academy First Section 5 | 70 | 109 | 67 | 72 | 6 | 52 | 64 | 52 |
| Exempt School Inspection | - | - | - | - | 2 | 16 | 10 | 68 |
| Maintained Academy and School Short inspection | - | - | - | - | 243 | 470 | 418 | 193 |
| Notice to Improve Section 5 Reinspection | 38 | 52 | 1 | - | - | - | - | - |
| Requires Improvement Section 5 Reinspection Visit 1 | - | - | - | - | 263 | 159 | 109 | 104 |
| Requires Improvement Section 5 Reinspection Visit 2 | - | - | - | - | 36 | 79 | 52 | 54 |
| Requires Improvement Section 5 Reinspection Visit 3 | - | - | - | - | - | - | - | 13 |
| Requires Improvement monitoring Visit 1 | - | - | - | - | 119 | 89 | 89 | 69 |
| Requires Improvement monitoring Visit 2 | - | - | - | - | 32 | 19 | 2 | 2 |
| Requires Improvement monitoring Visit 3 | - | - | - | - | 4 | - | - | - |
| Section 5 Inspection | 784 | 1,152 | 645 | 302 | 11 | 73 | 232 | 194 |
| Section 5 Requires Improvement 1st Re-Inspection | - | - | 107 | 315 | - | - | - | - |
| Section 5 Requires Improvement 2nd Re-Inspection | - | - | - | 1 | - | - | - | - |
| Section 5 Serious Weaknesses Re-Inspection | - | - | 25 | 28 | - | - | - | - |
| Section 8 Deemed Section 5 | 30 | 23 | 203 | 140 | - | - | - | - |
| Section 8 No Formal Designation Visit | - | - | - | - | 35 | 52 | 58 | 47 |
| Schools into Special Measures Visit 1 | - | - | - | - | 35 | 38 | 60 | 34 |
| Schools into Special Measures Visit 2 | - | - | - | - | 36 | 24 | 42 | 23 |
| Schools into Special Measures Visit 3 | - | - | - | - | 57 | 19 | 25 | 14 |
| Schools into Special Measures Visit 4 | - | - | - | - | 55 | 17 | 16 | 15 |
| Schools into Special Measures Visit 5 | - | - | - | - | 53 | 15 | 7 | 7 |
| Schools with Serious Weaknesses Visit 1 | - | - | - | - | 9 | 13 | 16 | 24 |
| Schools with Serious Weaknesses Visit 2 | - | - | - | - | 11 | 2 | 3 | 3 |
| Schools with Serious Weaknesses Visit 3 | - | - | - | - | 9 | - | 1 | - |
| Section 8 Inspection due to Parental Complaint | - | - | - | - | 25 | 13 | 16 | 9 |
| Serious Weaknesses Section 5 Reinspection | - | - | - | - | 22 | 8 | 8 | 14 |
| Special Measures Section 5 Reinspection | 4 | 1 | - | 12 | 26 | 11 | 8 | 17 |
| ***n*** | **926** | **1,337** | **1,048** | **870** | **1,089** | **1,169** | **1,236** | **956** |

**Appendix 2. Sample selection**

Primary inspections

Between September 2011 and August 2019 there were 35,566 primary inspections conducted, based upon the management information published on the Ofsted website[[11]](#footnote-11). We have extracted information on the lead inspector from 29,850 (84%) of these inspections from the “Watchsted” website. Of the remaining 5,716 inspections, we can access information on the inspectors involved in the inspection from 3,776 via our own scraping of the published Ofsted reports. Thus, when added together, we can observe information on the lead inspector from (29,850 + 3,776) / 35,566 = 94.5% of all primary inspections conducted between September 2011 and August 2019[[12]](#footnote-12). Moreover, of the 1,940 (5.5%) of primary inspections we have been unable to match, 1,489 are Require Improvement monitoring visits or special measures/serious weakness visits. Importantly, most (1,664 – 86%) of the 1,940 unmatched did not lead to an overall effectiveness judgement. Together, this provides reassurance that we have managed to access the relevant information on the vast majority of primary inspections conducted over this period, that issues of missing / unlinked data are limited, and that our analytic sample is representative of the population of primary inspections conducted over this period.

Secondary inspections

Between September 2011 and August 2019 there were 8,631 secondary inspections conducted, based upon the management information published on the Ofsted website. We have extracted information on the lead inspector from 5,901 (68%) of these inspections from the “Watchsted” website. Of the remaining 2,730 inspections, we can access information on the inspectors involved in the inspection from 1,432 via our own scraping of the published Ofsted reports. Thus, when added together, we can observe information on the lead inspector from (5,901 + 1,432) / 8,631 = 85% of all secondary inspections conducted between September 2011 and August 2019. Moreover, of the 1,298 (15%) of secondary inspections we have been unable to match, 1,009 are Requires Improvement monitoring visits or special measures/serious weakness visits. Importantly, most (1,166 – 90%) of the 1,298 unmatched did not lead to an overall effectiveness judgement. Together, this provides reassurance that we have managed to access the relevant information on the vast majority of secondary inspections conducted over this period, that issues of missing / unlinked data are limited, and that our analytic sample is representative of the population of secondary inspections conducted over this period.

Robustness test to using an alternative sample

In the main body of the paper we present results based upon data we have extracted from the Watchsted database alone. However, as noted above, we have also performed our own scraping of inspector names from the published Ofsted inspection reports, which we can add onto the Watchsted database. Tables 2.1 (primary) and 2.2 (secondary) below provides a comparison of the results across these two analytic samples. The estimated odds ratios and associated t-statistics are very similar, regardless of which sample is used. In other words, the estimates reported in the main text appear robust to further extension of our analytic sample via adding in data from our own scrapping of inspector names into the Watchsted database.

**Appendix Table 2.1.** A comparison of ordinal logistic regression estimates across alternative sample selections. Estimates for primary schools.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Main sample** | | **Alternative sample** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Gender (Ref: female)** |  |  |  |  |
| Male | 0.86\* | -2.83 | 0.85\* | -3.32 |
| **Contract (Ref: Ofsted inspector)** |  |  |  |  |
| Her Majesty’s Inspector | 1.43\* | 6.24 | 1.41\* | 6.21 |
| **Outside home region (Ref: No)** |  |  |  |  |
| Yes | 1.12\* | 2.22 | 1.11\* | 2.18 |
| **Phase specialism (Ref: primary only)** |  |  |  |  |
| 30-69% primary | 0.88 | -1.71 | 0.96 | -0.48 |
| 70-99% primary | 0.96 | -0.55 | 1.00 | -0.01 |
| Secondary only | 1.27 | 1.21 | 0.75 | -1.74 |
| **Experience (Ref: Quintile 1)** |  |  |  |  |
| Quintile 2 | 0.98 | -0.52 | 1.00 | 0.09 |
| Quintile 3 | 0.86\* | -3.07 | 0.86\* | -2.99 |
| Quintile 4 | 0.97 | -0.57 | 0.93 | -1.32 |
| Quintile 5 | 0.88\* | -1.99 | 0.88\* | -1.98 |
| **Team size (Ref: 1 inspector)** |  |  |  |  |
| 2 inspectors | 1.20\* | 4.63 | 1.17\* | 4.41 |
| 3 inspectors | 1.17\* | 3.94 | 1.15\* | 3.80 |
| 4 inspectors | 0.96 | -0.53 | 0.92 | -1.04 |
| 5 inspectors | 0.57\* | -2.28 | 0.60\* | -2.24 |
| ***n (Inspections)*** | 22,743 | | 25,936 | |
| ***n (Inspectors)*** | 986 | | 1,407 | |

Notes: Models include controls for percent of pupils eligible for FSM, Ofsted region, previous inspection rating, inspection type and Key Stage 2 English and mathematics test scores.

\**p* < .05.

**Appendix Table 2.2.** A comparison of ordinal logistic regression estimates across alternative sample selections. Estimates for secondary schools.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Main sample** | | **Alternative sample** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Gender (Ref: female)** |  |  |  |  |
| Male | 1.08 | 1.08 | 1.05 | 0.76 |
| **Contract (Ref: Ofsted inspector)** |  |  |  |  |
| Her Majesty’s Inspector | 1.24\* | 2.72 | 1.18\* | 2.24 |
| **Outside home region (Ref: No)** |  |  |  |  |
| Yes | 1.03 | 0.30 | 1.00 | 0.00 |
| **Phase specialism (Ref: secondary only)** |  |  |  |  |
| 30-69% primary | 0.92 | -0.85 | 1.01 | 0.14 |
| 70-99% primary | 0.96 | -0.34 | 1.10 | 0.89 |
| **Experience (Ref: Quintile 1)** |  |  |  |  |
| Quintile 2 | 0.89 | -1.39 | 0.88 | -1.64 |
| Quintile 3 | 0.83\* | -1.99 | 0.77\* | -3.03 |
| Quintile 4 | 0.88 | -1.23 | 0.79\* | -2.18 |
| Quintile 5 | 0.94 | -0.49 | 0.83 | -1.47 |
| **Team size (Ref: 4 inspectors)** |  |  |  |  |
| 1 inspector | 0.37\* | -6.34 | 0.51\* | -5.12 |
| 2 inspectors | 0.87 | -0.97 | 0.93 | -0.69 |
| 3 inspectors | 1.11 | 1.27 | 1.02 | 0.22 |
| 5 inspectors | 0.83\* | -2.17 | 0.80\* | -2.83 |
| ***n (Inspections)*** | 4,899 | | 6,191 | |
| ***n (Inspectors)*** | 564 | | 733 | |

Notes: Models include controls for percent of pupils eligible for FSM, religious denomination of the school, gender composition of school, Ofsted region, previous overall inspection rating, inspection type, Key Stage 2 scores of intake, percent of pupils achieving five A\*-C grades, Key Stage 4 total points score, Progress 8 scores, percent of pupils absent, percent of pupils who speak English as an Additional Language, percent of pupils with special educational needs.

\**p* < .05.

**Appendix 3. Manual checks of the data**

To check the quality of the data, we have performed some “manual” checks, returning to the initially published Ofsted reports to cross-reference the data we have extracted against.

To begin, we conducted a power calculation to understand the sample size required from our manual checks to give us a reasonable degree of accuracy. These power calculations were conducted assuming that there would be around 90% agreement between the Watchsted data (plus our automated inspector name extraction where the Watchsted data is missing) and our manual approach. These power calculations revealed that a sample size of 150 would yield a standard error of 2.4 percentage points[[13]](#footnote-13), and thus resulting in a confidence interval between 85% and 95%. We deemed this sufficient to understand the likely degree of measurement error within our data.

Two sets of random samples were drawn. The first random sample was 150 short inspections. The second was 150 not-short inspections (108 of these were an S5 inspection)[[14]](#footnote-14). For each of these 300 inspections, we attempted to find the relevant inspection report on the Ofsted website and manually recorded (a) the name of all inspectors (including the lead inspector) and (b) whether the lead inspector (or any other inspector) was an HMI. These are then used as a basis to check the quality of the full database we use in our analysis.

Non-short inspections

Of the 150 inspections in our initial random sample, the original inspection report was available from the Ofsted website on 138 (92%) occasions, for which we can manually check our data against. Of these, the name of the lead inspector matches on 134 (97%) of occasions (95% confidence interval spans 94% to 100%). Moreover, two of the four instances where the sources did not agree may be due to typos (“June Robinson” rather than “Jean Robinson” and “Christine Huard” rather than “Christine Howard”). The level of agreement for whether an HMI or OI led an inspection was also high (93% with a confidence interval spanning from 89% to 98%). In other words, the level of agreement is extremely high.

Short inspection results

All 150 of the short inspections in our random sample were found and accessed from the Ofsted website. Of these, the lead inspector matched on 145 (97%) of occasions. This is again a very high level of agreement and is reassuring regarding the quality of the data available.

The level of agreement of whether an HMI or OI led the short inspection was somewhat lower at 130 (86%) out of the 150 (confidence interval spanning from 80% to 91%). Further investigations of the data suggest that this may be due to individual inspectors changing contracts type over time (i.e. moving from being an OI to an HMI, or vice-versa). As the Watchsted database only includes a fixed flag at the inspector level for whether the named inspector is an HMI or not, this time dimension to contract status will not be captured.

We hence also investigate the level of agreement (for whether an HMI was involved in the inspection or not) between our manually extracted random sample and our own automated extraction of inspector names (and HMI status). An important advantage of our own extraction of inspector names (and HMI status) is that it has been done at the individual inspection level – and hence captures potential changes in OI/HMI status of individual inspectors over time.

Of the 150 short inspections in our random sample, we have managed to perform our own manual extraction successfully on 145 occasions. Of these, there was agreement on 141 occasions (97%) as to whether an HMI was involved in the inspection (confidence interval 95% to 100%)[[15]](#footnote-15). Hence data from our own automated extraction of inspector names – and, in particular, whether an HMI was involved in the inspection – provides a useful additional source of information that can be further used to investigate the robustness of our results (most notably, differences between HMI and OIs in short inspection outcomes).

**Appendix 4. Alternative estimates using multi-level modelling (random effects)**

In the main body of the paper we use ordinal logistic regression – with standard errors clustered by inspector – to examine the association between various inspector characteristics and school inspection outcomes. An alternative approach to taking account the “clustering” of inspections within lead inspectors would be to estimate a multilevel model (with inspections as the level 1 unit and lead inspectors as the second level). In this appendix, we explore the similarity of results under these two approaches, focusing on the results for Overall Effectiveness judgements.

Appendix Table 4.1 presents results from such a comparison of methodological approaches for primary schools, referring to a model that controls for percent of pupils eligible for FSM, Ofsted region, previous inspection rating, inspection type and Key Stage 2 English and mathematics test scores. All inspector characteristics are included in this model simultaneously. Figures on the left-hand side are from a multilevel (random effects) ordinal logistic regression model, while those on the right are from an ordinal logistic regression model with standard errors clustered within inspectors. Overall, parameter estimates (presented as odds ratios) and the associated t-statistics are very similar across the two approaches. The substantive conclusions reached are thus robust.

An analogous comparison across methodological approaches for secondary schools is presented in Appendix Table 4.2. The model used controls for gender composition of school, Ofsted region, previous overall inspection rating, inspection type, Key Stage 2 scores of intake, percent of pupils achieving five A\*-C grades, Key Stage 4 total points score, Progress 8 scores, percent of pupils absent, percent of pupils who speak English as an Additional Language, percent of pupils with special educational needs. Again, the estimated odds-ratios and the associated t-statistics do not substantive differ across the two approaches.

**Appendix 4.1.** A comparison of estimates from multilevel ordinal logistic regressions to ordinal logistic regression with clustered standard errors. Primary school results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Multi-level model** | | **Clustered SE** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Gender (Ref: female)** |  |  |  |  |
| Male | 0.88\* | -2.60 | 0.86\* | -2.83 |
| **Contract (Ref: Ofsted inspector)** |  |  |  |  |
| Her Majesty’s Inspector | 1.45\* | 6.35 | 1.43\* | 6.24 |
| **Outside home region (Ref: No)** |  |  |  |  |
| Yes | 1.16\* | 3.40 | 1.12\* | 2.22 |
| **Phase specialism (Ref: primary only)** |  |  |  |  |
| 30-69% primary | 0.88 | -1.61 | 0.88 | -1.71 |
| 70-99% primary | 0.94 | -0.95 | 0.96 | -0.55 |
| Secondary only | 1.32 | 1.07 | 1.27 | 1.21 |
| **Experience (Ref: Quintile 1)** |  |  |  |  |
| Quintile 2 | 0.96 | -0.83 | 0.98 | -0.52 |
| Quintile 3 | 0.84\* | -3.72 | 0.86\* | -3.07 |
| Quintile 4 | 0.97 | -0.71 | 0.97 | -0.57 |
| Quintile 5 | 0.83\* | -3.53 | 0.88\* | -1.99 |
| **Team size (Ref: 1 inspector)** |  |  |  |  |
| 2 inspectors | 1.21\* | 4.97 | 1.20\* | 4.63 |
| 3 inspectors | 1.19\* | 4.27 | 1.17\* | 3.94 |
| 4 inspectors | 0.98 | -0.25 | 0.96 | -0.53 |
| 5 inspectors | 0.57\* | -2.95 | 0.57\* | -2.28 |
| ***n (Inspections)*** | 22,743 | | 22,743 | |
| ***n (Inspectors)*** | 986 | | 986 | |

Notes: Models include controls for percent of pupils eligible for FSM, Ofsted region, previous inspection rating, inspection type and Key Stage 2 English and mathematics test scores.

\**p* < .05.

**Appendix 4.2.** A comparison of estimates from multilevel ordinal logistic regressions to ordinal logistic regression with clustered standard errors. Secondary school results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Multi-level model** | | **Clustered SE** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Gender (Ref: female)** |  |  |  |  |
| Male | 1.08 | 1.08 | 1.09 | 1.15 |
| **Contract (Ref: Ofsted inspector)** |  |  |  |  |
| Her Majesty’s Inspector | 1.24\* | 2.72 | 1.25\* | 2.77 |
| **Outside home region (Ref: No)** |  |  |  |  |
| Yes | 1.03 | 0.30 | 1.02 | 0.17 |
| **Phase specialism (Ref: secondary only)** |  |  |  |  |
| 30-69% primary | 0.92 | -0.85 | 0.91 | -1.03 |
| 70-99% primary | 0.96 | -0.34 | 1.01 | 0.08 |
| **Experience (Ref: Quintile 1)** |  |  |  |  |
| Quintile 2 | 0.89 | -1.39 | 0.92 | -1.00 |
| Quintile 3 | 0.83\* | -1.99 | 0.84 | -1.80 |
| Quintile 4 | 0.88 | -1.23 | 0.93 | -0.67 |
| Quintile 5 | 0.94 | -0.49 | 1.03 | 0.24 |
| **Team size (Ref: 1 inspector)** |  |  |  |  |
| 2 inspectors | 0.37\* | -6.34 | 0.38\* | -5.81 |
| 3 inspectors | 0.87 | -0.97 | 0.85 | -1.38 |
| 4 inspectors | 1.11 | 1.27 | 1.08 | 0.93 |
| 5 inspectors | 0.83\* | -2.17 | 0.83\* | -2.05 |
| ***n (Inspections)*** | 4,899 | | 4,899 | |
| ***n (Inspectors)*** | 564 | | 564 | |

Notes: Models include controls for percent of pupils eligible for FSM, religious denomination of the school, gender composition of school, Ofsted region, previous overall inspection rating, inspection type, Key Stage 2 scores of intake, percent of pupils achieving five A\*-C grades, Key Stage 4 total points score, Progress 8 scores, percent of pupils absent, percent of pupils who speak English as an Additional Language, percent of pupils with special educational needs.

\**p* < .05.

**Appendix 5. Sub-group ordinal logistic regression estimates for gender and contract status**

Gender

**Table 5.1.** The link between inspector gender and primary school inspection outcomes. Ordinal regression estimates for sub-groups.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | ***n*** | ***OR*** | ***t*** | ***n*** | ***OR*** | ***t*** |
| **Academic year** |  |  |  |  |  |  |
| 2011/12 | 3,587 | 0.79\* | -2.45 | 583 | 0.73 | -1.53 |
| 2012/13 | 5,092 | 0.80\* | -2.43 | 1,034 | 1.09 | 0.56 |
| 2013/14 | 4,272 | 1.00 | -0.02 | 800 | 1.10 | 0.64 |
| 2014/15 | 3,286 | 0.91 | -1.07 | 693 | 1.25 | 1.20 |
| 2015/16 | 1,550 | 0.84 | -1.39 | 406 | 1.33 | 1.15 |
| 2016/17 | 1,806 | 0.78\* | -2.37 | 500 | 1.29 | 1.26 |
| 2017/18 | 1,528 | 0.92 | -0.72 | 505 | 1.34 | 1.39 |
| 2018/19 | 1,621 | 0.84 | -1.65 | 436 | 1.00 | -0.01 |
| **Contract type** |  |  |  |  |  |  |
| Ofsted inspector (OI) | 17,617 | 0.87\* | -2.10 | 2,625 | 1.18 | 1.50 |
| Her Majesty's Inspector (HMI) | 5,126 | 0.83\* | -2.36 | 2,333 | 0.99 | -0.06 |
| **Inspection type** |  |  |  |  |  |  |
| Section 5 inspection | 14,876 | 0.87\* | -2.29 | 2,595 | 1.10 | 0.98 |
| **Ofsted region** |  |  |  |  |  |  |
| East Midlands | 2,320 | 0.76\* | -2.04 | 452 | 0.93 | -0.27 |
| East of England | 2,788 | 0.97 | -0.24 | 593 | 1.20 | 0.87 |
| London | 2,192 | 0.79\* | -2.10 | 568 | 1.19 | 0.89 |
| North East, Yorkshire and Humber | 3,675 | 0.86 | -1.14 | 746 | 1.02 | 0.08 |
| North West | 3,414 | 0.80\* | -2.12 | 740 | 1.26 | 1.43 |
| South East | 3,454 | 0.86 | -1.16 | 711 | 0.92 | -0.49 |
| South West | 2,256 | 0.96 | -0.27 | 510 | 1.06 | 0.20 |
| West Midlands | 2,644 | 0.98 | -0.21 | 638 | 1.10 | 0.46 |

Notes: Estimates based upon ordered logistic regression models. The models control for percentage of children eligible for free school meals, Ofsted region, previous inspection rating, inspection type, school performance measures, whether the inspector is an HMI and total amount of inspection experience. Separate models have been estimated for each sub-group. Standard errors have been clustered at the inspector level.

\**p* < .05.

Contract status

**Table 5.2.** The link between inspector contractual status and inspection outcomes. Ordinal regression estimates for sub-groups.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | ***n*** | ***OR*** | ***t*** | ***n*** | ***OR*** | ***t*** |
| **Academic year** |  |  |  |  |  |  |
| 2011/12 | 3,587 | 1.93\* | 3.95 | 583 | 2.69\* | 4.78 |
| 2012/13 | 5,092 | 1.40\* | 2.74 | 1,034 | 0.79 | -1.35 |
| 2013/14 | 4,272 | 2.11\* | 6.08 | 800 | 2.26\* | 3.94 |
| 2014/15 | 3,286 | 1.53\* | 2.84 | 693 | 2.40\* | 3.56 |
| 2015/16 | 1,550 | 1.31 | 1.8 | 406 | 0.95 | -0.20 |
| 2016/17 | 1,806 | 1.38\* | 2.69 | 500 | 0.91 | -0.47 |
| 2017/18 | 1,528 | 1.50\* | 3.28 | 505 | 1.08 | 0.40 |
| 2018/19 | 1,621 | 1.09 | 0.64 | 436 | 0.96 | -0.17 |
| **Inspection type** |  |  |  |  |  |  |
| Section 5 inspection | 14,876 | 1.54\* | 5.84 | 2,595 | 1.21\* | 1.97 |
| **Ofsted region** |  |  |  |  |  |  |
| East Midlands | 2,320 | 1.46\* | 2.81 | 452 | 1.28 | 0.80 |
| East of England | 2,788 | 1.20 | 0.99 | 593 | 1.05 | 0.20 |
| London | 2,192 | 1.09 | 0.54 | 568 | 1.62\* | 2.70 |
| North East, Yorkshire and Humber | 3,675 | 1.49\* | 2.93 | 746 | 1.33 | 1.31 |
| North West | 3,414 | 1.49\* | 2.97 | 740 | 1.10 | 0.51 |
| South East | 3,454 | 1.09 | 0.58 | 711 | 1.00 | 0.01 |
| South West | 2,256 | 2.23\* | 4.93 | 510 | 1.43 | 1.37 |
| West Midlands | 2,644 | 1.74\* | 3.93 | 638 | 1.04 | 0.19 |

Notes: Estimates based upon ordered logistic regression models controlling for percentage of children eligible for free school meals, Ofsted region, previous inspection rating, inspection type, school performance measures, whether the inspector is male and total amount of inspection experience. Separate models have been estimated for each sub-group. Standard errors have been clustered at the inspector level.

\**p* < .05.

**Appendix 6. Multinomial logistic regression model estimates**

Gender

**Table 6.1.** The link between inspector gender and primary school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Impact of female inspector (Ref: Good)** |  |  |  |  |  |  |
| Outstanding | 1.00 | -0.05 | 1.04 | 0.62 | 1.05 | 0.69 |
| Requires improvement | 0.89\* | -2.11 | 0.88\* | -2.20 | 0.89\* | -2.03 |
| Inadequate | 0.73\* | -3.28 | 0.71\* | -3.44 | 0.73\* | -3.17 |
| **Inspection-level controls** |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | |
| Inspection type | - | | Y | | Y | |
| Prior Ofsted rating | - | | Y | | Y | |
| School performance data | - | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |
| Her Majesty’s Inspector | - | | - | | Y | |

1. Predicted probabilities

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | |
|  | **Female** | **Male** | **Female** | **Male** | **Female** | **Male** |
| **Impact of female inspector** |  |  |  |  |  |  |
| Outstanding | 7.8% | 8.2% | 7.7% | 8.2% | 7.7% | 8.2% |
| Good | 55.9% | 58.7% | 55.6% | 58.6% | 56.1% | 58.5% |
| Requires improvement | 30.5% | 28.6% | 30.4% | 28.7% | 30.4% | 28.7% |
| Inadequate | 5.9% | 4.5% | 5.8% | 4.5% | 5.8% | 4.5% |
| **Inspection-level controls** |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | |
| Inspection type | - | | Y | | Y | |
| Prior Ofsted rating | - | | Y | | Y | |
| School performance data | - | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |
| Her Majesty’s Inspector | - | | - | | Y | |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level. Models based upon 22,736 inspections conducted by 983 inspectors.

\**p* < .05.

**Table 6.2.** The link between inspector gender and secondary school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Impact of female inspector (Ref: Good)** |  |  |  |  |  |  |
| Outstanding | 0.93 | -0.64 | 0.99 | -0.10 | 0.99 | -0.06 |
| Requires improvement | 1.01 | 0.11 | 1.01 | 0.08 | 1.00 | -0.03 |
| Inadequate | 1.16 | 1.15 | 1.16 | 1.19 | 1.11 | 0.86 |
| **Inspection-level controls** |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | |
| Inspection type | - | | Y | | Y | |
| Prior Ofsted rating | - | | Y | | Y | |
| School performance data | - | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |
| Her Majesty’s Inspector | - | | - | | Y | |

1. Predicted probabilities

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | |
|  | **Female** | **Male** | **Female** | **Male** | **Female** | **Male** |
| **Impact of female inspector** |  |  |  |  |  |  |
| Outstanding | 10.9% | 10.1% | 10.5% | 10.4% | 10.5% | 10.4% |
| Good | 45.4% | 44.9% | 45.5% | 44.9% | 45.3% | 45.0% |
| Requires improvement | 34.6% | 34.6% | 34.8% | 34.3% | 34.8% | 34.3% |
| Inadequate | 9.1% | 10.5% | 9.2% | 10.4% | 9.4% | 10.2% |
| **Inspection-level controls** |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | |
| Inspection type | - | | Y | | Y | |
| Prior Ofsted rating | - | | Y | | Y | |
| School performance data | - | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |
| Her Majesty’s Inspector | - | | - | | Y | |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level. Models based upon 4,947 inspections conducted by 560 inspectors.

Contract status (OI versus HMI)

**Table 6.3.** The link between inspector gender and school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | | **Secondary** | | | |
|  | **M0** | | **M1** | | **M0** | | **M1** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Impact of HMI (Ref: Good)** | |  |  |  |  |  |  |  |
| Outstanding | 1.51\* | 5.89 | 1.14 | 1.65 | 1.11 | 0.98 | 0.91 | -0.72 |
| Requires improvement | 1.63\* | 9.04 | 1.39\* | 5.76 | 1.18\* | 2.07 | 1.14 | 1.53 |
| Inadequate | 2.62\* | 10.08 | 2.39\* | 8.25 | 1.76\* | 4.37 | 1.79\* | 4.14 |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | - | | Y | |
| Inspection type | - | | Y | | - | | - | |
| Prior Ofsted rating | - | | Y | | - | | Y | |
| School performance data | - | | Y | | - | | Y | |
| School absences | - | | Y | | - | | - | |

1. Predicted probabilities

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | | **Secondary** | | | |
|  | **M0** | | **M1** | | **M0** | | **M1** | |
|  | **OI** | **HMI** | **OI** | **HMI** | **OI** | **HMI** | **OI** | **HMI** |
| Outstanding | 7.7% | 9.0% | 8.0% | 8.1% | 10.5% | 10.3% | 11.0% | 9.9% |
| Good | 60.3% | 47.1% | 59.1% | 50.9% | 47.8% | 42.2% | 46.7% | 43.4% |
| Requires improvement | 27.8% | 35.4% | 28.6% | 32.8% | 33.9% | 35.4% | 34.5% | 34.7% |
| Inadequate | 4.2% | 8.6% | 4.3% | 8.1% | 7.8% | 12.1% | 7.9% | 12.0% |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | - | | Y | |
| Inspection type | - | | Y | | - | | - | |
| Prior Ofsted rating | - | | Y | | - | | Y | |
| School performance data | - | | Y | | - | | Y | |
| School absences | - | | Y | | - | | - | |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level. Primary/Secondary models based upon 22,743/4,970 inspections conducted by 986/565 inspectors. OI = Ofsted Inspector. HMI = Her Majesty’s Inspector.

\**p* < .05.

Outside home region

**Table 6.4.** The link between the inspection being conducted outside of the lead inspectors home region and school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Primary** | | **Secondary** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Impact of inspection outside home region (Ref: Good)** |  |  |  |  |
| Outstanding | 0.82\* | -2.20 | 1.40 | 1.93 |
| Requires improvement | 1.06 | 1.02 | 1.22 | 1.91 |
| Inadequate | 1.22 | 1.94 | 0.99 | -0.03 |
| **Inspection-level controls** |  |  |  |  |
| School % Free School Meals | Y | | Y | |
| Inspection type | Y | | - | |
| Prior Ofsted rating | Y | | Y | |
| School performance data | Y | | Y | |
| School absences | Y | | - | |
| Inspector gender | Y | | Y | |
| Inspector contract | Y | | Y | |

1. Predicted probabilities

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Primary** | | **Secondary** | |
|  | **Inside** | **Outside** | **Inside** | **Outside** |
| Outstanding | 8% | 7% | 10% | 12% |
| Good | 57% | 57% | 45% | 41% |
| Requires improvement | 29% | 30% | 35% | 38% |
| Inadequate | 5% | 6% | 10% | 9% |
| ***n*** | **15,070** | **3,175** | **3,123** | **712** |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level. Primary/Secondary models based upon 18,245/3,847 inspections conducted by 760/322 inspectors.

\**p* < .05.

Phase specialism

**Table 6.5.** The link between inspectors’ phase specialism (primary / secondary) and primary school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***n* = 21,437** | **30-69% primary** | | **70-99% primary** | |
| ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Reference outcome = Good** |  |  |  |  |
| Outstanding | 1.53\* | 3.95 | 1.17\* | 2.14 |
| Requires improvement | 0.95 | -0.67 | 0.98 | -0.25 |
| Inadequate | 1.02 | 0.16 | 0.98 | -0.14 |
| **Controls** |  |  |  |  |
| School % Free School Meals | Yes | | | |
| Inspection type | Yes | | | |
| Prior Ofsted rating | Yes | | | |
| School performance data | Yes | | | |
| School absences | Yes | | | |
| Inspector gender | Yes | | | |
| Inspector contract status | Yes | | | |

1. Predicted probabilities

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Primary only** | **30-69% primary** | **70-99% primary** |
| Outstanding | 8% | 10% | 9% |
| Good | 57% | 56% | 57% |
| Requires improvement | 30% | 28% | 29% |
| Inadequate | 5% | 5% | 5% |
| ***n*** | **15,000** | **1,815** | **4,622** |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level.

\**p* < .05.

**Table 6.6.** The link between inspectors’ phase specialism (primary / secondary) and secondary school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***n* = 4,970** | **30-69% primary** | | **70-99% primary** | |
| ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Reference outcome = Good** |  |  |  |  |
| Outstanding | 1.02 | 0.12 | 0.85 | -0.85 |
| Requires improvement | 1.11 | 1.16 | 1.20 | 1.39 |
| Inadequate | 1.09 | 0.64 | 0.89 | -0.47 |
| **Controls** |  |  |  |  |
| School % Free School Meals | Yes | | | |
| Prior Ofsted rating | Yes | | | |
| School performance data | Yes | | | |
| Inspector gender | Yes | | | |
| Inspector contract status | Yes | | | |

1. Predicted probabilities

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Secondary only** | **30-69% primary** | **70-99% primary** |
| Outstanding | 11% | 11% | 9% |
| Good | 46% | 44% | 45% |
| Requires improvement | 33% | 35% | 38% |
| Inadequate | 10% | 10% | 8% |
| ***n*** | **2,307** | **1,953** | **710** |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level.

Team size primary

**Table 6.7.** The link between inspection team size and primary school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***n* = 21,131** | **2 inspectors** | | **3 inspectors** | | **4 inspectors** | | **5 inspectors** | |
| ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Reference outcome = Good** |  |  |  |  |  |  |  |  |
| Outstanding | 0.90 | -1.30 | 1.20\* | 2.23 | 1.94\* | 4.80 | 3.17\* | 4.50 |
| Requires improvement | 1.11\* | 2.18 | 1.12\* | 2.29 | 1.01 | 0.12 | 1.25 | 1.02 |
| Inadequate | 1.81\* | 5.42 | 2.03\* | 6.08 | 1.71\* | 3.21 | 0.50 | -1.28 |
| **Controls** |  |  |  |  |  |  |  |  |
| School % Free School Meals | Yes | | | | | | | |
| Inspection type | Yes | | | | | | | |
| Prior Ofsted rating | Yes | | | | | | | |
| School performance data | Yes | | | | | | | |
| School absences | Yes | | | | | | | |
| Inspector gender | Yes | | | | | | | |
| Inspector contract status | Yes | | | | | | | |

1. Predicted probabilities

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Number of inspectors** | | | | |
|  | **1** | **2** | **3** | **4** | **5** |
| Outstanding | 8% | 7% | 9% | 13% | 17% |
| Good | 60% | 57% | 55% | 55% | 50% |
| Requires improvement | 29% | 30% | 30% | 27% | 31% |
| Inadequate | 3% | 6% | 6% | 5% | 2% |
| ***n*** | **5,546** | **7,184** | **7,158** | **1,093** | **150** |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level.

\**p* < .05.

**Table 6.8.** The link between inspection team size and secondary school inspection outcomes. Multinomial regression estimates.

1. Regression model estimates

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **1 inspector** | | **2 inspectors** | | **3 inspectors** | | **5 inspectors** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Reference outcome = Good** |  |  |  |  |  |  |  |  |
| Outstanding | 2.58\* | -4.39 | 0.91 | 0.35 | 0.66\* | 2.73 | 1.15 | -0.93 |
| Requires improvement | 0.87 | 0.93 | 0.97 | 0.22 | 0.89 | 1.41 | 0.96 | 0.39 |
| Inadequate | 0.53\* | 3.10 | 0.82 | 1.04 | 0.97 | 0.32 | 0.60\* | 3.42 |
| **Controls** |  |  |  |  |  |  |  |  |
| School % Free School Meals | Yes | | | | | | | |
| Prior Ofsted rating | Yes | | | | | | | |
| School performance data | Yes | | | | | | | |
| Inspector contract status | Yes | | | | | | | |

1. Predicted probabilities

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Number of inspectors** | | | | |
|  | **1** | **2** | **3** | **4** | **5** |
| Outstanding | 17% | 9% | 7% | 9% | 10% |
| Good | 39% | 42% | 44% | 41% | 42% |
| Requires improvement | 35% | 36% | 34% | 36% | 38% |
| Inadequate | 9% | 12% | 15% | 14% | 10% |
| ***n*** | **404** | **407** | **1,503** | **2,593** | **1,035** |

Notes: Predicted probabilities generated holding other values of the covariates to their mean. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. Standard errors have been clustered at the inspector level.

\**p* < .05.

**Appendix 7. Ordinal logistic regression model estimates of differences in Overall Effectiveness judgements between OIs and HMIs**

**Table 7.1.** Ordinal regression model estimates of the link between contract status and inspection outcomes. Secondary school results.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| HMI (ref: OI) | 1.26\* | 3.05 | 1.20\* | 2.72 | 1.13 | 1.69 | 1.18\* | 2.14 | 1.18\* | 2.09 | 1.18\* | 2.03 | 1.32\* | 3.49 |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School religion | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School gender | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Ofsted region | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Prior Ofsted rating | - | | - | | Y | | Y | | Y | | Y | | Y | |
| School performance data | - | | - | | - | | Y | | Y | | Y | | Y | |
| School absences | - | | - | | - | | - | | Y | | Y | | Y | |
| School % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | |
| School % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Inspector gender | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector phase specialism | - | | - | | - | | - | | - | | - | | Y | |
| Inspecting inside home region | - | | - | | - | | - | | - | | - | | Y | |
| Inspection experience | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds-ratios above one indicate that being inspected by an HMI is associated with a worse inspection outcome. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. M0-M3 based upon 5,024 inspections conducted by 586 inspectors. M4-M6 based upon 4,899 inspections conducted by 564 inspectors. Standard errors have been clustered at the inspector level. HMI = Her Majesty’s Inspector. OI = Ofsted inspector.

\**p* < .05.

**Appendix 8. Alternative estimates for the link between lead inspector gender and short inspection outcomes for primary schools**

1. September 2015- August 2019

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Watchsted sample** | |  | **Extended sample** | |
|  | ***n*** | ***OR*** | ***t*** | ***n*** | ***OR*** | ***t*** |
| M0 | 8,302 | 0.84 | -1.71 | 8,860 | 0.81\* | -2.15 |
| M1 | 8,302 | 0.83 | -1.87 | 8,860 | 0.80\* | -2.40 |
| M2 | 8,302 | 0.83 | -1.87 | 8,860 | 0.80\* | -2.40 |
| M3 | 8,302 | 0.83 | -1.81 | 8,860 | 0.80\* | -2.35 |
| M4 | 8,302 | 0.83 | -1.81 | 8,860 | 0.80\* | -2.34 |
| M5 | 8,302 | 0.82 | -1.94 | 8,860 | 0.79\* | -2.52 |
| M6 | 8,302 | 0.82\* | -2.00 | 8,860 | 0.77\* | -2.68 |
| M7 | 8,302 | 0.81\* | -2.08 | 8,860 | 0.77\* | -2.78 |

Notes: Model M7 adds a control for academic year, in addition to the variables controlled in model M6. Based upon short inspections conducted between September 2015 and August 2019. The outcome measure is a “negative” short inspection outcome (conversion to a full inspection leading to a downgrade in the Overall Effectiveness judgement or recommendation of an S5 inspection next due to concerns). Odds ratio below one indicates male inspectors award more lenient short inspection outcomes than their female counterparts.

\**p* < .05.

1. September 2015 – December 2017

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Watchsted sample** | |  | **Extended sample** | |
|  | ***n*** | ***OR*** | ***t*** | ***n*** | ***OR*** | ***t*** |
| M0 | 4,697 | 0.86 | -1.35 | 4,993 | 0.85 | -1.56 |
| M1 | 4,697 | 0.86 | -1.34 | 4,993 | 0.85 | -1.62 |
| M2 | 4,697 | 0.86 | -1.34 | 4,993 | 0.85 | -1.62 |
| M3 | 4,697 | 0.82 | -1.69 | 4,993 | 0.81\* | -1.98 |
| M4 | 4,697 | 0.83 | -1.55 | 4,993 | 0.82 | -1.81 |
| M5 | 4,697 | 0.83 | -1.57 | 4,993 | 0.82 | -1.84 |
| M6 | 4,697 | 0.85 | -1.31 | 4,993 | 0.82 | -1.80 |
| M7 | 4,697 | 0.85 | -1.35 | 4,993 | 0.81 | -1.87 |

1. January 2018 – August 2019

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Watchsted sample** | |  | **Extended sample** | |
|  | ***n*** | ***OR*** | ***t*** | ***n*** | ***OR*** | ***t*** |
| M0 | 3,605 | 0.79 | -1.56 | 3,867 | 0.73\* | -2.11 |
| M1 | 3,605 | 0.79 | -1.70 | 3,867 | 0.74\* | -2.32 |
| M2 | 3,605 | 0.79 | -1.70 | 3,867 | 0.74\* | -2.32 |
| M3 | 3,605 | 0.83 | -1.35 | 3,867 | 0.76\* | -2.01 |
| M4 | 3,605 | 0.82 | -1.46 | 3,867 | 0.75\* | -2.12 |
| M5 | 3,605 | 0.80 | -1.70 | 3,867 | 0.73\* | -2.40 |
| M6 | 3,605 | 0.75\* | -2.19 | 3,867 | 0.69\* | -2.86 |
| M7 | 3,605 | 0.75\* | -2.17 | 3,867 | 0.69\* | -2.84 |

\**p* < .05.

**Appendix 9. Additional estimates**

**Table 9.1**. The percentage of the variation in inspection outcomes that occurs between different inspectors.

1. Primary

|  |  |  |
| --- | --- | --- |
| **Primary** | | |
|  | **Unconditional** | **Conditional** |
| Ordinal | 9.6% | 9.4% |
| Outstanding | 9.6% | 8.3% |
| Good | 6.3% | 5.1% |
| RI | 7.2% | 6.9% |
| Inadequate | 17.5% | 16.0% |
| Short inspection | 12.2% | 10.8% |

1. Secondary

|  |  |  |
| --- | --- | --- |
| **Secondary** | | |
|  | **Unconditional** | **Conditional** |
| Ordinal | 7.2% | 5.1% |
| Outstanding | 9.2% | 12.0% |
| Good | 4.9% | 2.7% |
| RI | 4.0% | 2.2% |
| Inadequate | 11.5% | 10.2% |
| Short inspection | 5.0% | 0.4% |

Notes: Figures refer to the percent of the variation in inspection outcomes that occurs between different inspectors. Estimates based upon multi-level (random effects) ordinal or binary logistic regression models, with inspections being nested within inspectors. Unconditional estimates refer to results from an empty model with no controls. Conditional estimates include controls for percent of pupils eligible for Free School Meals, Ofsted region, previous Ofsted inspection rating, inspection type, school gender composition (secondary only) and school performance measures (average Key Stage 2 maths and English scores for primary schools and average Key Stage 4 grades and progress measures for secondary schools). Primary estimates based upon 22,761 inspections conducted by 996 inspectors (other than short inspections, which is based upon 8,329 inspections conducted by 565 inspectors). Secondary estimates based upon 5,024 inspections conducted by 586 inspectors (other than short inspections, which is based upon 1,199 inspections conducted by 253 inspectors). Analysis based upon all inspections conducted between the 2011/12 and 2018/19 academic years.

**Table 9.2.** Crosstabulation between the gender of the lead inspector and overall effectiveness judgements.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | **Female** | **Male** | **Difference** | **Female** | **Male** | **Difference** |
| Outstanding | 7.8% | 8.2% | 0.4% | 10.9% | 10.1% | -0.9% |
| Good | 55.9% | 58.7% | 2.9% | 45.4% | 44.9% | -0.5% |
| Requires improvement | 30.5% | 28.6% | -1.9% | 34.6% | 34.6% | -0.1% |
| Inadequate | 5.9% | 4.5% | -1.4% | 9.1% | 10.5% | 1.4% |
| ***n (inspections)*** | **11,056** | **11,698** |  | **2,188** | **2,813** |  |

Notes: Figures refer to column percentages.

**Table 9.3.** Ordinal regression model estimates of the link between inspector gender and inspection outcomes. Primary school results.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***T*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| Female inspector (ref: Male) | 0.87\* | -2.60 | 0.86\* | -2.87 | 0.85\* | -3.05 | 0.84\* | -3.18 | 0.83\* | -3.39 | 0.84\* | -3.22 | 0.84\* | -3.18 |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School religion | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School gender | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Ofsted region | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Prior Ofsted rating | - | | - | | Y | | Y | | Y | | Y | | Y | |
| School performance data | - | | - | | - | | Y | | Y | | Y | | Y | |
| School absences | - | | - | | - | | - | | Y | | Y | | Y | |
| School % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | |
| School % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Her Majesty’s Inspector | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector phase specialism | - | | - | | - | | - | | - | | - | | Y | |
| Inspecting inside home region | - | | - | | - | | - | | - | | - | | Y | |
| Inspection experience | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds-ratios below one indicate that being inspected by a female lead inspector is associated with a worse inspection outcome. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. M0-M3 based upon 22,754 inspections conducted by 983 inspectors. M4-M6 based upon 21,366 inspections conducted by 983 inspectors. Standard errors have been clustered at the inspector level.

\**p* < .05.

**Table 9.4.** The link between inspector gender and primary school inspection outcomes. Sub-judgements.

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***N*** | ***OR*** | ***t*** |
| Behaviour | 14,863 | 0.84\* | -2.48 |
| Development | 6,503 | 0.81\* | -3.02 |
| Leadership & Management | 21,366 | 0.83\* | -3.44 |
| Outcomes | 7,588 | 0.81\* | -3.21 |
| Quality | 6,503 | 0.84\* | -2.41 |
| Teaching | 14,863 | 0.86\* | -2.19 |
| Overall effectiveness | 21,366 | 0.84\* | -3.18 |

Notes: Number of observations differs due to sub-domains changing over time. Estimates based model specification M6, which controls for percentage of children eligible for FSM, school religion, school gender composition, Ofsted region, inspection type, prior school inspection rating, school performance data, school absences, percentage of pupils with SEN, percentage of pupils with EAL, whether the inspector is an HMI, inspectors amount of inspection experience, inspector phase specialism (primary versus secondary) and whether the inspector is inspecting in their home region.

\**p* < .05.

**Table 9.5.** The association between lead inspector gender and short inspection outcomes.

1. Primary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Time-period** | ***N*** | ***OR*** | **Confidence interval** |
| Section 5 next due to concerns or conversion with a subsequent downgrade in judgement | September 2015 – August 2019 | 8,302 | 0.81 | 0.66 – 0.99 |
| Conversion leading to Requires Improvement or Inadequate | September 2015 – December 2017 | 4,697 | 0.85 | 0.67 – 1.08 |
| Conversion or Section 5 recommended next due to concerns | January 2018 – August 2019 | 3,605 | 0.75 | 0.57 – 0.97 |

1. Secondary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Time-period** | ***N*** | ***OR*** | **Confidence interval** |
| Section 5 next due to concerns or conversion with a subsequent downgrade in judgement | September 2015 – August 2019 | 1,184 | 0.93 | 0.65-1.29 |
| Conversion leading to Requires Improvement or Inadequate | September 2015 – December 2017 | 753 | 1.19 | 0.75-1.75 |
| Conversion or Section 5 recommended next due to concerns | January 2018 – August 2019 | 431 | 0.84 | 0.45-1.39 |

Notes: Odds ratios below one indicates that male lead inspectors are less likely to convert or recommend a Section 5 inspection next than their female counterparts. Estimates based upon model M6.

**Table 9.6.** The unconditional association between inspector contract status and Ofsted inspection judgements.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | **Ofsted inspector** | **Her Majesty’s Inspector** | **Difference** | **Ofsted Inspector** | **Her Majesty’s Inspector** | **Difference** |
| Outstanding | 7.7% | 9.0% | -1.3% | 10.5% | 10.3% | -0.2% |
| Good | 60.3% | 47.0% | -13.3% | 47.8% | 42.2% | -5.6% |
| Requires improvement | 27.8% | 35.4% | 7.7% | 33.9% | 35.4% | 1.6% |
| Inadequate | 4.2% | 8.6% | 4.4% | 7.8% | 12.1% | 4.3% |
| ***n* (inspections)** | **17,622** | **5,139** |  | **2,654** | **2,370** |  |

Notes: Difference column refers to percentage for HMI minus percentage for OI. Analysis based upon data from 986 primary and 586 secondary lead inspectors. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years.

**Table 9.7.** Ordinal regression model estimates of the link between contract status and inspection outcomes. Primary school results.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| HMI (ref: OI) | 1.54\* | 7.23 | 1.53\* | 7.55 | 1.49\* | 6.94 | 1.43\* | 6.12 | 1.43\* | 6.07 | 1.42\* | 5.95 | 1.45\* | 6.21 |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School religion | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School gender | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Ofsted region | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Prior Ofsted rating | - | | - | | Y | | Y | | Y | | Y | | Y | |
| School performance data | - | | - | | - | | Y | | Y | | Y | | Y | |
| School absences | - | | - | | - | | - | | Y | | Y | | Y | |
| School % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | |
| School % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Inspector gender | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector phase specialism | - | | - | | - | | - | | - | | - | | Y | |
| Inspecting inside home region | - | | - | | - | | - | | - | | - | | Y | |
| Inspection experience | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds-ratios above one indicate that being inspected by an HMI is associated with a worse inspection outcome. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. M0-M3 based upon 22,761 inspections conducted by 986 inspectors. M4-M6 based upon 21,372 inspections conducted by 986 inspectors. Standard errors have been clustered at the inspector level. HMI = Her Majesty’s Inspector. OI = Ofsted inspector.

\**p* < .05.

**Table 9.8.** Ordinal regression model estimates of the link between contract status and inspection outcomes. Secondary school results.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| HMI (ref: OI) | 1.26\* | 3.05 | 1.20\* | 2.72 | 1.13 | 1.69 | 1.18\* | 2.14 | 1.18\* | 2.09 | 1.18\* | 2.03 | 1.32\* | 3.49 |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School religion | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School gender | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Ofsted region | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Prior Ofsted rating | - | | - | | Y | | Y | | Y | | Y | | Y | |
| School performance data | - | | - | | - | | Y | | Y | | Y | | Y | |
| School absences | - | | - | | - | | - | | Y | | Y | | Y | |
| School % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | |
| School % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Inspector gender | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector phase specialism | - | | - | | - | | - | | - | | - | | Y | |
| Inspecting inside home region | - | | - | | - | | - | | - | | - | | Y | |
| Inspection experience | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds-ratios above one indicate that being inspected by an HMI is associated with a worse inspection outcome. Data based upon inspections conducted between the 2011/12 to 2018/19 academic years. M0-M3 based upon 5,024 inspections conducted by 586 inspectors. M4-M6 based upon 4,899 inspections conducted by 564 inspectors. Standard errors have been clustered at the inspector level. HMI = Her Majesty’s Inspector. OI = Ofsted inspector.

\**p* < .05.

**Table 9.9.** The link between inspector contract status and inspection outcomes. Sub-judgements.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Primary** | | | **Secondary** | | |
|  | ***n*** | ***OR*** | ***t*** | ***n*** | ***OR*** | ***t*** |
| Behaviour | 14,868 | 1.59\* | 4.69 | 3,053 | 1.66\* | 4.40 |
| Development | 6,504 | 1.15 | 1.87 | 1,846 | 0.97 | -0.20 |
| Leadership & Management | 21,372 | 1.38\* | 5.26 | 4,899 | 1.31\* | 3.10 |
| Outcomes | 7,589 | 1.33\* | 4.00 | 2,028 | 1.10 | 0.79 |
| Quality | 6,504 | 1.23\* | 2.63 | 1,846 | 1.00 | 0.00 |
| Teaching | 14,868 | 1.56\* | 4.93 | 3,053 | 1.53\* | 4.24 |
| **Overall effectiveness** | **21,372** | **1.45**\* | **6.21** | 4,899 | **1.32**\* | **3.49** |

Notes: Number of observations differs due to sub-domains changing over time. Estimates based model specification M6.

\**p* < .05.

**Table 9.10.** The unconditional association between inspector contract status and short inspection outcomes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Primary** | | **Secondary** | |
|  | **Ofsted inspector** | **Her Majesty’s Inspector** | **Ofsted inspector** | **Her Majesty’s Inspector** |
| Existing grade retained | 90% | 86% | 82% | 78% |
| Conversion with downgrade or Section 5 next due to concerns | 10% | 14% | 18% | 22% |
| ***n*** | **3,369** | **4,960** | **469** | **730** |

**Table 9.11.** The association between lead inspector contract status and a negative outcome from a short inspection.

1. Primary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Time-period** | ***n*** | ***OR*** | **Confidence interval** |
| Section 5 next due to concerns or conversion with a subsequent downgrade in judgement | September 2015 - August 2019 | 8,329 | 1.44 | 1.18-1.76 |
| Conversion leading to Requires improvement or Inadequate | September 2015 - December 2017 | 4,702 | 1.27 | 1.00-1.62 |
| Conversion or Section 5 recommended next due to concerns | January 2018 - August 2019 | 3,627 | 1.63 | 1.25-2.14 |

1. Secondary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Time-period** | ***n*** | ***OR*** | **Confidence interval** |
| Section 5 next due to concerns or conversion with a subsequent downgrade in judgement | September 2015 - August 2019 | 1,184 | 1.40 | 0.88-2.22 |
| Conversion leading to Requires improvement or Inadequate | September 2015 - December 2017 | 753 | 1.56 | 0.89-2.71 |
| Conversion or Section 5 recommended next due to concerns | January 2018 - August 2019 | 431 | 1.41 | 0.66-3.01 |

Notes: Odds ratios above one indicate that short inspection led by an HMI more likely to lead to a negative outcome for the school than a short inspection led by an OI. Estimated based upon model M6.

**Table 9.12.** Ordinal regression model estimates of the link between inspector experience and overall effectiveness judgements.

1. Primary schools

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Quintile 2** | | **Quintile 3** | | **Quintile 4** | | **Quintile 5** | |  |
| **Model** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***n*** |
| 0 | 0.94 | -0.63 | 0.99 | -0.09 | 1.13 | 1.40 | 0.89 | -1.26 | 6,523 |
| 1 | 1.00 | -0.03 | 1.06 | 0.69 | 1.15 | 1.55 | 0.93 | -0.70 | 6,523 |
| 2 | 1.10 | 1.00 | 1.14 | 1.46 | 1.21\* | 2.12 | 0.99 | -0.09 | 6,523 |
| 3 | 1.00 | 0.00 | 1.04 | 0.39 | 1.12 | 1.26 | 0.94 | -0.72 | 6,505 |
| 4 | 0.97 | -0.28 | 1.02 | 0.24 | 1.09 | 0.95 | 0.87 | -1.44 | 6,505 |
| 5 | 0.95 | -0.50 | 1.02 | 0.24 | 1.08 | 0.84 | 0.87 | -1.43 | 6,505 |
| 6 | 0.96 | -0.47 | 1.03 | 0.31 | 1.09 | 0.95 | 0.89 | -1.22 | 6,505 |
| 7 | 0.95 | -0.56 | 1.01 | 0.08 | 1.07 | 0.70 | 0.92 | -0.89 | 6,505 |
| 8 | 0.95 | -0.48 | 1.02 | 0.22 | 1.09 | 0.91 | 0.95 | -0.51 | 6,505 |
| 9 | 0.94 | -0.60 | 1.00 | 0.03 | 1.07 | 0.69 | 0.93 | -0.64 | 6,505 |

1. Secondary schools

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Quintile 2** | | **Quintile 3** | | **Quintile 4** | | **Quintile 5** | |  |
| **Model** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***n*** |
| 0 | 0.90 | -0.75 | 0.95 | -0.41 | 0.94 | -0.45 | 1.11 | 0.68 | 1,884 |
| 1 | 0.92 | -0.56 | 0.97 | -0.26 | 0.95 | -0.36 | 1.14 | 0.77 | 1,884 |
| 2 | 0.84 | -1.22 | 0.96 | -0.34 | 0.94 | -0.43 | 1.04 | 0.27 | 1,884 |
| 3 | 0.80 | -1.53 | 0.91 | -0.75 | 0.93 | -0.48 | 1.03 | 0.18 | 1,872 |
| 4 | 0.80 | -1.49 | 0.75\* | -2.02 | 0.85 | -1.02 | 0.89 | -0.59 | 1,847 |
| 5 | 0.76 | -1.83 | 0.73\* | -2.15 | 0.83 | -1.17 | 0.90 | -0.51 | 1,847 |
| 6 | 0.76 | -1.83 | 0.74\* | -2.06 | 0.83 | -1.15 | 0.90 | -0.51 | 1,847 |
| 7 | 0.76 | -1.84 | 0.74\* | -2.07 | 0.83 | -1.17 | 0.90 | -0.50 | 1,847 |
| 8 | 0.76 | -1.77 | 0.75 | -1.95 | 0.85 | -0.89 | 0.94 | -0.32 | 1,847 |
| 9 | 0.72 | -1.95 | 0.71\* | -2.13 | 0.82 | -1.08 | 0.89 | -0.53 | 1,847 |

Notes: Sample of inspections from September 2015 to August 2019. Experience measured as total number of inspections conducted since September 2011. Low experience (Quintile 1) is the reference group. OR refers to the estimated odds ratio. M0 has no controls. M1 adds a control for academic year. M2 controls for school religion, gender, FSM and Ofsted region. M3 adds controls for prior inspection outcome and inspection type. M4 adds school performance data. M5 controls for school absence, EAL and SEN. M6 adds inspector gender. M7 adds inspector contract status (HMI / OI). M8 adds inspector phase specialism (primary/secondary). M9 controls for whether inspection was conducted outside the inspector’s home region.

\**p* < .05.

**Table 9.13.** Cross-tabulation for whether an inspection took place outside of the lead inspector’s home region and inspection outcomes.

1. Primary

|  |  |  |
| --- | --- | --- |
|  | **Inspection outside of home region** | |
|  | **No %** | **Yes %** |
| **Overall effectiveness** |  |  |
| Outstanding | 8 | 7 |
| Good | 57 | 59 |
| Requires improvement | 30 | 28 |
| Inadequate | 5 | 5 |
| ***n*** | **15,925** | **3,347** |
| **Section 5 next due to concerns or conversion with subsequent downgrade** |  |  |
| No | 87 | 86 |
| Yes | 13 | 14 |
| ***n*** | **7,627** | **183** |

1. Secondary

|  |  |  |
| --- | --- | --- |
|  | **Inspection outside of home region** | |
|  | **No %** | **Yes %** |
| **Overall effectiveness** |  |  |
| Outstanding | 9 | 15 |
| Good | 45 | 44 |
| Requires improvement | 36 | 34 |
| Inadequate | 11 | 7 |
| ***n*** | **3,161** | **735** |
| **Section 5 next due to concerns or conversion with subsequent downgrade** |  |  |
| No | 80 | 72 |
| Yes | 20 | 28 |
| ***n*** | **988** | **43** |

Notes: Figures refer to column percentages. Lower panel captures whether the short inspection was converted to a full inspection with a subsequent downgrade or a recommendation was made for a Section 5 inspection to be conducted next.

**Table 9.14.** Ordinal logistic regression model estimates of the link between whether the inspection was within the lead inspector’s home region and overall effectiveness judgements. Results for primary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | | **M7** | | **M8** | |
|  | ***OR*** | ***T*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| Outside region | 0.99 | -0.20 | 1.02 | 0.42 | 1.06 | 1.24 | 1.08 | 1.61 | 1.09 | 1.78 | 1.11\* | 2.03 | 1.13\* | 2.50 | 1.13\* | 2.53 | 1.13\* | 2.50 |
| ***n*** | **19,272** | | **19,272** | | **19,254** | | **19,254** | | **18,122** | | **18,122** | | **18,122** | | **18,122** | | **18,122** | |
| Controls |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School FSM | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School background | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Region | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Previous rating | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School performance | - | | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Male inspector | - | | - | | - | | - | | - | | Y | | Y | | Y | | Y | |
| Her Majesty’s Inspector | - | | - | | - | | - | | - | | - | | Y | | Y | | Y | |
| Phase specialism | - | | - | | - | | - | | - | | - | | - | | Y | | Y | |
| Post 2018 | - | | - | | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate that inspections conducted outside of the lead inspector’s home region have worse inspection outcomes.

\**p* < .05.

**Table 9.15.** Ordinal logistic regression model estimates of the link between whether the inspection was within the lead inspector’s home region and overall effectiveness judgements. Results for secondary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | | **M7** | | **M8** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| Outside region | 0.73\* | -3.44 | 0.92 | -0.93 | 0.96 | -0.46 | 0.97 | -0.37 | 0.96 | -0.39 | 0.96 | -0.38 | 1.00 | -0.04 | 1.00 | 0.00 | 0.99 | -0.14 |
| ***N*** | **3,896** | | **3,896** | | **3,884** | | **3,835** | | **3,793** | | **3,793** | | **3,793** | | **3,793** | | **3,793** | |
| Controls |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School FSM | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School background | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Region | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Previous rating | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School performance | - | | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Male inspector | - | | - | | - | | - | | - | | Y | | Y | | Y | | Y | |
| Her Majesty’s Inspector | - | | - | | - | | - | | - | | - | | Y | | Y | | Y | |
| Phase specialism | - | | - | | - | | - | | - | | - | | - | | Y | | Y | |
| Post 2018 | - | | - | | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate that inspections conducted outside of the lead inspector’s home region have worse inspection outcomes.

\**p* < .05.

**Table 9.16.** Cross-tabulation between the percent of primary school inspections an inspector conducts throughout their inspection career and Ofsted judgements.

1. Primary

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Phase specialism** | | |
|  | **30-69% primary** | **70-99% primary** | **Primary only** |
| **Overall effectiveness** |  |  |  |
| Outstanding | 11 | 9 | 7 |
| Good | 53 | 56 | 58 |
| Requires improvement | 30 | 29 | 30 |
| Inadequate | 6 | 5 | 5 |
| ***n*** | **1,912** | **4,880** | **15,871** |
| **Section 5 next due to concerns or conversion with downgrade** |  |  |  |
| No | 87 | 85 | 88 |
| Yes | 13 | 15 | 12 |
| ***n*** | **1,062** | **1,833** | **5,378** |

1. Secondary

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Phase specialism** | | |
|  | **30-69% primary** | **70-99% primary** | **Secondary only** |
| **Overall effectiveness** |  |  |  |
| Outstanding | 10 | 9 | 11 |
| Good | 44 | 46 | 46 |
| Requires improvement | 35 | 37 | 33 |
| Inadequate | 11 | 8 | 10 |
| ***n*** | **1,976** | **740** | **2,308** |
| **Section 5 next due to concerns or conversion with downgrade** |  |  |  |
| No | 77 | 75 | 82 |
| Yes | 23 | 25 | 18 |
| ***n*** | **425** | **107** | **667** |

Notes: Figures refer to column percentages. Lower panel captures whether the short inspection was converted to a full inspection with a subsequent downgrade or a recommendation was made for a Section 5 inspection to be conducted next.

**Table 9.17.** Ordinal logistic regression model estimates of the link between inspector phase specialism and overall effectiveness judgements. Results for primary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | | **M7** | | **M8** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Primary only (Reference)** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 30-69% primary | 0.97 | -0.31 | 0.96 | -0.50 | 0.91 | -1.17 | 0.89 | -1.49 | 0.87 | -1.73 | 0.89 | -1.44 | 0.85\* | -2.04 | 0.86 | -1.92 | 0.86 | -1.91 |
| 70-99% primary | 0.96 | -0.54 | 0.97 | -0.41 | 0.96 | -0.65 | 0.95 | -0.76 | 0.94 | -0.88 | 0.95 | -0.85 | 0.95 | -0.86 | 0.95 | -0.83 | 0.95 | -0.83 |
| ***n*** | **22,663** | | **22,663** | | **22,645** | | **22,645** | | **21,277** | | **21,277** | | **21,277** | | **21,277** | | **21,277** | |
| Controls |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School background | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Region | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Previous rating | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School performance | - | | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Male inspector | - | | - | | - | | - | | - | | Y | | Y | | Y | | Y | |
| Her Majesty’s Inspector | - | | - | | - | | - | | - | | - | | Y | | Y | | Y | |
| Phase specialism | - | | - | | - | | - | | - | | - | | - | | Y | | Y | |
| Post 2018 | - | | - | | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate that those inspectors who have ever conducted secondary school inspections have worse inspection outcomes.

\**p* < .05.

**Table 9.18.** Ordinal logistic regression model estimates of the link between inspector phase specialism and overall effectiveness judgements. Results for secondary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | | **M7** | | **M8** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Secondary only (Reference)** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 30-69% primary | 1.12 | 1.33 | 0.96 | -0.55 | 0.92 | -1.09 | 0.92 | -1.03 | 0.92 | -1.02 | 0.92 | -1.06 | 0.90 | -1.36 | 0.92 | -1.06 | 0.92 | -1.06 |
| 70-99% primary | 1.08 | 0.63 | 1.01 | 0.14 | 1.00 | 0.00 | 0.98 | -0.19 | 0.99 | -0.05 | 1.00 | -0.02 | 1.01 | 0.11 | 1.03 | 0.24 | 1.03 | 0.21 |
| ***N*** | **5,024** | | **5,024** | | **5,012** | | **4,958** | | **4,899** | | **4,899** | | **4,899** | | **4,899** | | **4,899** | |
| Controls |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School background | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Region | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Previous rating | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School performance | - | | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Male inspector | - | | - | | - | | - | | - | | Y | | Y | | Y | | Y | |
| Her Majesty’s Inspector | - | | - | | - | | - | | - | | - | | Y | | Y | | Y | |
| Phase specialism | - | | - | | - | | - | | - | | - | | - | | Y | | Y | |
| Post 2018 | - | | - | | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate that those inspectors who have ever conducted secondary school inspections have worse inspection outcomes.

**Table 9.19.** Ordinal logistic regression model estimates of the link between inspector phase specialism and a negative outcome from the short inspection. Results for primary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | | **M7** | | **M8** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Primary only (Reference)** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 30-69% primary | 1.09 | 0.71 | 1.07 | 0.60 | 1.07 | 0.60 | 1.08 | 0.55 | 1.08 | 0.57 | 1.14 | 0.96 | 1.03 | 0.23 | 1.02 | 0.12 | 1.00 | 0.01 |
| 70-99% primary | 1.38\* | 2.20 | 1.30 | 1.88 | 1.30 | 1.88 | 1.30 | 1.85 | 1.30 | 1.84 | 1.32 | 1.95 | 1.29 | 1.88 | 1.27 | 1.71 | 1.24 | 1.58 |
| ***N*** | **8,273** | | **8,273** | | **8,273** | | **8,273** | | **8,273** | | **8,273** | | **8,273** | | **8,273** | | **8,273** | |
| Controls |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School background | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Region | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Previous rating | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School performance | - | | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Male inspector | - | | - | | - | | - | | - | | Y | | Y | | Y | | Y | |
| Her Majesty’s Inspector | - | | - | | - | | - | | - | | - | | Y | | Y | | Y | |
| Outside region | - | | - | | - | | - | | - | | - | | - | | Y | | Y | |
| Post 2018 | - | | - | | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate that those inspectors who have ever conducted secondary school inspections are more likely to convert to full inspection leading to a downgrade or recommend a Section 5 inspection next due to concerns.

\**p* < .05.

**Table 9.20.** Ordinal logistic regression model estimates of the link between inspector phase specialism and a negative outcome from the short inspection. Results for secondary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | | **M7** | | **M8** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Secondary only (Reference)** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 30-69% primary | 1.32 | 1.66 | 1.23 | 1.23 | 1.23 | 1.23 | 1.45 | 2.00 | 1.49 | 2.16 | 1.51 | 2.18 | 1.37 | 1.55 | 1.55 | 2.10 | 1.54\* | 2.07 |
| 70-99% primary | 1.51 | 1.78 | 1.86\* | 2.68 | 1.86\* | 2.68 | 1.83\* | 2.62 | 1.75\* | 2.31 | 1.75\* | 2.32 | 1.69\* | 2.15 | 1.96\* | 2.62 | 2.00\* | 2.70 |
| ***N*** | **1,199** | | **1,199** | | **1,199** | | **1,184** | | **1,184** | | **1,184** | | **1,184** | | **1,184** | | **1,184** | |
| Controls |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School background | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Region | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Previous rating | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | | Y | |
| School performance | - | | - | | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Male inspector | - | | - | | - | | - | | - | | Y | | Y | | Y | | Y | |
| Her Majesty’s Inspector | - | | - | | - | | - | | - | | - | | Y | | Y | | Y | |
| Outside region | - | | - | | - | | - | | - | | - | | - | | Y | | Y | |
| Post 2018 | - | | - | | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate that those inspectors who have ever conducted secondary school inspections are more likely to convert to full inspection leading to a downgrade or recommend a Section 5 inspection next due to concerns.

\**p* < .05.

**Table 9.21.** Predicted probability of a negative outcome from the short inspection by inspector phase specialism.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Phase specialism** | | | |
|  | **30-69% primary** | **70-99% primary** | **Primary only** | **Secondary only** |
| **Ofsted Phase** |  |  |  |  |
| Primary | 12% | 14% | 12% | - |
| Secondary | 22% | 26% | - | 19% |

Notes: Model controls for percentage of pupils eligible for free school meals, Ofsted region, prior inspection rating, inspection type, school performance measures, school absences, percentage of pupils who have English as an Additional Language, gender of the inspector, contract status of the inspector (OI or HMI) and whether the short inspection was conducted before or after January 2018.

**Table 9.22.** Cross-tabulation between inspection team size and inspection outcomes.

1. Primary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Team size** | | | | |
|  | **1** | **2** | **3** | **4** | **5** |
| **Overall effectiveness** |  |  |  |  |  |
| Outstanding | 9% | 7% | 8% | 14% | 26% |
| Good | 61% | 59% | 55% | 46% | 30% |
| Requires improvement | 27% | 29% | 31% | 33% | 41% |
| Inadequate | 3% | 5% | 7% | 8% | 3% |
| ***n*** | **5,546** | **7,184** | **7,158** | **1,093** | **150** |
| **Short conversion with downgrade (September 2015 - December 17)** |  |  |  |  |  |
| No | 96% | 73% | 48% | 54% | 54% |
| Yes | 4% | 27% | 52% | 46% | 46% |
| ***n*** | **3,458** | **241** | **444** | **341** | **103** |
| **Section 5 next due to concerns or conversion with downgrade (January 18 - August 19)** |  |  |  |  |  |
| No | 89% | 90% | - | - | - |
| Yes | 11% | 10% | - | - | - |
| ***n*** | **3,279** | **206** | **-** | **-** | **-** |

1. Secondary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Team size** | | | | |
|  | **1** | **2** | **3** | **4** | **5** |
| **Overall effectiveness** |  |  |  |  |  |
| Outstanding | 22% | 9% | 7% | 10% | 15% |
| Good | 43% | 47% | 46% | 44% | 46% |
| Requires improvement | 30% | 36% | 35% | 35% | 32% |
| Inadequate | 4% | 9% | 12% | 11% | 7% |
| ***n*** | **233** | **273** | **1,148** | **2,072** | **889** |
| **Short conversion with downgrade (September 2015 - December 17)** |  |  |  |  |  |
| No | 100% | 99% | 40% | 52% | 52% |
| Yes | 0% | 1% | 60% | 48% | 48% |
| ***n*** | **74** | **353** | **40** | **95** | **168** |
| **Section 5 next due to concerns or conversion with downgrade (January 18 - August 19)** |  |  |  |  |  |
| No | 85% | 79% | 69% | **-** | **-** |
| Yes | 15% | 21% | 31% | **-** | **-** |
| ***n*** | **71** | **307** | **45** | **-** | **-** |

Notes: Figures refer to column percentages. Lower panel captures whether there was a negative outcome from the short inspection (conversion with a downgrade in Overall Effectiveness rating or recommendation of S5 next due to concerns).

**Table 9.23.** Ordinal logistic regression model estimates of the link between inspection team size and Overall Effectiveness judgements. Results for primary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Team size (Ref: 1 inspector)** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 inspectors | 1.24\* | 5.93 | 1.15\* | 3.64 | 1.20\* | 4.88 | 1.18\* | 4.25 | 1.21\* | 4.64 | 1.21\* | 4.67 | 1.25\* | 5.29 |
| 3 inspectors | 1.37\* | 8.13 | 1.25\* | 5.67 | 1.32\* | 7.02 | 1.18\* | 3.91 | 1.22\* | 4.63 | 1.23\* | 4.73 | 1.26\* | 5.23 |
| 4 inspectors | 1.33\* | 3.43 | 1.33\* | 3.47 | 1.21\* | 2.34 | 0.98 | -0.19 | 1.03 | 0.35 | 1.03 | 0.33 | 1.05 | 0.56 |
| 5 inspectors | 0.99 | -0.06 | 1.07 | 0.27 | 0.75 | -1.19 | 0.61\* | -2.04 | 0.66 | -1.74 | 0.64 | -1.87 | 0.68 | -1.60 |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School religion | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School gender | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Ofsted region | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Prior Ofsted rating | - | | - | | Y | | Y | | Y | | Y | | Y | |
| School performance data | - | | - | | - | | Y | | Y | | Y | | Y | |
| School absences | - | | - | | - | | - | | Y | | Y | | Y | |
| School % English as Additional Language | - | | - | | - | | - | | Y | | Y | | Y | |
| School % Special Educational Needs | - | | - | | - | | - | | Y | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Inspector gender | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector Her Majesty’s Inspector | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector phase specialism | - | | - | | - | | - | | - | | - | | Y | |
| Inspecting inside home region | - | | - | | - | | - | | - | | - | | Y | |
| Academic year | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate a lower Overall Effectiveness judgement is reached than for the reference group (one inspector).

\**p* < .05.

**Table 9.24.** Ordinal logistic regression model estimates of the link between inspection team size and Overall Effectiveness judgements. Results for secondary schools.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M0** | | **M1** | | **M2** | | **M3** | | **M4** | | **M5** | | **M6** | |
|  | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** | ***OR*** | ***t*** |
| **Team size (Ref: 4 inspectors)** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 inspector | 0.50\* | -4.60 | 0.59\* | -3.55 | 0.52\* | -4.43 | 0.46\* | -4.83 | 0.44\* | -5.07 | 0.41\* | -5.51 | 0.43\* | -4.92 |
| 2 inspectors | 0.97 | -0.25 | 0.91 | -0.80 | 0.89 | -1.01 | 0.91 | -0.78 | 0.87 | -1.20 | 0.86 | -1.32 | 0.90 | -0.93 |
| 3 inspectors | 1.14\* | 1.98 | 1.10 | 1.44 | 1.05 | 0.75 | 1.08 | 0.97 | 1.08 | 0.99 | 1.08 | 0.90 | 1.08 | 0.93 |
| 5 inspectors | 0.71\* | -4.19 | 0.82\* | -2.44 | 0.79\* | -2.72 | 0.84\* | -2.05 | 0.84\* | -2.01 | 0.83\* | -2.04 | 0.83\* | -2.10 |
| **Inspection-level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| School % Free School Meals | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School religion | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| School gender | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Ofsted region | - | | Y | | Y | | Y | | Y | | Y | | Y | |
| Inspection type | - | | - | | Y | | Y | | Y | | Y | | Y | |
| Prior Ofsted rating | - | | - | | Y | | Y | | Y | | Y | | Y | |
| School performance data | - | | - | | - | | Y | | Y | | Y | | Y | |
| School absences | - | | - | | - | | - | | Y | | Y | | Y | |
| School % Free School Meals | - | | - | | - | | - | | Y | | Y | | Y | |
| School % SEN | - | | - | | - | | - | | Y | | Y | | Y | |
| **Inspector level controls** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Inspector gender | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector Her Majesty’s Inspector | - | | - | | - | | - | | - | | Y | | Y | |
| Inspector phase specialism | - | | - | | - | | - | | - | | - | | Y | |
| Inspecting inside home region | - | | - | | - | | - | | - | | - | | Y | |
| Academic year | - | | - | | - | | - | | - | | - | | Y | |

Notes: Odds ratios above one indicate a lower Overall Effectiveness judgement is reached than for the reference group (four inspectors).

\**p* < .05.

**Table 9.25.** Predicted distribution of primary school inspection outcomes for two hypothetical inspectors.

1. Multi-nominal logistic regression estimates

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Inspector A** | **Inspector B** | **Risk ratio (A/B)** |
| **Overall effectiveness** |  |  |  |
| Outstanding | 6.0% | 7.8% | 0.77 |
| Good | 45.4% | 60.1% | 0.76 |
| Requires improvement | 35.3% | 28.7% | 1.23 |
| Inadequate | 13.3% | 3.4% | 3.94 |
| **Short inspection** |  |  |  |
| Conversion with downgrade or Section 5 next due to concerns. (January 2018 – August 2019) | 15.5% | 9.7% | 1.60 |
| **Inspector characteristics** |  |  |  |
| Team size | 2 inspectors | 1 inspector |  |
| Contract status | HMI | OI |  |
| Gender | Female | Male |  |

1. Ordinal logistic regression estimates

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Inspector A** | **Inspector B** | **Risk ratio (A/B)** |
| **Overall effectiveness** |  |  |  |
| Outstanding | 4.5% | 9.0% | 0.50 |
| Good | 48.0% | 59.3% | 0.81 |
| Requires improvement | 38.4% | 27.2% | 1.41 |
| Inadequate | 9.1% | 4.5% | 2.03 |
| **Short inspection** |  |  |  |
| Conversion with downgrade or Section 5 next due to concerns. (January 2018 – August 2019) | 15.5% | 9.7% | 1.60 |
| **Inspector characteristics** |  |  |  |
| Team size | 2 inspectors | 1 inspector |  |
| Contract status | HMI | OI |  |
| Gender | Female | Male |  |

Notes: Multinomial logistic estimates control 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. Ordinal logistic regression models additionally control for school religion, school gender composition, Key Stage 1 scores and percent of pupils with special educational needs. Team size not included in the short inspection models. HMI = Her Majesty’s Inspector. OI = Ofsted Inspector.

**Appendix 10. Key changes to Ofsted inspections between September 2011 and August 2019**

Our analysis focuses on inspections conducted between September 2011 and August 2019. There were several changes to Ofsted inspections during (and since) this period which may, at times, have implications for the interpretation of our results. The key points, in chronological order, are as follows:

* 2012. Changes were made from the previous school inspection framework in January and September 2012. This included exempting schools previously judged to be outstanding from routine inspections. See Richards (2012) for further discussion.
* September 2012. Prior to this date, the third rung on Ofsted’s four-point judgement scale was labelled “satisfactory”. In September 2012 this changed to “requires improvement”. This to some extent changed expectations, with it made clear to schools that they should not remain at this level.
* January 2013. Ofsted moved to its current regional structure. England was divided into eight regions, each led by a regional director, who was responsible for managing and delivering inspections. This replaced more centralised direction of these activities.
* September 2015. Several changes were made to Ofsted inspections. This included the introduction of a common inspection framework across its different remits, the introduction of short inspections, and OIs now directly contracting with Ofsted (rather than through commercial providers). It was also when there was a substantial decline in the number of OIs inspecting on behalf of Ofsted.
* January 2018. Ofsted noted how “*the process for converting short inspections to full section 5 inspections has proven challenging for both schools and inspectors*” (Ofsted 2017a). This was because conversion required short notice changes to the inspection schedules of OIs, many of whom also hold other jobs. Indeed, Ofsted noted how “*OIs are typically busy school leaders who have booked time off to inspect, and these last minute changes are frustrating and impractical*” (Ofsted 2017b). Consequently, from January 2018, rather than issuing an immediate conversion, inspectors also had the option of recommending a full inspection to be conducted next (within the next year or two). This coincided with a notable reduction in the number of short inspections being immediately converted.
* September 2019. The new Education Inspection Framework (EIF) was introduced. This led to several significant changes to how schools were inspected, including a greater emphasis on the curriculum and less weight given to school test results.

Short inspections

Short inspections were introduced in September 2015 and reserved for schools judged to be good or outstanding during their last inspection. Between September 2015 and August 2019 these inspections typically lasted for a single day, although this increased from September 2019 to two days for schools with more than 150 pupils. Short inspections tend to be lighter touch, and do not require the lead inspector to award an Overall Effectiveness grade. Rather, they can either:

1. Confirm that the school remains at its previous grade (e.g. a school previously judged as good remains good).
2. Should convert to a full inspection “immediately” (within 48 hours up until January 2018, changed to within a maximum of seven days after January 2018).
3. Recommend that a full S5 inspection should be conducted within the next one to two years. This option was introduced in January 2018, with an inspector using it if they felt that the school was either at risk of declining to “requires improvement” or improving to “outstanding”.

Schools judged to require improvement or to be inadequate may also be subject to other forms of inspection, such as re-inspections or monitoring visits. Ofsted may also conduct no-notice inspections where they have concerns (e.g. because of their risk assessment or receiving complaints). The distribution of all types of inspections conducted between September 2011 and August 2019 by academic year can be found in Appendix 1.

Ofsted. 2017a. Ofsted confirms new arrangements for short inspections. Accessed 17/02/2022 from <https://www.gov.uk/government/news/ofsted-confirms-new-arrangements-for-short-inspections>

# Ofsted. 2017b. Ofsted seeking views on improved approach to short inspections. Accessed 18/02/2022 from <https://www.gov.uk/government/news/ofsted-seeking-views-on-improved-approach-to-short-inspections>

Richards, C. 2012. Ofsted Inspection Inspected: an examination of the 2012 framework for school inspection and its accompanying evaluation schedule. *FORUM* 54(2): 247-272.

1. Up until September 2015 OIs were employed by private sector organisations such as Serco. They have however since been directly contracted by Ofsted. This led to a sharp decline in number of OIs - from around 3,000 to 1,600 ([Richardson](https://www.bbc.co.uk/news/education-33198707), 2015). [↑](#footnote-ref-1)
2. In 2020, 88% of primary schools were rated as good or outstanding, compared to 76% of secondary schools (Ofsted 2020). [↑](#footnote-ref-2)
3. Available from <https://perspective.angelsolutions.co.uk/Perspective/Login.aspx?ReturnUrl=%2fPerspective%2fLiteUsers%2fOfsted%2f> [↑](#footnote-ref-3)
4. One would ideally have access to additional information about inspectors to ensure that the merged cases refer to the same individual. Unfortunately, very little such time-invariant information about inspectors is available for us to use. As our analysis focuses on inspector characteristics – rather than individual inspectors – the impact of any incorrect merges (e.g., Ash Rahman and Ashfaq Rahman being different people) is likely to be small. In particular, our point estimates will be largely unchanged, while reported standard errors are likely to be slightly conservative (due to underestimation of the “cluster” – i.e., inspector – sample size). [↑](#footnote-ref-4)
5. Available from <https://www.gov.uk/government/statistical-data-sets/monthly-management-information-ofsteds-school-inspections-outcomes> [↑](#footnote-ref-5)
6. The ‘Satisfactory’ grade was replaced with the ‘Requires Improvement’ grade in 2012, on the basis that the original label was thought to be lacking in ambition (Ofsted, 2012). [↑](#footnote-ref-6)
7. Outcomes from short inspections were different between September 2015 and December 2017, when they were either immediately converted into a full inspection or the Good grade was retained. [↑](#footnote-ref-7)
8. Inspectors who have conducted more than half of their inspections outside of their ‘home’ region have been recoded into a separate category of ‘no home region’. [↑](#footnote-ref-8)
9. We find about 9.5% of the variation in Overall Effectiveness judgements occurs between inspectors for primary schools (regardless of whether controls are included in the model or not), compared to between 5% and 7% for secondary schools (depending on whether controls are included). The estimated Intra-Class Correlation is slightly higher for short primary inspections (around 11%-12%), but lower for secondary inspections (between 0 and 5%). [↑](#footnote-ref-9)
10. We have also re-estimated our analytic models using multinomial (rather than ordinal) logistic regression – see Appendix 6. These confirm that there is little evidence of a gender difference when it comes to the Good/Outstanding distinction, but more so for Good/RI/Inadequate judgements. This is consistent with the descriptive results presented in Table 2, panel (a). [↑](#footnote-ref-10)
11. This is based upon Excel files published by Ofsted at <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/920755/Management_information_-_state-funded_schools_1_September_2015_to_31_August_2019.xlsx> and <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/485634/Management_information_-_schools_-_1_Sept_2005_to_31_August_2015.xlsx> [↑](#footnote-ref-11)
12. If we focus only upon primary inspections that led to an overall effectiveness rating, we have been able to access 22,760 from a total of 26,360 via the Watchsted website. Of the remaining 3,600, we have managed to access information from a further 3,324 via our own scraping of the Ofsted reports. We have hence been able to access the relevant information for 99% of all primary inspections conducted between September 2011 and August 2019 that led to an overall effectiveness judgement. [↑](#footnote-ref-12)
13. This can be computed via the formula sqrt((p\*(1-p))/n). [↑](#footnote-ref-13)
14. 6 of the 108 were Academy first section 5. [↑](#footnote-ref-14)
15. When performing a similar analysis for “non-short” inspections, we get 99% agreement (confidence interval 98% to 100%) between our automated extraction of HMI involvement in the inspections and our manual coding such information from the inspection reports. This is based upon 128 of the random sample of 150 “non-short” inspections where data is available from across the two sources. [↑](#footnote-ref-15)