

# Watching the Watchmen



The future of school inspections  
in England

Harriet Waldegrave and Jonathan Simons



---

# Watching the Watchmen

The future of school inspections in England

Harriet Waldegrave  
and Jonathan Simons



---

**Policy Exchange** is the UK's leading think tank. We are an educational charity whose mission is to develop and promote new policy ideas that will deliver better public services, a stronger society and a more dynamic economy. Registered charity no: 1096300.

Policy Exchange is committed to an evidence-based approach to policy development. We work in partnership with academics and other experts and commission major studies involving thorough empirical research of alternative policy outcomes. We believe that the policy experience of other countries offers important lessons for government in the UK. We also believe that government has much to learn from business and the voluntary sector.

#### **Trustees**

Daniel Finkelstein (Chairman of the Board), Richard Ehrman (Deputy Chair), Theodore Agnew, Richard Briance, Simon Brocklebank-Fowler, Robin Edwards, Virginia Fraser, David Frum, Edward Heathcoat Amory, David Meller, Krishna Rao, George Robinson, Robert Rosenkranz, Charles Stewart-Smith and Simon Wolfson.

---

## About the Authors

**Harriet Waldegrave** has been a Research Fellow in Policy Exchange since 2011. She is the author of “Quality Childcare: Improving early years childcare” and “Centres of Excellence? The role of Children’s Centres in Early Intervention”. Prior to joining Policy Exchange, Harriet worked at a Citizen’s Advice Bureau and the educational charity First Story. She also has experience volunteering at a children’s centre.

**Jonathan Simons** is Head of Education at Policy Exchange where he directs all research within the Unit. He is the editor of “Centres of Excellence? The role of Children’s Centres in Early Intervention” and “Reversing the ‘Widget Effect’: The introduction of performance-related pay for all teachers in English schools”. Prior to joining Policy Exchange, Jonathan worked at Serco Group, and has also been Head of Education in the Prime Minister’s Strategy Unit. He is chair of Governors of the Greenwich Free School and Vice Chair of Governors and Vice Chair of Governors, and Trustee of WWV Volunteering.

© Policy Exchange 2014

Published by

Policy Exchange, Clutha House, 10 Storey’s Gate, London SW1P 3AY

[www.policyexchange.org.uk](http://www.policyexchange.org.uk)

ISBN: 978-1-907689-73-4

Printed by Heron, Dawson and Sawyer

Designed by Soapbox, [www.soapbox.co.uk](http://www.soapbox.co.uk)

---

# Contents

	About the Authors	2
	Acknowledgments	4
	Executive Summary	5
1	Why We Need Ofsted and What it Does	12
2	The Practice of Lesson Observations	17
3	Other Methods Used by Ofsted to Inspect Schools	26
4	The Quality of the Inspectorate	34
5	Effect of Ofsted	43
6	Recommendations	50
	Appendix 1: Summary of Call for Evidence Responses	63
	Appendix 2: A Dummy 2 Page Report Card for a Short Inspection	66

---

# Acknowledgments

We are extremely grateful to Mr Krishna Rao and The J Isaacs Charitable Trust for their generous support of this project, as well as all in Policy Exchange who commented during the editing process.

We would like to express thanks for all those who contributed to the research phase of this report, including the over 300 responses to the Call for Evidence, and the teachers and headteachers who gave up time to attend our roundtables. Particular thanks are due to Sir David Carter, Professor Rob Coe, and David Didau for their thoughts and explanations of their and other research!

Finally, we would like to thank all the teachers and other educationalists who have continued to debate the role of Ofsted on blogs and on Twitter and in doing so, influenced our work - even if they didn't know they were! Social media is a democratic phenomenon which offers a tremendous opportunity for closing the gap between practitioners and policymakers. If ideas are good and arguments are compelling, then it has never been as easy as now to shape what politicians and policymakers are thinking. The role of Ofsted is a particularly contentious issue in education, and we are grateful to all for their insights in helping us shape our own messages for reform.

---

## Executive Summary

Government spending on schools in England will be almost £47bn in this academic year. Ensuring that this public money is spent effectively, that children and young people receive the best possible education and that underperforming schools are turned round are the key reasons why a regulator is required. That has been the case ever since 1839 and the appointment of the first two of Her Majesty's Inspectors, and remains the case today. In fact, in a system where schools are increasingly autonomous and diverse in their legal and operational structure, the case for an independent regulator to safeguard the public both as taxpayers and as parents is stronger than ever.

It would also be wrong not to recognise that Ofsted has in many ways been highly successful in its core mission. Around half of all schools graded Satisfactory subsequently improved to a Good rating in their next inspection. New analysis for this report suggests that a Satisfactory Ofsted rating for secondary schools also most commonly leads to a increase in children passing GCSEs in subsequent years over and above general national increases. Ofsted has also performed well in identifying the worst performing schools, and through placing them into Special Measures, has ensured they have either been closed or taken over and reopened as Academies. As one of the respondents to our consultation noted, when discussing a poor performing school that they had had dealings with in the past *"It felt like Ofsted coming in and failing this school was the only thing that would give it the push needed to get it on the path to improvement"*.

But Ofsted, as with all public bodies, needs to be consistently challenged and itself held up to scrutiny. The watchmen must themselves be watched. So whilst this report starts from the principle that an external independent inspectorate is needed, it also finds that significant changes should to be made to the way in which Ofsted conducts school inspections to make it as effective as it both should be and needs to be in future if educational standards are to increase.

These changes also need to recognise and work within the system referred to above – a system of increasing autonomy for all schools, but particularly in the light of vastly increased numbers of Academies, and the establishment of numerous new free schools. Any changes also need to reflect the fact that 8 out of 10 schools are already judged as Good or Outstanding by Ofsted. Important too is that schools are the leaders of their own destiny and should decide themselves what are the best methods of teaching their young people. 'What does Ofsted want?' should be a phrase that is not banished completely, but rather is one that withers away through lack of use because it is schools who increasingly decide what they want. Ofsted also needs to operate in an environment where public finances are tight, and where, as an organisation, it can achieve more with less.

Finally, Ofsted needs to address head on the criticism it receives within the system. All organisations will inevitably grumble about their regulator – that

conflict between what can be termed ‘producer interest’ and ‘consumer interest’ is at the heart of regulation. And a regulator that does not make its sector feel scrutinised is almost certainly not doing its job effectively. But the genesis of this report – and indeed its conclusions – come from a belief that schools’ lack of faith and confidence in Ofsted goes beyond this, and represents an issue for Ofsted to need to address. The fact that, during the research phase for this report, Policy Exchange received over 300 responses to the Call for Evidence from a wide cross section of schools – more than many national Government consultations – shows a widespread interest in this issue (though, of course, as a self selecting group the results should not and have not necessarily be taken as representative).

This report explored two main themes within the current model of school inspections. First, are the judgements made by inspectors accurate? This includes an assessment of what Ofsted looks at in schools (including lesson observations), and also the quality of the inspectorate. And secondly, what are the impacts of Ofsted judgments on schools, and are these fair and proportionate?

The conclusions of the report are clear. Overall, the report concludes that the four main judgements which Ofsted makes about each school – around the Achievement of Pupils, Quality of Teaching, Leadership and Management and Behaviour and Safety – are the right ones. However, there is a real variance in the respective importance of each of these criteria. There is more data in schools than ever before about progress and achievement of pupils, and both the data analysis undertaken for this report and responses to the call for evidence were clear that the data on Achievement is a significant driver for the overall result of the school. Indeed, the Achievement of Pupils sub grade correlates with the overall grade in around 99% of cases. For schools graded as a 3 (Satisfactory/Requires Improvement) or a 4 (Inadequate), this is the strongest correlation by some way. An analysis of a subset of Ofsted inspection results between September 2012 and June 2013 (around 1,200 reports) also found that in schools where the subgrades varied, the Achievement score was the one which correlated with the overall grade – indeed, in a number of cases, this held even when the Achievement grade was the only subgrade that correlated with the overall judgement. In other words, this analysis, and the feedback from the consultation, seem to suggest that pupil data drives the Achievement sub grade, and the Achievement subgrade drives the overall grade. Quality of Teaching, and Leadership and Management, also correlate quite highly at the top end (in many Outstanding schools, all three of these subgrades are typically Outstanding) but less so at the lower end. Behaviour and Safety remains the only subgrade with limited impact on the overall judgement. This data analysis was supported by Ofsted inspectors interviewed during this report, and responses from our Call for Evidence, where some schools reported being told explicitly the initial judgement formed by the inspection team before they had even stepped through the gates.

The report concludes that an over-reliance of data is not without risks – particularly for smaller schools, and for schools that do not use National Curriculum levels. But on balance, the conclusion is that the data – when properly moderated, and when scrutinised by inspectors fluent in data analysis – offers much potential for assessing schools in a valid and reliable way.

The report also concludes that lesson observations – which take the majority of an inspection in terms of time and money – are neither valid nor reliable in their

present form. Although the purpose seems sensible – to validate the quality of teaching in a school and check how young people are learning – the report is unequivocal in concluding that observations in their current format cannot make such a judgement, and that the consequences that flow from the practice of observations – whether it is schools preparing checklists of ‘Outstanding lessons’, conducting mock inspections, or teachers preparing ‘Ofsted lessons’ – are all both nugatory and avoidable. Ofsted has recently clarified its approach towards observations, and in particular that these should not be graded, noting rightly that a maximum 20 minute observation with little context cannot make an accurate judgement. Yet this report concludes that this clarification does not go far enough. So long as inspectors are observing lessons, they will – in practice – still be making judgements on the teacher and on the quality of teaching in a school, even if such individual judgements are not graded or even shared. Such judgements also run the risk of – as the long running debate over eliminating the ‘Ofsted preferred teaching style’ shows – always falling prey to confirmation bias and the whims of individual inspectors. Moreover, the mere fact of having a lesson observation at all – and Ofsted using this method to assess Quality of Teaching – belies the oft repeated statements that schools should be leading their own judgements of their schools, and for Ofsted to play a scrutinising and validating role. Why should a school conduct carefully planned and exhaustive processes around monitoring and assessing its own staff’s performance, if Ofsted can make a judgement that trumps with a series of 20 minute drop ins? And if the Ofsted judgement does not trump a school’s, why are they even doing it in the first place? This report concludes that the practice of lesson observations is symptomatic of many of the issues related to the balance of power between inspectors and schools, **and recommends – as part of a wider reform to the structure of school inspections – the total abolition of all routine lesson observations by Ofsted in the course of their standard inspections.**

“So long as inspectors are observing lessons, they will – in practice – still be making judgements on the teacher and on the quality of teaching in a school, even if such individual judgements are not graded or even shared”

The report also considered issues around the quality of the inspectorate. It finds concerns with two specific areas in particular – an ability to understand, interpret and draw conclusions from statistical data, and a mindset in some inspectors of a way in which schools should run and an unwillingness or inability to engage with different structures. The first of these is particularly problematic given this report’s belief in the increasing importance of data, and it is concerning that nowhere in the current requirements for Ofsted inspector recruitment is there an explicit reference to data or statistics. On the second, this goes beyond the caricature of ‘the Blob’ used by many to describe Ofsted inspectors – though it is worth noting the work of people like Daisy Christodoulou and blogger Andrew Old who have both catalogued previous Ofsted reports extensively to show a consistent preference for a certain type of teaching style or organisational format. The more pertinent point is around a variability in ability to engage with schools who do not follow what might be termed a ‘mainstream’ model – either because of their size, or their newness, or their particular pedagogical style. As one of the Headteachers who responded to the call for evidence put it, “you don’t know who’s



going to walk through the door so you have to be ready for anything and prepare for the lowest common denominator”. Such an approach is – understandably – one that makes many schools furious.

The report also notes the particular structure of Ofsted in school inspections with regards to this complaint. The vast majority of inspectors, called Additional Inspectors (AIs), do not work directly for Ofsted but are outsourced to one of three regional contractors (known as Regional Schools Inspection Providers or RISPs) – currently CfBT, Serco, and Tribal. The current value of these contracts is around £30m a year in total and there are around 3000 AIs in total, of which around 1,500 carry out school inspections. Ofsted directly employs some inspectors, known as Her Majesty’s Inspectors (HMIs), but there are only 300-400 of these, of whom around 140 work on schools. For the vast majority of schools therefore, their experience with ‘Ofsted’ will in fact be with Additional Inspectors working for contractors. Although outsourcing models can and often do work effectively in public services, this report raises significant concerns with the current structure of the contracts, and in particular what it terms the ‘twice removed’ form of accountability; as many AIs will themselves work under contract to the RISP and commit merely a few days a month in amongst various other jobs. This report therefore suggests that one of the reasons for the relative lack of success of the current HMCI – and indeed his predecessors – in driving much needed change through the inspectorate, is the loose levers of accountability and responsibility he has to the ultimate front line inspectors. This is worrying both from an efficiency and effectiveness point of view, and **this report recommends that, as part of the wider shake up of school inspections, Ofsted consider carefully either abolishing or drastically reducing the dependence on Additional Inspectors for future school inspections – and if the latter, requiring AIs to work full time directly for the RISP, to strengthen cohesiveness of the inspectorate.**

The report also summarises the concerns that many schools raise about the impact of Ofsted before and after an inspection. Such concerns range from high levels of stress and nugatory planning immediately beforehand, to more substantial concerns around a lack of incentives for many schools to conduct major strategic change or innovation within a system, purely because of a fear of how Ofsted may judge it. This lack of innovation is particularly problematic for new schools entering the system – whether free schools or new sponsored Academies – and any schools wishing to make changes, which may be either ‘traditional’ or ‘progressive’. The consequences after an inspection are often very severe, with a very common occurrence being a change in Senior Leadership team roles as well as wider staffing. To be clear, such consequences are often merited and indeed welcome from a system improvement point of view. The report is also careful to distinguish – as were many respondents – between practice driven by Ofsted, and poor practice by weaker schools who have a tendency to over plan, over guess and over interpret Ofsted’s requirements with adverse consequences where Ofsted should not be blamed. However, the overall conclusion here is that such consequences seem to affect Good and Outstanding schools as well in a way which seems unhelpful or unnecessary – even more so when based on flawed inspection methods and/or inspectors.

Ofsted have made a series of important changes changes recently – a renewed effort to stamp out the preferred teaching style, bringing AI training back in house, and clarifying how inspectors grade lessons during observations. It is welcome that they recognise the need for change. But there still remain issues to resolve: that, at its core, the model of school inspections needs to structurally change, with a more tailored inspection process for weaker schools and allowing stronger schools to lead more with their process of self evaluation. Ofsted have very recently indicated a willingness to consider that. Our report’s conclusion is that this cannot be half hearted. It must fundamentally recast the relationship between the inspectorate and schools.

“It is welcome that Ofsted recognise the need for change. But there still remain issues to resolve: that, at its core, the model of school inspections needs to structurally change”

This report sets out how such a significant shift could take place. The starting point is to move away from a view that every school should undergo the same inspection process, albeit with different frequency. **Ofsted should move away from the standard section 5 inspection which every school undergoes at some point, and replace it with a 2 stage inspection process.** The first of these which the report terms a ‘Short Inspection’, should happen to every school at least once every two years. This should be heavily data driven, but also include a one day visit by a single inspector to the school. The inspector’s job should be strictly defined as to validate the head and governing body’s own assessment of their school through its own process of self evaluation. Importantly, it would include no lesson observations at all of teaching staff. The inspector would be expected to validate the school’s judgements on the four subgrades through discussion with the headteacher, scrutinising the data, and checking all the school’s processes and judgements on things like performance related pay to assess their judgement of quality of teaching. An inspector could also meet with a selection of staff, parents and pupils if they wished. They would be able to visit lessons if they wished but would not complete any lesson observation forms at all and, specifically, not make any judgements whilst observing. In practice, given time and manpower constraints, this would be minimal.

At the end of a Short Inspection, the inspector would produce a brief public judgement. This report recommends that this should include an overall graded judgement on the school, and together with a new combined judgement on what is termed ‘school capability’, which would grade a combination of Achievement, Leadership, Teaching and Behaviour.

Should this judgement be that the school is Good or Outstanding for both criteria, that would be end of the process. Should an inspector judge that either the overall grade or the capability be less than Good, or – crucially – that the data and the assessment cannot prove such a judgement – then a full inspection would take place, called a Tailored Inspection. This report recommends that this should sit somewhere between what is a normal inspection now (called a Section 5) and a more detailed Section 8 monitoring visit that follows a weak judgement on a school. Importantly, this Tailored Inspection would be longer than a normal inspection – approximately double the number of inspector days for a typical primary or secondary school. Such additional time would allow inspectors to truly understand the school, and its data, and explore issues that the school may

feel a Short Inspection could not cover (for example, over how the school assesses and tracks its pupils, in particular where schools use different methods of doing this or have an unusual pupil cohort). This Tailored Inspection would include reformed lesson observations, and this report lays out several recommendations as to how to strengthen the quality of the inspectorate and the protocols to ensure this would have benefit. This Tailored Inspection would also result in a full published report, where the school would be able to receive any grade, from Outstanding through to Inadequate. In other words, a Tailored Inspection should not be seen as a proxy for failure.

**Such a two step model would mean a significant shift of manpower away from Good or Outstanding schools to those who are at risk of being RI or below. This report estimates, however, that a considerable saving of time through abolishing routine lesson inspections would mean 12,000 fewer inspector hours would be needed across the system. Even after reallocating staff to double the coverage in schools for the new Tailored Inspections, Ofsted would save around 5% cash a year.** Such savings could be made through either abolishing or radically reducing the requirement for as many AIs to be recruited, as detailed above.

The report makes one final recommendation around strengthening the need for school led improvement. **In future, a school should not be able to be rated as Outstanding unless Ofsted judged that it was not just Outstanding in its own practice, but that it was engaged in a serious and meaningful way in some form of school to school improvement with other schools – as chosen by the school itself.** Such an approach, of course, is already practised by many schools through a myriad of routes – becoming a Teaching School, working within a Multi Academy Trust to support weaker schools, being part of a School Direct consortium, or being part of some of the sector led networks for school improvement. The intention of this change would be to systematise this model of school to school improvement across the country. Many of the future challenges in the school system will best be addressed by schools working in partnership, as set out in one of Policy Exchange’s previous reports on school chains. Most recently, some of the difficulties experienced by some of the larger chains has demonstrated the need for smaller, localised partnerships. This is likely to particularly be the case for primary schools, which is a future area of Policy Exchange work. Such an incentive on all schools to strive for meaningful partnership will help to provide the ‘supply’ of high performing schools to meet this demand. It is, however, vital that the method chosen by the school is at their discretion. It must be able to consider the demand in its local area, and the school’s own capacity to support other schools.

The report makes a series of supporting recommendations to reform the practice of Ofsted school inspections. These are:

- 1. The Ofsted inspector person specification should be tightened so that all Inspectors should only be allowed to inspect a school when they have relevant and recent teaching experience in Special, Primary or Secondary Schools, or a high knowledge of assessment and pedagogical practice in that area.**
- 2. Inspectors should have to pass a data interpretation test in order to become accredited. Such accreditation should be time limited and regularly**

renewed – perhaps every five years. For inspectors who wish to deliver tailored inspections, they will need to be trained in lesson observations to the extent set out by MET, or as the result of a specific UK study commissioned to identify the conditions necessary for a high level of validity and reliability.

3. Ofsted should consider how to introduce additional methods to test the reliability and validity of their inspections on a randomised basis, not just when complaints or appeals are raised by schools. This could include random sampling follow up moderation days.
4. Schools' internal assessment procedures should be validated by Ofsted as to their rigour and frequency, to ensure moderation is reliable.
5. Ofsted should pilot a survey of students' school experiences, including views on teaching, bullying and safety. This pilot should be tested against other judgements made on these elements to explore its reliability and validity ahead of a possible wider roll out.
6. Ofsted should exercise more caution in publications which seem to endorse certain teaching methods.
7. Ofsted should work with the Behavioural Insights Team to trial different models of ensuring high level of parental sign up to the Parent View survey, combined with low levels of fraudulent feedback.
8. Ofsted should design a system for inspecting Academy chains.
9. Ofsted should consider carefully whether it retenders its contracts for Additional Inspectors when the contracts are re-let in 2015. Should Ofsted retender the contracts, it should place a condition that AIs work full time for the contractor so as to ensure organisational loyalty and mechanisms for development and information flow.

This report sets out a radical new direction of travel for Ofsted. It seeks to balance the need for clear public accountability with a system in which schools must primarily be their own drivers of improvement. By introducing structural change, it moves beyond the current approach of seeking to simply request changes with no real levers to implement them, and makes Ofsted explicitly an organisation that validates, rather than makes judgements itself – moving it back to a function of a hygiene inspector, rather than that of a food critic. Importantly, such a shift would open up the potential of galvanising a new wave of school innovation and improvement, with Outstanding schools leading from the front.

---

# 1

## Why We Need Ofsted and What it Does

The names Hugh Seymour Tremenheere and Revd John Allen may not trip off the tongue of education reformers. But these two gentlemen were the first ever of Her Majesty's Inspectors of Schools (HMI), appointed in December 1839 with a brief to inspect elementary schools which had recently been given the support of public money, to ensure that this money was well spent. The grants made to the elementary schools were specifically made "conditional upon inspection". The 1902 Education Act extended this responsibility to secondary schools. Alongside the HMIs, inspections were carried out by Local Education Authority inspectors, with HMIs focussing largely on national education issues. In 1992, inspectors were centralised under a new Office for Standards in Education, or Ofsted, with a mandate to report directly into Parliament, and publish their reports on schools.

The principle of school inspections since 1839 has therefore been first and foremost of accountability for public money, and to ensure that it is being spent on the education of children and young people for whom Parliament has voted this money be spent on, in an effective a way as possible. And this principle holds as true today as it has ever done. Although the explicit focus today is about standards of education and the knowledge and skills that are acquired, the starting point for government intervention through a regulator is to safeguard public money, and protect the interests of the ultimate users of the system. The independence of such a regulator ensures that it should be beholden neither to the sector which it regulates, nor – crucially – to government.

This report, then, starts from the presumption that it is right and proper to have an independent schools inspectorate. The question is how best to conduct such regulation in a way which is necessary, fair, effective, affordable, and that enjoys public confidence.

### The current focus of Ofsted within schools – what do they look for?

The coverage of what Ofsted looks for in schools has moved on some considerable distance from when Mr Tremenheere and Revd Allen started scrutinising what schools were doing.

When Ofsted inspects a school now (under what is called the current framework), it provides a single, graded judgment on the overall quality of the school in one of four categories: Outstanding, Good, Requires Improvement, or Inadequate. As well as the overall grade, Ofsted gives graded judgements in four subcategories:

- Achievement of pupils
- Quality of Teaching
- Leadership and Management
- Behaviour and Safety of pupils

The Ofsted handbook, which details how the framework should be assessed during an inspection, also sets out that inspectors must consider the spiritual, moral, social and cultural development of pupils, as well as whether the needs of children with SEN or disabilities are met. However these four sub elements are the key judgments and which are graded individually in a school's report.

Prior to 2012 there was a broader focus for school inspections, with 27 individual graded criteria for them to be judged against, including topics around whether children's lifestyles were healthy (for example). In his letter to Her Majesty's Chief Inspector explaining the latest revision to the Ofsted framework, the Secretary of State for Education Michael Gove made clear there was a need to re-focus inspection on those things that were most important to schools:

*"As we both agree, we need to refocus inspection on the principal purpose of schools improving teaching and learning and dramatically reduce the time and energy spent on other existing bureaucratic duties"*<sup>1</sup>

It is therefore worth exploring whether these four subgrades are actually the components of what makes for an effective school.

### Achievement of pupils

If Achievement is defined in the broadest possible sense, so not limited to achievement on tests, but also including personal and social achievement and successes, then it is difficult to conclude anything other than the Achievement of pupils being the single most important thing that a school can do. Indeed, it is difficult to conclude that any of the other subgrades – or indeed, the overall grade – could rank as more important than Achievement – something which this report has found, in practice, rarely happens.

### Leadership and Management

Research shows that this school leadership is also, unsurprisingly, important:

*'There are statistically significant empirical and qualitatively robust associations between heads' educational values, qualities and their strategic actions and improvement in school conditions leading to improvements in pupil outcomes'*<sup>2</sup>

Further evidence has suggested that school leadership – and in particular the processes undertaken consistently by effective school leaders – is second only to classroom teaching as an influence on pupil learning.<sup>3</sup> This research also shows that effective leadership is more commonly distributed throughout a school, including to middle leaders, rather than concentrated in a head. On a practical level when making judgments about a school, leadership seems to be an eminently sensible measure to include.<sup>4</sup> There is less evidence about the impact of effective governance on a school in the same way – much of it surrounds

<sup>1</sup> Harrison A, *Schools inspections Slimmed Down*, BBC News, September 2010

<sup>2</sup> Leithwood K and Levin B, *Assessing School Leader and Leadership Programme Effects on Pupil Learning*, DfES, 2005

<sup>3</sup> Leithwood et al, *Seven strong claims about successful school leadership*, NCSL 2006

<sup>4</sup> Sammons P, *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*, Institute of Education, 1995

self reported impact and a consensus on the key features of effective governing bodies as defined by schools which have been defined as Outstanding in their governance. In practice, the governance of a school will become increasingly important as schools become more autonomous, and as governing bodies become legally responsible for staff, land, contracts and the delivery of services, so it seems reasonable to assess their capacity and capabilities.

#### Quality of Teaching

The dominance of the Achievement subgrade within Ofsted's overall grade structure (discussed further below) reflects a widespread view that the learning achieved by young people within a school (defined broadly) is the most important outcome of that school. But the teaching within the school – as the primary transmission of knowledge and skills to young people – is also vital. The school effectiveness research is clear that the quality of teaching is the single biggest variable on performance within a school, and that there is a close and significant

“The dominance of the Achievement subgrade within Ofsted's overall grade structure reflects a widespread view that the learning achieved by young people within a school is the most important outcome of that school”

link between the quality of teaching in a school, and the learning experienced by children; studies have also shown that, for a child from a disadvantaged background, the difference between a good quality and a poor quality teacher can be the equivalent of a year's worth of learning.<sup>5</sup>

#### Behaviour and Safety of Pupils

Again, it is not stretching a point too far to argue that the safety of pupils – while not the purpose of education – is fundamentally important in facilitating education. Additionally it is perhaps in this category that the requirement to have consideration for ‘the spiritual, moral, social and cultural development of pupils’ is made most explicit; there is reference in the framework to whether leaders ‘have created a positive ethos in the school’ and judgement of ‘pupil's behaviour towards, and respect for, other young people’. It is this category which tries to grasp some one the subtleties of whether the school is a happy one, alongside judgments about whether it is a competent one.

However there is also an assumption in the current framework that behaviour is a key factor in learning. Inspectors are asked to judge ‘pupils' attitude to learning’ and how well teachers manage behaviour ‘to ensure that all pupils have an equal and fair chance to thrive and learn’. There is some evidence to show the link between poor behaviour and achievement, both from the individual:

‘Analysis of data from the Longitudinal Study of Young People in England (LSYPE) shows that after controlling for a wide range of pupil and school factors self-reported misbehaviour in most classes has a negative impact on predicted Key Stage 4 (KS4) attainment: predicting a capped GCSE score of 29 points fewer than those who did not report their own misbehaviour (equivalent to gaining one grade lower in five subjects).’<sup>6</sup>

<sup>5</sup> *Improving the impact of teachers on pupil achievement in the UK – interim findings*, The Sutton Trust, September 2011

<sup>6</sup> *Pupil behaviour in schools in England*, Research Report DFE-RR218, DfE, 2012



And for the behaviour of those in their class:

*'Analysis shows that after controlling for a range of factors there are some associations between reports of self and others' misbehaviour and being not in education, employment or training (NEET) between the ages of 16 and 18: for example, misbehaviour or troublemaking by others in classes was amongst factors associated with being NEET at age 17.'*<sup>7</sup>

However, there is a less clear cut case in the literature that behaviour and learning are correlated than there is with teaching or leadership – most often, because behaviour itself can be a proxy for other factors that hamper learning.

Overall, therefore, the report finds that the current framework with an overall grade and four sub graded criteria above are the correct ones for Ofsted to assess when looking at the quality of school, and it is not recommended that these change. The important issue here is the consideration of how these factors can be assessed.

### How does Ofsted evaluate how schools are performing?

When looking at Ofsted's methods, it is important to establish whether the techniques that they use produce results which are both valid and reliable.

#### Validity

If a judgement is **valid**, it just means that it corresponds with what is empirically the case – in other words, that is accurate. This might be easy to establish for some things (for example, it is easy to assess whether a set of weighing scales gives a valid judgment as to an object's weight), but is more difficult in the case of the quality of schools. To establish whether Ofsted judgements are valid, we need to be able to compare them to something completely separate to see if that gives the same judgement about a school, and this is very difficult.

One way of testing whether an Ofsted judgment is valid, or accurate, might be to see whether it correlates to exam results – as these could be seen as an independent measure of school quality. However, this would be problematic as exam results are always included as part of Ofsted's judgment process – if a school gets very good results, this will in (large) part determine Ofsted's judgment of the school. So exam results cannot be used as an external measure of validity. Moreover, even if the findings diverged, this would not necessarily show anything about validity – as an inspection is not simply looking at pupil outcomes but is also looking at the full set of processes in a school (and also considers pupils not covered by one set of exam data).

One piece of research does show a strong correlation between Ofsted ratings and the ratings that pupils and parents give to their school/the school their child attends.

*'The association between inspection ratings and student survey reports of teacher practice is economically meaningful and statistically significant, even after conditioning on the school's test rank, proportion of students eligible for a free lunch and other school characteristics. This implies that students enrolled in schools with better inspection ratings experience an environment where, according to student self – reports, teacher practices are superior. Similar findings hold for survey measures of parent satisfaction'*<sup>8</sup>

<sup>7</sup> Ibid

<sup>8</sup> Hussain I, *Subjective Performance Evaluation in the Public Sector: Evidence from School Inspections?*, Centre for the Economics of Education Discussion Paper No. 135, 2012



It is important to note that these findings control for previous Ofsted inspections – so the views of parents and children are not simply reflecting the ‘official’ quality rating of the school. There are of course problems here, as Ofsted also conducts some parental and student surveys and so we perhaps should not be surprised if there is a strong correlation between the two, although surveys tend to only cover a small proportion of the school community. Parents and students could also be making an ‘invalid’ judgement – but student surveys have also been shown as a valid measure when correlating with teacher effectiveness

So there is one control measurement to look at the validity of an Ofsted judgement. Overall, however, given the comprehensiveness of the Ofsted framework, it is difficult to prove in aggregate whether judgements are empirically valid, or empirically invalid.

### Reliability

The **reliability** of Ofsted reports has been a frequent cause for complaint. During the evidence gathering phase of this report, many teachers raised the concern that there was no consistency between different inspection teams, which made it difficult for schools to assess Ofsted judgements. It is important to stress here what reliability should and should not mean in an inspection. Occasionally, the report has heard evidence around reliability being defined as an expectation that two schools with similar circumstances and achievement ought to receive similar Ofsted grades. Ofsted themselves firmly reject this definition of reliability, emphasising that inspector discretion must be paramount. This report agrees with that. The definition of reliability used here is that *if two different inspection teams came into the school on consecutive days (or a suitably short time period so that no internal factors would significantly change), they would give the same judgement.*

Judging reliability – like validity – is also difficult to determine. Unlike some other regulators, for example the Health and Safety Executive, Ofsted does not have peer review days to verify inspections. It is also not possible to compare different inspections in the same school that are two or three years apart, as either the framework or the circumstances of the school, for example a leadership change (or often both) will make such a comparison flawed.

Determining reliability by comparing the judgements made by different inspectors is also difficult. The overall judgement is made by the Lead Inspector and there is no indication of whether a team was unified or dissented on a judgement. There is also no openly available information on Ofsted’s own quality control (as we shall discuss in more detail in the next chapter). While all reports do have to be quality assured by Ofsted senior management, to make sure that the overall judgment seems to be supported by evidence, there are real concerns from teachers and headteachers that inspectors’ own opinions and ability to collect and interpret evidence is very varied – in other words, that reliability between inspectors is low.

So for the inspections as a whole, there is limited evidence available to say whether Ofsted judgements are either valid or reliable. This is not the same, however, as saying there is evidence that they are invalid or unreliable. To further understand this issue the report now looks in more detail at the individual aspects of inspection methodology, to see whether they are likely to provide accurate assessments of quality within a school.

---

# 2

## The Practice of Lesson Observations

Lesson observations are perhaps the most symbolic and well recognised element of school inspections for most schools, particularly classroom teachers. They also take a significant proportion of time for a typical inspection. It is therefore important to understand exactly what the benefits of this approach are, be clear on how lesson observation is used by Ofsted, and consider how they might operate in the future.

Lesson observations can take place in a number of ways, with short visits to lots of lessons for a few minutes, short observations of group teaching, longer observations of more than 25 minutes, or by tracking a group of pupils throughout their school day. Lesson observations may be undertaken jointly with members of senior staff of the school.

When it comes to using the evidence from observations, the handbook states that:

- The key objective of observations is to evaluate teaching and its contribution to learning, particularly in the core subjects, and that these can be graded in a lesson observation. So while a lesson itself cannot be graded, an individual teacher may be given a grade for their teaching along with feedback based on an observation.
- To reach a judgement on teaching, inspectors must not simply aggregate the grades awarded following lesson observations. They must use other evidence on progress over time as well.
- If a lesson observation is short, inspectors might not give a grade but should use it as part of their evidence base.

There is a growing body of evidence around the effectiveness of evaluating teaching through observation, particularly since the seminal Measuring Effectiveness of Teachers (MET) study in the United States, which looked at (amongst other things) the reliability and validity of evaluation methods. This large study (which cost over \$50m) looked at 3,000 volunteer teachers and a variety of different ways of measuring their effectiveness, one of which was lesson observations. Of course the frameworks used in their observations were not the same as the Ofsted framework, and do not have the same number of judgements (one of the ones used, the CLASS framework for example, has seven grades, although these are separated into three descriptors of low, middle and high).<sup>9</sup> Nevertheless, the

9 Hamre et al, *CLASS Implementation Guide*, December 2009

research findings from MET are highly relevant in this context, in assessing both the validity and the reliability of lesson observations as a critical element of the current Ofsted school inspection framework.

### Validity of Observations

Recently there has been growing concern about the validity of lesson observations alone as a way of judging teaching quality. Much of this has been led by Professor Rob Coe of Durham who has worked using the MET data (and the MET calculations on the accuracy of observers, as well as other data <sup>10</sup>) to

“The stark conclusion is that when comparing a lesson judgement of teacher quality against its ‘actual’ quality, the lesson judgements often do not tally”

estimate the validity of observations carried out by Ofsted inspectors, making best case and worst case assumptions on Ofsted validity. His findings are set out below. The stark conclusion is that when comparing a lesson judgement of teacher quality against its ‘actual’ quality (as defined by the Value

Added progress made by pupils in that class), the lesson judgements often do not tally. In fact, overall the results are worse than flipping a coin – there is a 49% chance that the quality of the lesson will be empirically the same as judged, and a 51% chance that it will not be the same as that assessed by an observation. For lessons on either extreme (either Outstanding or Inadequate), the accuracy falls away even faster – with at best a 71% chance and an 83% chance that an observed judgement of a lesson into one of these two categories is wrong.<sup>11</sup>

**Table 1: Validity of lesson observations (Coe)**

1st rater gives	% of lessons considered this grade	Probability value-added data disagrees	
		Best case	Worst case
Outstanding	12%	71%	96%
Good	55%	40%	45%
Requires Improvement	29%	59%	79%
Inadequate	4%	83%	>99%
<b>Overall</b>		<b>51%</b>	<b>63%</b>

Other research supports this claim. In an overview of teacher evaluation methods the Sutton Trust state that:

‘Even when conducted by well-trained independent evaluators, classroom observations are the least predictive method of assessing teacher effectiveness’.<sup>12</sup>

There is some evidence to the contrary. The Effective Pre-school and Primary Education Project 3–11 (EPPE) conducted a range of observations of primary school lessons and found that overall the scores given in their observations for teaching quality were a significant predictor of greater cognitive progress between the ages of 6 and 10, suggesting that classroom observations can be

10 Strong, M., Gargani, J., & Hacifazlioglu, O. *Do we know a successful teacher when we see one? Experiments in the identification of effective teachers*, 2011

11 Coe, R, *Lesson Observation: It’s harder than you think*, TeachFirst TDT Meeting, 13th January 2014

12 Murphy R, *Testing Teachers: What works best for teacher evaluation and appraisal*, The Sutton Trust, March 2013

a valid measure.<sup>13</sup> However these observations were carried out according to internationally developed teacher observation frameworks, and by observers who had shown a relatively high level of ‘inter-rater reliability’.<sup>14</sup> Furthermore while some studies have found that headteachers are able to identify teachers at the extremes of effectiveness (if not in the middle of the distribution), this is based on headteachers evaluating teachers’ work ethic, parent satisfaction and a range of other factors, which will be very different to the process undergone during an Ofsted inspection.<sup>15</sup> In summary, therefore, it seems clear that the overall validity of lesson observations are – at best – questionable.

Additionally, the potential inaccuracy of lesson observations was drawn from the MET study which was based on 3,000 teachers volunteering to open up their classrooms to be part of the study; while there are still likely to be effects from being observed it seems reasonable that in the much more ‘high stakes’ and compulsory Ofsted observations behaviour will be more affected. In response to a questionnaire we sent to those working in schools, some teachers stated that they would never change their practice because Ofsted were in but many said that they felt the pressure to put on a good show, or change their teaching style ‘Teaching should not alter when Ofsted arrive but because of the pressure applied by Ofsted it invariably does’.<sup>16</sup> If we add the likelihood that at least some teachers are not teaching in their usual fashion to the already identified inaccuracy of lesson observations, it is possible that the accuracy is even lower.

### Reliability of observations

Coe’s research also explores the reliability of inspections by comparing the likelihood of different observers agreeing a judgement of the same lesson. At the top end estimate of reliability, this is based on the EPPE data for the reliability of highly trained observers, and at the bottom end Coe uses the lower end of the reliability ratings seen in the MET study, but again it is important to note that these are estimates based on his own simulations.<sup>17</sup> As with validity, the conclusion is clear that judgements on quality of teaching made by lesson observations are not necessarily reliable. Even in a best case scenario, observers would only agree 61% of the time, although this varies substantially depending on the grade of lesson.

**Table 2: Reliability of lesson observations (Coe)**

1st rater gives	% of lessons considered this grade	Probability that 2nd rater disagrees	
		Best case	Worst case
Outstanding	12%	51%	78%
Good	55%	31%	43%
Requires Improvement	29%	46%	64%
Inadequate	4%	62%	90%
<b>Overall</b>		39%	55%

13 Sylva K et al, *EPPE – Final Report from the Primary Phase*, DCSF-RB061, DCSF, November 2008

14 Sammons P et al, *EPPE – Variations in Teacher and Pupil Behaviours in Year 5 Classes*, DfES, 2006

15 Jacob B and Lefgren L, *Can Principals Identify Effective Teachers?*, Journal of Labour Economics, 2008

16 Ian Dutton, response to call for evidence

17 Professor Coe, *Lesson Observation: It’s harder than you think*, TeachFirst TDT Meeting, 13th January 2014

Importantly, the MET study found that to get to approximately the level of reliability used by EPPE – represented here as ‘best case scenario’ – would take

two 45 minute lesson observations from within the school, one 45 minute observation by an external observer and three fifteen minute observations by different external observers.<sup>18</sup>

MET also point out the importance of some full-length lesson observations, as:

‘Our results also suggest that it is important to have at least one or two full-length observations, given that some aspects of teaching scored on the Framework for Teaching (Danielson’s instrument) were frequently not observed during the first 15 minutes of class.’<sup>19</sup>

The question that emerges therefore is not just, ‘are lesson observations reliable?’, but ‘are lesson observations as conducted by Ofsted likely to be reliable?’ Although there is no publicly available data or research from Ofsted about the reliability of its observations, all the discussions we have had during our research and responses to our call for evidence show that it is highly unlikely – veering towards unprecedented – that during the course of any standard two day Section

5 inspection a single teacher would have one 45 minute observation, and then three further observations by different observers (which would be the requirement to get lesson observation to the kind of reliability levels seen in the EPPE project). This would be assuming that they had already had at least two internal lesson observations from colleagues. One headteacher we spoke to timed all the lesson observations during her school’s inspection and found that the longest lasted for fourteen minutes; many of the responses to our call for

evidence cited the brevity of lesson observations as one of the reasons they could not be used as a fair representation of a teacher’s competence.

So if lesson observations are neither valid nor reliable, why are they done, and specifically, why are – in many, many individual testimonies, even if not officially – graded judgements of teaching made on the back of them? In recent months, during the research phase of this project, something of a campaign has built up against this practice, largely through the growing use of Twitter and blogs to share pedagogical discussions. The Headteachers’ Roundtable group have set this as a specific topic for exploration<sup>20</sup> and other leading bloggers and educationalists such as Andrew Old and David Didau have also been writing about this issue. Most recently, this culminated in a clarification letter being issued by Mike Cladingbowl, Ofsted National Director for Schools, around why Ofsted inspectors observe lessons and the purposes of it. The key excerpts are set out below:

“Since 2009, inspectors have been instructed not to grade the overall quality of a lesson they visit. As you can see from [the evidence form completed by inspectors], the box for a graded ‘judgement on the overall quality of the lesson’ has been removed.

[Why do inspectors observe lessons?] It’s just one piece of a jigsaw of evidence about the work of the school that includes: the school’s own observations and self-evaluation, joint visits to classrooms with the headteacher or other staff, evidence about how teaching has improved,

“One headteacher timed all the lesson observations during her school’s inspection and found that the longest lasted for fourteen minutes; many responses cited the brevity of lesson observations as one of the reasons they could not be used as a fair representation of a teacher’s competence”

---

18 *Ensuring Fair and Reliable Measures of Effective Teaching, Culminating Findings from the MET Project’s Three-Year Study*, Bill and Melinda Gates Foundation, January 2013

19 *Ensuring Fair and Reliable Measures of Effective Teaching, Culminating Findings from the MET Project’s Three-Year Study*, Bill and Melinda Gates Foundation, January 2013

20 *Meeting the Michaels, HMCI, SoS and HTRT discuss Ofsted, Headteachers’ Roundtable*, February 2014

the quality of work seen in books, teachers' marking, discussions with pupils and staff and, of course, test results and so on. In my view, inspectors must always spend time in classrooms when they inspect.

Inspectors do not judge the overall lesson. But it is still possible for an inspector to record a graded evaluation on an evidence form under one or more of the four main judgement headings, including teaching, where there is sufficiently compelling evidence gathered by observing routines, looking in books, listening to students and so on. It might be possible, for example, to see evidence of the impact of a recent decision taken by the leadership, which has improved behaviour. But this is categorically not the same as judging a teacher, or even the teaching, and especially not a lesson overall, by evaluating the performance of the teacher in a lesson or a part of a lesson. Making a judgement about the quality of teaching, based on a wide variety of evidence gathered in the classroom and elsewhere, is not the same as judging how well a teacher performed. I know this may sound like splitting hairs – but it is an important difference.

Inspectors should not grade an aspect such as teaching, unless circumstances are exceptional, without considering the broad range of evidence that they can gather during a visit to a lesson – for example, the behaviour of the students and how well they are managed, subject knowledge, the standard of work completed in books, the quality of marking and so on – and use this to come to a view about what teaching is like for those students and its impact on their learning over time.<sup>21</sup>

This is undoubtedly a step forward. In particular, the explicit clarification (of existing policy since 2009, admittedly) that inspectors should not judge or grade an overall lesson is helpful. During this report's evidence gathering phase, dozens and dozens of responses were submitted discussing the difficulty of lesson grades – that they were felt unhelpful, that they drove consequences, that they were unreliable, but that all too often were used by schools and school leaders because “it is what Ofsted do” or “it is what Ofsted want”. To be clear, there should be a distinction made between genuine practice that Ofsted requirements drive, and poor leadership in schools misinterpreting Ofsted's approach. Ofsted cannot and should not be blamed for the latter – as was readily acknowledged during both our teacher roundtable and the heads' roundtable. But as will be discussed later in the report, given how much Ofsted drives school behaviours, an explicit clarification like this – which should leave schools in no doubt that Ofsted will not grade lessons formally, and they ought not to either – is helpful.

However, there remain three distinct problems with the current practice even after the “Claddingbowl clarification”:

There is still considerable nuance on what is graded, which may blunt its impact for teachers and schools

The key passage from the Ofsted letter is here:

It is still possible for an inspector to record a graded evaluation on an evidence form under one or more of the four main judgement headings, including teaching, where there is sufficiently

“Inspectors should not grade an aspect such as teaching, without considering the broad range of evidence that they can gather during a visit to a lesson”

21 Claddingbowl M, *Why do Ofsted inspectors observe individual lessons and how do they evaluate teaching in schools?*, Ofsted, February 2014

compelling evidence. . . . . But this is categorically not the same as judging a teacher, or even the teaching, and especially not a lesson overall, by evaluating the performance of the teacher in a lesson or a part of a lesson. Making a judgement about the quality of teaching, based on a wide variety of evidence gathered in the classroom and elsewhere, is not the same as judging how well a teacher performed.

Mike Cladingbowl (accurately) describes the risk of this distinction being seen as ‘splitting hairs’. Because under the current arrangement, lesson evidence forms will have a judgement space left for inspectors to be able to make a grade on Quality of Teaching. It is reasonable to assume that at least some teachers and schools, when receiving feedback on their lesson, may lose the distinction between “this lesson was Good” (ought not to happen), “you are a Good teacher

on the basis of this lesson (ought not to happen), and “I cannot pass judgement on you or the lesson, but I have marked in the Evidence Form that, having seen your lesson, I am minded to say that overall teaching in this school is Good” (could happen). In practice many teachers who receive teaching grades

may feel that this is – effectively – a judgement on their lesson or their quality overall. Similarly, if heads who are inspected know that the Quality of Teaching grade is made up of evidence formed – at least in part – from lesson observations, it is not an unreasonable leap to conclude that many of them will continue some form of lesson gradings in order to form their own self evaluation judgement of Quality of Teaching.

This nuance may also feed into confusion around inspection practice

Ofsted inspectors we have interviewed for this project have made it very clear that they understood (even before Mike Cladingbowl’s clarification) that lesson observation alone is not an effective way to judge progress. For example Mary Myatt, a school improvement adviser who leads inspections, writes on her blog:

*‘A 20 minute observation can only tell part of the story, so what else needs to be taken into account? In a nutshell it must relate to students’ progress over time. As a result the quality of teaching judgement links closely to the judgement on achievement. If a ‘good’ or even ‘outstanding’ lesson does not lead to good or better progress over time, then it follows that the quality of teaching is likely to require improvement. And the flip side of this is that if a lesson is observed which requires improvement but the progress is good, then the judgement on the quality of teaching over time will be good’.*<sup>22</sup>

And as Ofsted make clear, the fact that observed lessons are not graded has been the practice since 2009. Yet as noted above, the overwhelming feedback from our call for evidence responses and the teacher roundtable and head roundtable discussions was that graded lesson observations have been, and continue to, take place. The only reasonable conclusion that can be drawn from this is that Ofsted policy is not reliably feeding through into inspection practice on the ground.

The example of Ofsted trying to clarify that there is “no preferred teaching style” shows the difficulties that Ofsted (by which we mean here HMCI, his

“In practice many teachers who receive teaching grades may feel that this is – effectively – a judgement on their lesson or their quality overall”

---

22 Myatt M, *Why lesson observations only count for so much*, October 2013



senior team, and the central organisation HQ) faces in ensuring consistency amongst its inspectorate:<sup>23</sup>

- April 2012 – Michael Wilshaw, newly appointed HMCI, states that “We, and in that word “we” I include OFSTED, should be wary of trying to prescribe a particular style of teaching.”
- Summer 2012 – Ofsted handbook amended and phrases added including “Inspectors will not look for a preferred methodology” and “Inspectors must not expect teaching staff to teach in any specific way or follow a prescribed methodology.”
- November 2012 – HMCI speech at London Festival of Education states that “Let me emphasise again to anyone who hasn’t heard this from me or from anyone else in OFSTED. OFSTED does not have a preferred style of teaching”.
- March 2013 – a letter to inspectors complained that “inspection reports often contain phrases that give the false impression that Ofsted expects teaching to occur in a particular way.”
- September 2013 – a speech by the Secretary of State to Policy Exchange says that “Ofsted’s guidance provided too little clarity about what constituted good teaching; or allowed inspectors’ personal prejudices and preferences to be interpreted as ‘the Ofsted way’ ... The good news is that Ofsted – under its inspirational new leadership – is moving to address all these weaknesses.”
- December 2013 – new guidance to inspectors points out that “Inspectors must not give the impression that Ofsted favours a particular teaching style ... Do not focus on lesson structure at the expense of its content ... Do not expect to see ‘independent learning’ in all lessons and do not make the assumption that this is always necessary or desirable ... Do not criticise ‘passivity’ as a matter of course and certainly not unless it is evidently stopping pupils from learning new knowledge or gaining skills and understanding.”
- January 2014 – in a letter to inspectors, HMCI repeats – again – his statement (by now, explicitly pleading) around a preferred style “please, please, please think carefully before criticising a lesson because it doesn’t conform to a particular view of how children should be taught.”

And yet there are specific instances of recent reports – even post the January letter, at least the seventh time in two years that this point has been made – where judgements have been made that include specific comments on a teaching style. Without exception, these have sought to identify as areas of weakness a teaching style that is not, in the inspector’s judgement, as they would wish to see. There are also instances where reports have been rewritten, or held up prior to publication, because of these judgements. We heard frequently in our call for evidence about the dominance of a preferred teaching style narrative in many observations. A large number of respondents spoke of the need – or presumed need – to have plenaries, display progress in twenty minutes, do lots of activities and so on.

In other words, an assumption that this latest clarification around lesson observations grading will end poor practice amongst either individual inspectors or schools can be most kindly described as optimistic.

23 The timeline of previous instances of Ofsted guidance is taken, with thanks, from the work of Andrew Old, and specifically his expert view for the National Union of Teachers [www.teachers.org.uk/files/andrewold-nutarticle.docx](http://www.teachers.org.uk/files/andrewold-nutarticle.docx)



### The Ofsted handbook and other Ofsted guidance still offers some mixed message around what ought to be observed in lessons

The handbook does seem to place emphasis on making judgements of Quality of Teaching through actually observing progress in a lesson, referring to a requirement to:

*‘gather evidence about how well individual pupils and particular groups of pupils are learning and making progress, including those with special needs, those for whom the pupil premium provides support and the most able, and assess the extent to which pupils have grown in knowledge’*

Which could suggest that progress can be observed in an individual lesson, and also that when observing a lesson:

*‘inspectors should grade, where possible, key judgments such as achievement and teaching, indicating in particular the growth in students’ knowledge and the quality of their learning’.*

And consider whether:

*‘pupils’ responses demonstrate sufficient gains in their knowledge, skills and understanding, including of literacy and mathematics’.*

It is not hard to see how this could translate into wanting to see students ‘progress’ in a twenty minute observation, leading to fast-paced lessons and plenaries being employed. As one respondent to our survey put it:

*Ofsted’s “line” now is that they are not looking for any particular type of teaching. However that isn’t the case, everyone knows that their current buzzword is “pace”. This means that we have to assume children have the attention span of goldfish and plan incredibly energetic lessons with “bite size” activities. At the same time we are told not to change our lessons just because they’ve come in.<sup>24</sup>*

The Annual Report attempted to clarify what is meant by pace. It stated that *‘Pace – a belief that the faster the lesson, the better the learning. While pace is important – pupils may lose concentration in a slow lesson – teachers concentrate too often on the pace of the activity rather than the amount of the learning.’<sup>25</sup>* Yet all this clarification really does is suggest that pace should be balanced against other objectives. It reiterates that, in and of itself, pace is important. Again, given this confusion, it is not difficult to see why confusion may rise whereby teachers prioritise pace in a lesson. Furthermore, the letter from HMCI to inspectors this year asked them to consider *‘Is the pace of the lesson good because the teacher is proactive and dynamic in the classroom?’<sup>26</sup>* So there remain mixed messages.

Given the nuance in the current framework, the confusion over some of the messaging in the handbook, and using the example of the on-going debate about preferred teaching style to show just how hard it is to have consistent judgements on teaching, it is difficult to conclude anything other than the practice of graded lessons may well continue, despite all its flaws, if the current Ofsted inspection set up continues.

---

24 Teacher, response to Policy Exchange consultation

25 Ofsted Annual Report 2013

26 Letter reported by TES *‘Irritated’ Wilshaw writes to inspectors to tell them not to prescribe ‘teaching styles’*, TES, January 2014. Details of the letter at [www.heatherleatt.co.uk](http://www.heatherleatt.co.uk) Extract from *Sir Michael Wilshaw’s message to inspectors regarding teaching*

In Mike Cladingbowl's recent letter, Ofsted raise the possibility of changing the Evidence Form so that no element of the Quality of Teaching can be formally recorded as part of a lesson observation, and instead inspectors would "bring [the evidence gathered about teaching] all together at the end of the inspection in a plenary before discussing the single overall judgement on teaching with the school". This would move the agenda further on, in that it would address the first point of the three raised above. Teachers would not receive any graded feedback after a lesson because no grade (even on the overall Quality of Teaching) would have been made. Yet the two other problems would still remain.

Most notably, any single judgement made by inspectors would still remain – must remain – susceptible to the flaws outlined above, both in terms of its lack of validity and reliability, and the danger of inspectors applying their own personal preferences to what they have sense. In fact, in a worst case scenario all that would happen is that a series of flawed judgements would be made throughout the visit during observations, which would be stored in inspectors' heads, not shared with the school, and an overall judgement that compounded all those flaws would emerge from an opaque discussion between inspectors right at the end of the inspection and handed down to the school with little chance of discussion or debate.

At its heart, any practice of an external observer judging Quality of Teaching in the way it is structured at present is unlikely to command the confidence of schools, and is likely to drive perverse consequences. This is why this report recommends a more widespread change to the practice of school inspections, with the removal of lesson observations as standard practice from all schools during an initial inspection.

“At its heart, any practice of an external observer judging Quality of Teaching in the way it is structured at present is unlikely to command the confidence of schools, and is likely to drive perverse consequences”

---

# 3

## Other Methods Used by Ofsted to Inspect Schools

### Scrutiny of data

Before an inspection team even gets to a school, they will be presented with a vast array of data from the school to help them form an opinion; not simply exam or key stage test or assessment results but also a breakdown of how all pupils are progressing through their education, and a particular focus on children eligible for the pupil premium and those with SEN. This data is used by Ofsted to gain an understanding of how successful – principally – the achievement of the pupils in the school is.

When it comes to the weight it is given at present, the strong consensus was that it played a very significant role. Many respondents to our Call for Evidence

and in the roundtables argued that Ofsted used the data to form a judgement which they then used the inspection in school to disprove or (more likely) prove. Some gave evidence of having heard this explicitly, with one respondent saying “The inspector came in, and met the staff on the first morning...he said ‘I’m sure this is a Good school, I’m here to find evidence to

support that fact’”.<sup>27</sup> In other words, Ofsted inspectors operate an explicit or implicit confirmation bias – they know what they think of the school before they come in, and look for evidence to match that assessment. One teacher told us that “Ofsted are results driven – if the data is positive, they want to give a good judgement, and vice versa”.<sup>28</sup> In discussions with Ofsted inspectors, some argued that data is only ever one part of how a judgement is reached, but others saying that it is very hard to get a judgment approved if it does not match what the data says. Headteachers generally echoed that last point – a typical observation was that “Ofsted’s own quality assurance is so tight that they will only let judgements through where they’re supported by data”.<sup>29</sup>

New research for this report backs up this view. The charts below show the proportion of schools where the overall grade for the school agrees with each of the subgrades (using a snapshot of data from inspections that took place between September 2012 and June 2013).

Here we can see that the Achievement subgrade agrees most strongly with the overall grade, for both primaries and secondaries, followed very closely by the Quality of Teaching subgrade. Although these judgements are not solely determined by data (for teaching, in particular, lesson observations are meant

“Ofsted inspectors operate an explicit or implicit confirmation bias – they know what they think of the school before they come in, and look for evidence to match that assessment”

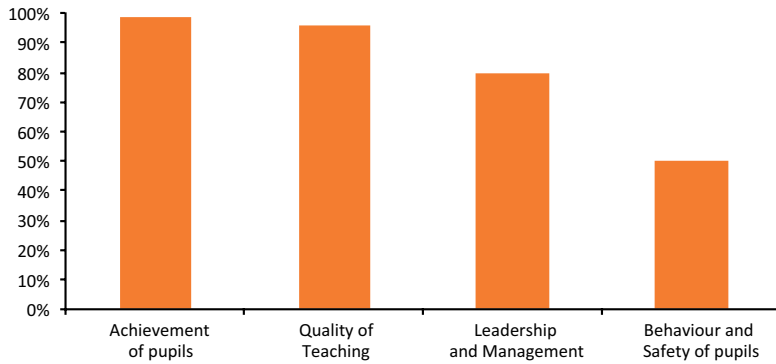
27 Attendee at Policy Exchange roundtable

28 Attendee at Policy Exchange roundtable

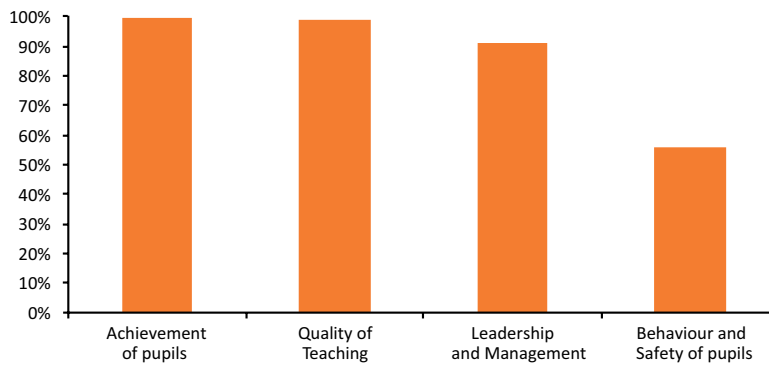
29 Attendee at Policy Exchange roundtable

to play a role), the handbook makes clear that the data – from external tests or internal school data, including progress made by various sub groups – ought to inform these two judgements heavily.

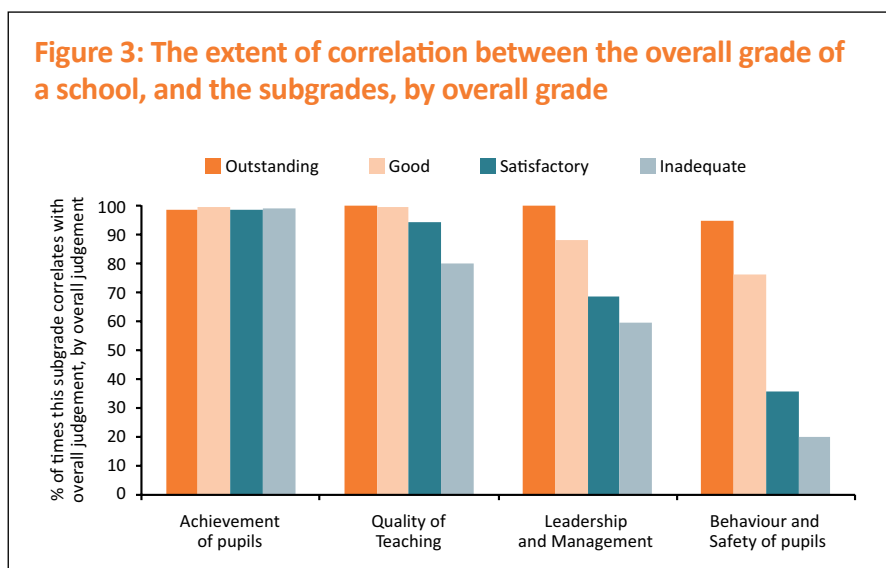
**Figure 1: Agreement of subgrades with overall grade – secondary schools**



**Figure 2: Agreement of subgrades with overall grade – primary schools**



If this data is broken down by overall grade of the school (where primaries and secondaries are combined), we can see the following:



This shows that the lower rated the school overall, the more likely it is that the Achievement grade has been a driver of the overall grade – whether the other grades have been more positive or more negative. For example:

- Of the 109 Inadequate schools in this dataset, there are 18 where the Achievement score was the **only** subgrade rated Inadequate, with the rest being Satisfactory (or in one case Good). This compares to just 1 school where only the Behaviour was Inadequate, and 0 schools where only the Quality of Teaching or the Leadership and Management is Inadequate. In other words, in poorly performing schools where the subgrades do not paint a consistent picture, the Achievement score is by far the dominant subgrade.
- For the 419 Requires Improvement schools, there are 24 where the Achievement subgrade is the only Satisfactory grade, with the rest different (all good or better). There are 3 where only Quality of Teaching is Satisfactory, 1 where only Leadership and Management is rated Satisfactory, and 0 where only Behaviour is Satisfactory. So for Satisfactory schools, a lone Achievement subgrade is 8 times as likely to match the overall score as the next subgrade.

At the higher ends, the effect is more muted, and Achievement, Quality of Teaching and (for Outstanding schools) Leadership and Management are much more likely to all correlate.

- For 563 Good schools, there are 31 schools where both Leadership and Management and Behaviour and Safety are Outstanding, but Achievement and Quality of Teaching are Good (and which drive the overall grade). Conversely, there are 0 schools where Achievement and Quality of Teaching are both Outstanding but Leadership and Management and Behaviour and Safety are only Good, where the school is not rated Outstanding overall.

- For the 129 Outstanding schools, the driver breaks down further because Quality of Teaching score correlates 100% of the time as a requirement of Outstanding. So in almost every instance, Achievement and Quality of Teaching – and also Leadership and Management – are all ranked Outstanding. Behaviour and Safety still exists largely as a standalone score (although it also correlates 93% of the time).

In other words, for many schools – particularly at the lower end – the data drives the Achievement subgrade, and perhaps also the Quality of Teaching subgrade, and these drive the overall grade for the school.

So the evidence is clear that data drives inspection results. But ought it to so much? Here the opinions of the call for evidence were more mixed. Many teachers responding to our call for evidence welcomed a focus on data as a way of showing what a school does in a way that was rational and defensible. This was particularly the case amongst some respondents who felt that the ‘visible’ elements of their school might be less appealing, so the data provided a welcome empirical counterpoint. However, many other teachers and heads suggested that data could be misleading. Three main concerns were raised: either

- that data could be actively manipulated
- that data does not measure important things
- that data does not sufficiently take account of school’s context.

The first of these is inevitably a sensitive area. Lord Bew’s report into Key Stage 2 assessment particularly cites complaints from junior schools that infant schools inflated results, leaving them with impossible progress targets.<sup>30</sup> While some of the respondents to our call for evidence admitted that they themselves had inflated levels, these were a small minority. However more did state that, while they had not done so, they knew other schools where this happened. This was particularly the case from secondary schools referring to their feeder primaries. Outright fraud and manipulation of any system is difficult to design policies against, as it is by definition conducted by those acting in breach of any policy. It is of course not only a question of gaming the system; assessment is a difficult task, and as the review by Professor Harlen notes:

“there is bias in teachers’ assessment (TA) relating to student characteristics, including behaviour (for young children), gender and special educational needs; overall academic achievement and verbal ability may influence judgement when assessing specific skills.”<sup>31</sup>

To this list could be added the warning that teachers are liable to under-assess students from some ethnic groups and over-assess students from others.<sup>32</sup>

Lord Bew’s report recommended that moderation be focused on schools where results are inconsistent should be moderated more frequently, a recommendation that has been adopted. However there are no formal requirements around this. Currently, it is only a requirement that schools will receive key stage moderation visits once every four years.<sup>33</sup> Effective and regular moderation, with a random sample of work assessed by external experts, is

30 Lord Bew, *Independent Review of Key Stage 2 testing, assessment and accountability*, June 2011

31 Harlen, W., *A Systematic Review of Evidence of the Impact on Students, Teachers and the Curriculum of the Process of Using Assessment by Teachers for Summative Purposes in Research Evidence*, EPPI-Centre, Social Science Research Unit, Institute of Education, 2004

32 Burgess, S. and Greaves, E., *Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities*, Centre for Market and Public Organisation, University of Bristol, 2009

33 *Guidance for key stage 1 teacher assessment moderation*, Standards and Testing Agency, 2014

one of the few ways to defend against internal assessment which is inaccurate, whether intentionally so or not.

The proposed new Ofsted system as outlined in this paper also acts as a mitigating check against this in two ways – firstly, through the option for an inspector to call in a tailored inspection if they don't trust the data, and secondly the QA process of a random sample of schools undertaking tailored inspection.

In terms of data not measuring the most important things, the new accountability measures at secondary level from 2016, will mean that schools have to report their students' average grade across their best eight GCSE subjects as well as how each pupil has progressed since primary school compared to children of a similar starting ability, and how their GCSE grades differ. There are hopes that this will reduce the pressure from league tables to focus on C/D borderline children, but perhaps more importantly in the context of Ofsted, it makes for a straightforward comparison for how children are progressing at the school they are inspecting compared to those in other schools.

At primary, the issue is more uncertain. There is currently the Early Years Foundation Stage Profile (EYFSP) at age five, at the end of reception. However, primary schools usually take children at age four, at the start of reception. This means that the EYFSP happens after a year of being at that school. In effect it is not a true baseline against which to measure further progress. This seems to be one of the major drivers for Sir Michael Wilshaw calling for the introduction of a baseline assessment of four year-olds upon entry to primary school, which this report endorses as a method of more accurately tracking progress, although it must be very carefully designed and must measure more than just literacy and numeracy.<sup>34</sup>

One respondent also raised a technical point, that this data may in fact be quite out of date 'when an inspection occurs before the validated RAISE report about the most recent exam results is published – this means that inspectors will arrive in a school having formed opinions on data which could be over 12 months out of date.'<sup>35</sup> The other issue often raised here is one about the wider elements of what makes a school successful – the extent to which it enriches pupils' moral, spiritual, social and cultural development in ways not picked up on test data. It is undoubtedly true that such elements are important to schooling. But a focus on data ought not to take away an appreciation of these wider elements. Often inspection reports now balance an assessment of pupils' progress with commentary on the wider development they receive, and there is no reason why this ought not continue in theory – though as the report discusses later, there are concerns around the quality of inspectors to assess it.

Lastly, there are concerns around the context of a school not being picked up. Consultation responses specifically raised issues of small schools, of new schools, around treatment of the pupil premium cohort, and of special schools. This will be discussed in more detail later in the report.

There is some validity in many of these concerns – particularly around the moderated quality of the data, and the quality of the inspectors who are using it and interrogating it. Later chapters will explore the ways in which these concerns can and should be addressed. Data will never tell the whole picture, and it is always open to manipulation or misinterpretation. But nor will it go away, and – if properly moderated, and used by inspectors who have the statistical capacity to understand it and the ability to take context and special circumstances into

---

<sup>34</sup> *Unseen Children – HMCI Speech*, Ofsted, June 2013

<sup>35</sup> Attendee, Policy Exchange roundtable

account – it can be one of the most powerful indicators of whether a school is doing well for its pupils.

### Headteacher discussion and school self evaluation

The evidence does suggest that headteachers are good judges of their own teachers' quality, and they are of course best placed to be informed about everything going on in their school.<sup>36</sup> It makes sense therefore that the self-evaluation by senior leadership forms a key part of the judgment on a school, and particularly feeds into assessing the Leadership and Management. However, as with so many of the current Ofsted methods, it is almost impossible to judge how accurate these self-evaluations are, and how good inspectors are at interrogating them. On the first point, self-evaluations are not made public, so it has not been possible for the report to determine if they tend to agree with the judgements given by Ofsted. Self-evaluation, and the discussions between headteacher and inspector were much discussed during the headteachers' roundtable which informed this research. During conversation it emerged that many headteachers – unsurprisingly – felt that their own self evaluation was the single most important element of an inspection, being based on professional expertise but also knowledge of the school. One head commented that “the inspector ought to be looking at the school's own quality assurance processes and...checking these judgements”.<sup>37</sup> Heads seemed comfortable with these self evaluations being challenged and scrutinised. One other element that was discussed was that the likelihood of Ofsted inspectors agreeing with their own evaluations came down to a personality issue, with more forceful heads who had more experience of Ofsted inspections able to argue their case and receive better grades. There was some concern felt for new headteachers, suggesting that they would be less able to make their case and might bow to the judgment on an inspector.

“Responses to consultation showed that many headteachers – unsurprisingly – felt that their own self evaluation was the single most important element of an inspection”

### Work Scrutiny

One of the key balancing tools which Ofsted inspectors have reported employing is that of work scrutiny; this involves looking through the books of certain students to judge their overall progress. It is worth noting at this point that schools should be moderated for their assessment at least once every four years; moderators tend to be experienced classroom teachers specialising in certain key stages which enables them to judge the level and progress that children are displaying. Whether it is possible to expect inspectors who are able to work across special, early years, primary and secondary phases is more questionable to be similarly adept at making these judgements is questionable. Accurate assessment of children's work is a real skill and one which requires knowledge of the levels and abilities of children at different stages of development. As the Ofsted analysis of the pilot program 'Assessing Pupils' Progress' identified, teachers need assessment guidelines for assessing work at different levels, and other supporting materials; it is not clear how much training or guidance on assessing work Ofsted inspectors are given, and there is no guidance in the handbook or subsidiary guidance.<sup>38</sup> Additionally, when schools are assessed early in the school year

<sup>36</sup> Jacob B and Lefgren L, *Can Principals Identify Effective Teachers?*, Journal of Labour Economics, 2008

<sup>37</sup> Attendee, Policy Exchange roundtable

<sup>38</sup> *The impact of the 'Assessing Pupils' Progress' initiative*, Ofsted, April 2011



inspectors will only have the last year's books to go on – which may be of limited relevance if there is high turnover of staff or children.

There is not much available research into book scrutiny, and that which has been done tends to rely on highly trained inspectors with subject-specific knowledge following a specific rubric to analyse assignments – for example the IQA and IDAP rubrics. This is quite different to the system of examining children's entire books to judge their progress, as it is more about judging the quality of teacher's assignments. However, the research does show that it is possible to get to relatively high levels of reliability when trained in a specific method, and that this does correlate well with value-added test scores over the year.<sup>39</sup> It seems therefore that thorough analysis of students' work, by well-trained inspectors who know the subject and phase well, may be a valuable way of judging the quality of teaching and learning.

### Student surveys

Another method that inspectors use to gain an understanding of a school is in discussion with students, although this is most frequently cited by teachers and inspectors as a way of getting a better understanding of their learning or progress. The handbook is clear that this can be done in a formal or informal way on the days of inspection, but given time constraints there will of course only be limited numbers of students chosen for interview.<sup>40</sup> Headteachers we spoke to mentioned how comments from one of the few students chosen could directly inform the wording of a report. While of course the views and experiences of every student matter, it may be an unfair representation of the general views of students in the school. For the most part, teachers we spoke to were somewhat sceptical about asking students to evaluate them – fearing perhaps that they would be overly negative, or have 'favourite' teachers for reasons unrelated to their ability.

However research has shown that when using a large sample of students, with questions designed appropriately, it is possible to get a high level of validity (compared to value added test scores from student surveys). In fact, some studies have shown that student surveys were a better predictor of student achievement than lesson observation, teacher or headteacher effectiveness ratings.<sup>41,42</sup> While there are some things these evaluations will not be able to grasp, questions about how challenged students are, and how responsive their teachers are, give an impressive source of information. Currently it seems that this is a significantly under-used resource, with no student questionnaires being systematically distributed.

There are also concerns that behaviour issues are not an important factor when it comes to the overall grade, which Ofsted has gone some way towards addressing with its announcement of no-notice inspections when there are concerns about behaviour (alongside an increased emphasis on behaviour). However, organisations such as BeatBullying suggest more could be done, and that an inspection regime that engaged more with children is needed to ascertain the true level of bullying in schools.<sup>43</sup> More engagement with students, perhaps through surveys, could perhaps go some way to addressing these issues as well.

39 Junker B et al, *Overview of the Instructional Quality Assessment*, January 2006

40 Ofsted Handbook

41 Little O et al, *A Practical Guide to Evaluating Teacher Effectiveness*, National Comprehensive Center for Teacher Quality, April 2009

42 *Ensuring Fair and Reliable Measures of Effective Teaching*, Culminating Findings from the MET Project's Three-Year Study, Bill and Melinda Gates Foundation, January 2013

43 BeatBullying response to Call for Evidence

## ParentView

Given that one of the objectives of Ofsted is to provide information to parents when choosing their school, it seems sensible to draw on existing parents' views of the school and their levels of satisfaction. However, under the current model ParentView may be problematic; it relies purely on goodwill to ensure that non-parents/carers do not fill in the questionnaire as anybody is able to register and choose a school. This is potentially opening it up to misuse. Additionally, there is no way for a school, or Ofsted, to know if the responses it receives are from a reasonable cross-section of the school community, including those who might struggle to engage with schools.<sup>44</sup> Unfortunately up to date figures on parental response figures are not available, although a Freedom of Information request from March 2013 showed that there had been fewer responses on ParentView than there were schools, but of course this was only a few months after it had been established.<sup>45</sup> However more recent small scale studies have found many schools failing to get the number of required responses.<sup>46</sup> This research did suggest, though, that concerns that parents were likely to use it as a means to complain, which some headteachers voiced as a potential problem, were perhaps unfounded as 6 out of 10 inadequate schools had too few responses.<sup>47</sup> Of course, this is far from conclusive, as there may have been other reasons those schools had low responses, and parents may still use it as a complaint mechanism.

44 Cowley A and Cowley D, *An investigation into Ofsted's approach to Parental Engagement within the inspection frameworks introduced in January and September 2012*, Engagement in Education Ltd, March 2013

45 O'Brien s, Number of Parent Responses received on ParentView, FOI request, January 2012

46 *ParentView: Getting your parents online*, Headteacher Update, May 2013

47 Ibid

---

# 4

## The Quality of the Inspectorate

48 The House of Commons report into Ofsted in session 2010-2011 quoted a figure of 400 across Ofsted (Education Select Committee, Second report, *The role and performance of Ofsted*, March 2011), and a TES article in 2012 quoted 318 (Exley S, *Are they experienced*, TES, April 2012)

49 House of Commons Hansard, 9th July 2009, Column 997W, *Letter from Christine Gilbert dated 6th July 2009*

50 List of additional inspectors completing inspections on behalf of Her Majesty's Chief Inspector of Education, Children's Services and Skills, Ofsted, October 2013

51 Sources for value of total contracts: Serco press release announcing a contract value of £55m over 6 years (*Serco selected by Ofsted to run Inspections*, Serco, 25th March 2009); *Tribal half year results for 2009* show a value of £75m over 6 years on slide 18 (*Half year results for the six months ended 30th June 2009*, Tribal Group plc, August 2009) and a CfBT broker note reports contract value of £57.5m over 6 years (*Income down 13% at CfBT*, *The Assignment Report*, January 2012). In other words, the total annual value across the three outsourced contracts is £31m a year assuming a flat profile. Last year there were almost 8,000 inspections of schools, compared to 423 FE inspections and around 200 work based learning inspections. It seems fair to assume therefore that the vast majority of the value of the outsourced contracts is for the schools phase. Ofsted's total budget in 2012/13 was £157m, which will include all programme expenditure for inspections of all areas Ofsted inspects (around £140m) as well as administration costs (paybill plus corporate services) which totalled £16m in 2012/13. Ofsted annual accounts do not break expenditure down between the different frameworks. Schools make up approximately 1/3 of all total inspections in a year (8,000 out of c 25,000) but are much more labour intensive than, for example, child minder inspections (of which there are around 10,000 a year) so it is not possible to estimate accurately what proportion of Ofsted in house spend is on schools.

As well as concerns around what Ofsted look at during a school inspection, for many of those teachers and heads that responded to the call for evidence, their concern was as much or more with the ability of inspectors to actually assess a school's performance in a high quality way.

It is first of all worth noting how Ofsted are structured for the purposes of school inspections:

- Only a very small number of inspectors are employed directly by Ofsted. These are the Her Majesty's Inspectors, or HMI. There are somewhere between 300-400 HMIs employed directly by Ofsted, of whom 141 work within the schools framework.<sup>48</sup> Clearly, as a very small number, HMI cannot carry out much routine inspection. Their role is mixed between leading higher risk section 5 inspections, overseeing section 8 monitoring visits and follow up inspections, and training and otherwise addressing national issues across schools inspection.
- The vast majority of the 6,690 maintained school inspections that took place last year were therefore carried out by Additional Inspectors (AIs) who work for outsourced contractors called Regional Inspection Service Providers (RISPs). These are currently Tribal, Serco and CfBT. There were 1567 AIs undertaking school inspections in 2009.<sup>49</sup> The latest list of Additional Inspectors on the Ofsted website has a total of 3001 additional inspectors, of whom 344 are still awaiting authorisation to inspect schools. However this list covers those completing inspections of maintained and independent schools, learning and skills, and initial teacher training, so is not a directly comparable figure.<sup>50</sup> These AIs often work as associates, or freelancers, for the RISPs themselves, alongside other jobs, rather than as full time employees of the contractors.

In 2009, Ofsted signed a six year contract with the three RISPs to carry out inspections across a third of the country each. The contract covers inspections across schools, further education and skills and work based learning. As a commercial contract, the exact terms of the deal are confidential but this report estimates the total value of the contracts for schools to be close to £30m a year.<sup>51</sup>

This structure of the inspectorate is vitally important because it relates to the quality control of inspectors (both HMIs and AIs), their training, and the accountability of their work both to Ofsted and also to schools.

The main concerns around the quality of the inspectorate fall into two categories:

- whether Ofsted and the RISPs recruit the right people with the right skill sets to become inspectors, and
- whether the training and quality assurance they receive is appropriate.

## Who are the inspectors?

Ofsted produces a document detailing the standards required of inspectors for the RISPs.

‘2.2 An inspector will always have:

- j) In the case of inspectors undertaking inspections of schools subject to section 5 of the Education Act 2005, a relevant degree and/or teaching qualification;
- k) In the case of inspectors undertaking inspections of schools subject to section 5 of the Education Act 2005, a minimum of five years’ successful teaching experience;
- l) credibility and up-to-date professional knowledge, for example of the remit, curriculum, recent developments in the sector, and statutory requirements (where appropriate) within the relevant area;
- m) competence in the use of IT; and
- n) a clear criminal records bureau check that is refreshed every three years.

2.3 An inspector will normally have:

- a) In the case of inspectors undertaking inspections other than those of schools subject to section 5 of the Education Act 2005, a relevant degree and/or equivalent professional qualification. For example, a teaching qualification, a degree, a relevant social care qualification, a relevant vocational and/or teaching qualification in further education (FE) or adult skills, and/or a leadership and management qualification;
- b) appropriate occupational/industrial experience for inspectors of FE colleges, work-based learning and adult skills;
- c) a minimum of five years’ successful experience within the relevant setting, for example teaching, training, inspecting or advising;
- d) a minimum of two years’ successful and substantial management experience in the relevant area; and
- e) a wide range of experience within the relevant area, for example in more than one institution.<sup>52</sup>

There is also a list of competencies required:

‘In order to be deemed ready for training in the new frameworks, all inspectors should already possess the basic knowledge, skills and abilities necessary to inspect effectively. They should be competent in their ability to:

- a) gather, analyse and interpret relevant evidence;
- b) make judgements that are objective, fair and based securely on evidence;
- c) communicate clearly, convincingly and succinctly, both orally and in writing;
- d) display high levels of professional conduct; and
- e) lead others and manage their work effectively to achieve high quality outcomes.<sup>53</sup>

52 Ofsted, *Qualifications, experience and standards required of additional inspectors undertaking inspections on behalf of Her Majesty’s Chief Inspector of Education, Children’s Services and Skills*, December 2013

53 Ofsted, *Qualifications, experience and standards required of additional inspectors undertaking inspections on behalf of Her Majesty’s Chief Inspector of Education, Children’s Services and Skills*, December 2013

Interestingly, there are a couple of omissions from this list. The vague reference to “gather, analyse and interpret relevant evidence” and “make judgements based on evidence” seems to underplay the absolute centrality of data and other analysis to a successful Ofsted inspection. Furthermore, the only requirement on familiarity with current practice is again a vague reference that inspectors should have “credibility and up to date knowledge on the sector”, which seems far too loose – there is no requirement, for example, that all inspectors must have taught in a maintained school within the past, say 5 years. Thirdly, there is nothing about matching by phase – no requirement that inspectors from a primary background, for example, are deployed to primary schools. In response to the calls for evidence, we heard anecdotal claims that RISPs are continually searching for additional inspectors to ensure they can meet the number of inspections they are contracted to do, which may mean that matching by phase is not possible with the limited numbers at current.

### How are they trained?

Even if the right people are recruited, it is then essential to ensure they are trained in the correct manner. Serco gives the most comprehensive outline of its inspector training which consists of five days of face to face learning, formal assessment, a mentored inspection, a final workshop as well as distance learning and completion of the Professional Qualification of School Inspectors which involves further study and sign off by an HMI.<sup>54</sup> However, in conversation with one inspector who had undergone their training with one of the RISPs, it became clear that

there are potentially some serious concerns about the quality. Firstly, little emphasis was placed in the training on ensuring that their judgments were reliable. Would be inspectors were shown videos of lessons, and asked to give their grades for teaching and learning based on the lesson (interestingly, in complete contrast to the guidance around making judgements based on snapshot lessons – let alone with the added difficulty and impact of validity of making a judgement via a video)

but no ‘right answer’ was given. Secondly, there was no training given on the assessment of students’ work which is a very skilled task and one which teachers or headteachers may have knowledge of for the particular phase or subject they teach, but not necessarily beyond that. Ofsted have acknowledged the importance of guidelines for teachers, but have no guidance on it in their own handbook.<sup>55</sup> Finally, while RAISEonline was discussed, there was no detailed explanations around significance, sample sizes or other common statistical issues. When it comes to continued training, inspectors receive an occasional update on any new Ofsted guidance, but no ongoing assessment.

Additionally responses to our consultation raised concerns around how thorough the induction and professional development process is. One headteacher commented that “A colleague did one shadow inspection, then did two secondary school inspections as part of a team, and has now been invited to be a lead inspector – having never done a primary inspection”.<sup>56</sup> Another commented that “get inspections under your belt as soon as you can and then go for Lead Inspector ... this is standard advice you get from the RISPs”.<sup>57</sup>

“The vague reference to “gather, analyse and interpret relevant evidence” and “make judgements based on evidence” seems to underplay the absolute centrality of data and other analysis to a successful Ofsted inspection”

---

<sup>54</sup> *New school inspector training overview v1.3*, Serco, July 2013

<sup>55</sup> *The impact of the ‘Assessing Pupils’ Progress’ initiative*, Ofsted, April 2011

<sup>56</sup> Attendee, Policy Exchange roundtable

There are also potential pitfalls in recruitment beyond the description of requirements – when it comes to recruiting serving headteachers, many told us that they were not inclined to do so because of the cost to their school (Serco shows a cost for its training as £3,200 + VAT), or because they did not want to take the time away from their own school.<sup>58</sup> Another common concern was that the day rates were not high enough to attract the calibre of professional required.

Since beginning the process of research for this report, Sir Michael Wilshaw has announced that all Ofsted training for AIs will be taken back in house. This is very important, as it shows that Ofsted itself has concerns about the quality of training that has been provided so far. However this raises some further questions, namely how can we be sure Ofsted will do it better, and what are the risks around the lack of transparency with RISPs?

When it comes to Ofsted's own training, it is certainly true that the majority of teachers and headteachers spoken to for this research held HMIs in much higher esteem than AIs. However, this does not mean there are no concerns about Ofsted's own training regime. The information for National Leaders of Education shows that training consists of nine days including shadowing one inspection and being on the team for another. This still does not seem like a very intensive training period (for example the EPPE study included 12 days of in-house training on lesson observation alone<sup>59</sup>). However, more important than time is what actually happens at that training, particularly around ensuring that individual inspectors will reliably implement the framework in the same way. Professor Coe, among others, has made suggestions for how to ensure that on-going reliability is independently assessed and openly reported, with inter-rater reliability scores being tested and reported. This would help address many likely concerns about the quality and consistency of lesson observations, and the judgments they feed into, and help answer questions about whether (for example) the time of day, year, characteristics of children or inspectors' own experience play a part in determining their judgments. The case study on the Health and Safety Executive's approach to ensuring consistency of judgments may also provide some useful lessons, whereby inspectors have on-going development to check the reliability of judgments they have made, and moderation sessions to do the same.

### Assessing the performance of the outsourced contractors

As shown by Ofsted's decision to take training in-house, there have clearly been concerns about the outsourced training as an element of the contract. This could be shown as performance management in action, with a faulty training system being rectified. However it also reveals how little is known about the performance of outsourced contractor(s) in this important public service. For example, things that are not known include:

- the amount which contractors pay to their AIs for conducting Ofsted inspections
- the number of AIs who have applied to each RISP, their qualifications, and their experience, and how many of these AIs were accepted onto the programme
- the way in which AIs are recruited, (mostly) trained, and performance managed by the contractors

<sup>57</sup> Attendee, Policy Exchange roundtable

<sup>58</sup> *New school inspector training overview v1.3*, Serco, July 2013

<sup>59</sup> Sammons P et al, *EPPE – Variations in Teacher and Pupil Behaviours in Year 5 Classes*, DfES, 2006

- the extent to which the RISPs monitor the performance of their AIs and the judgements they form, and whether new inspections are allocated out to AIs on this basis
- the extent to which Ofsted holds the RISPs accountable through contract performance indicators, including what these are
- what the performance of the RISPs is against these performance indicators and whether there is variance between the RISPs
- whether there has been any financial penalty to RISPs because of any element of poor practice
- whether any inspectors who have been disciplined or otherwise had action taken against them for poor quality practice in inspections
- whether the training element was one stated requirement of the contract and whether Ofsted had to exercise a break clause in order to bring back the training in house, and what the costs of this was to Ofsted, if any

In fact, the sum total that is known about the contracts Ofsted holds with the RISPs is the total value of the contract to each organisation (which is largely a requirement of listed companies more than from Ofsted), the number of inspectors employed by the RISPs and their names, the central lists of requirements placed by Ofsted as to the qualities of AIs, and some exceptionally bland ‘pen portraits’ of selected AIs, which are often written in the first person (hence presumably self authored), not standardised, and do not contain information on previous inspections, their level of inspection training, or anything on their reliability or validity of previous judgements. Moreover, as private companies, and because of a commercial contract between them and (FOI eligible) Ofsted, none of this information is subject to FOI requests.

This is profoundly unsatisfactory from an accountability point of view. And it is made more so because of the specific “twice removed” system of accountability with regards Additional Inspectors. If, say, a prison is outsourced, then the staff of that prison will not work for the public contracting body but instead for the

contractor – and will transfer under TUPE if existing staff or be employed on a new contract by the contractor. The staff are therefore ‘one step removed’ from the ultimate payer – the government. However, the contractor has a direct employment responsibility to these

staff, and all the standard organisational tools to ensure the workforce is well developed and knowledgeable in the task.

Ofsted inspectors, by contrast, are “twice removed”, as they will not work full time for the contractor, but instead will typically work on a contracted basis, undergoing specific tasks on a day rate basis. This has the advantage for the contractor of flexibility of workforce and minimal costs. However, it does mean that their employment responsibilities and oversight functions are diminished. There will be, for example, no objectives set, and limited performance management or professional development. And, crucially, the organisational loyalty will be weaker – particularly if the AI is also undergoing various other roles for the contractor or another contractor in a different organisation. A self employed individual who works (for example) 10 days a month for a contractor doing Ofsted inspections, as

“What is known about the RISPs is profoundly unsatisfactory from an accountability point of view”



well as undergoing various other jobs, will likely feel only a partial loyalty to that contractor. His or her loyalty to the ultimate payer – Ofsted – is likely to be even weaker, as he or she is “twice removed”. This – crucially – is why we conclude that all the laudable efforts by Sir Michael Wilshaw to impose consistency on the workforce have been so mixed. When the dissemination of a message is required to make several pitstops – first permeating the organisation within which is made before being passed to a contractor and then again to the contractor’s loose workforce – it is no surprise these Chinese whispers result in diluted forms of the original. This is perhaps particularly the case when the normal organisational tools that would be deployed to make such a change a reality (eg organisational wide training, changing of objectives and so on) are also not available.

To be clear, this report is not saying that outsourcing of public services cannot work. Indeed, there have been many successful examples – including in education – where outside contractors have delivered improved services at lower cost. The key conclusion is that such an outsourced process must always be tailored to the specific circumstances of the market.<sup>60</sup> In this instance, the conclusion is that the current outsourced model does not provide such assurance.

The reason for worrying about the selection, training and transparency on quality of these inspectors is because there have been serious concerns voiced over their ability to make valid and reliable judgments. As one Headteacher put to us “the fundamental problem is variance of inspectors, and variance of schools to manage them”<sup>61</sup>

These concerns break down into different areas:

- Ability to understand and interrogate data, especially for schools that do not for whatever reason fit a standard pattern
- Background and preferred style of inspectors, including particularly around lesson observations and judgements on teaching

### Ability to understand data

One of the most significant concerns raised by headteachers and schools in responses to the call for evidence was that inspectors simply did not understand their data, in particular progress measures. This is deeply concerning, given the reliance placed by inspectors on data. In particular, we heard concerns from specific types of schools who felt that they were particularly at a disadvantage when it came to their data being assessed:

- Schools which are rapidly changing in performance, either for the better or for the worse. In this instance, looking at historic data will likely not tell the full picture as to the school’s current trajectory and capability to improve.
- Small schools, where the cohort size is smaller. The report discussed earlier the weakness of data analysis in this regard. The recently published Ofsted guidance also hints that there may have previously been difficulties in interpreting data, pointing out to inspectors the difficulties of small sample sizes. It explains how, for example, if you are trying to judge the progress of high achieving children on free school meals, and there are only four at the school, your only percentages can be 0%, 25%, 50%, 75% or 100%, so

<sup>60</sup> For an exhaustive discussion of the benefits of a mixed delivery model for public services, and the need to tailor specific solutions to the market in question, see Worth S, *Better Public Services: A roadmap for revolution*, Policy Exchange, 2013

<sup>61</sup> Attendee, Policy Exchange roundtable



comparing to a national average of say 64% making expected progress can be difficult.<sup>62</sup> This suggests that this was not fully understood by inspectors before. One of the respondents to the call for evidence stated “It is very apparent that not all Ofsted inspectors understand the statistical unreliability of outcomes from small groups. One recent inspection (I was interviewed as part of the Governor team) 5 boys (year group >200), who would have been excluded from many schools, swayed the inspectors judgement! Inspector refused to look at detailed case files for the pupils concerned.”<sup>63</sup>

- New schools – particularly free schools – which will not have three years’ worth of previous data with which to share with inspectors. The handbook is clear that in such instances inspectors should look carefully at all internal data provided by the school. But two free schools raised in the Call for Evidence the issues of their internal data not being seen as robust as external data, with one in particular being told by Ofsted that they were likely to get a better inspection result once the first cohort had taken their GCSEs.
- Schools who track pupil performance data in ways other than through National Curriculum levels. In time, all schools will be designing bespoke systems for tracking pupils and assessing their progress. But we heard evidence that at present time, inspectors can show a preference for levels and systems they understand, and do not always appreciate the ways in which different systems work.
- Non mainstream schools – in particular, special schools. A special school headteacher raised a particular concern around the way in which his pupils’ data can be measured in RaiseOnline: “there is a real issue with tracking pupils making expected levels of progress and thresholds in special schools because so few of ours do given their starting points. There’s also an issue with comparative data – we look around and we either find other schools but where the data is three years out of date, or we can’t find a comparator because our kids have multiple needs. RaiseOnline doesn’t work for us – you either need to buy a different commercial package or design your own moderated system to show partial levels of progress”.<sup>64</sup>

Taken together – and accounting for the potential for this only being one side of the story in any specific instance – this adds up to a catalogue of schools that fall into one or more of these categories. The conclusion that can be drawn is that inspectors are very comfortable with stable schools with years worth of historic data neatly tabulated into National Curriculum progress levels. But any data that is different to that has the potential to cause issues. And the variance of inspectors in understanding and interrogating that data is unacceptable.

### Background and preferred style of inspectors

There has been much talk by some that Ofsted inspectors represent part of what is sometimes termed ‘the Blob’. This rather unflattering term describes a loose coalition of educationalists who are hostile to much of the current direction of travel of government policy and who by accident or design seek to undermine it. In that case, runs the thesis, Ofsted inspectors are almost by definition incapable of making judgements that fairly represent school progress in any way that runs counter to their preferences. The only solution is some form of standalone inspectorate for innovative schools or abolition of the inspectorate.

A more nuanced and credible version of this thesis seeks to demonstrate how there is a consistent pattern in the preferred style of inspectors that comes from

---

62 *Subsidiary guidance supporting the inspection of maintained schools and academies*, Ofsted, December 2013

63 Carole Thomson, respondent to Policy Exchange Call for Evidence

64 Attendee, Policy Exchange roundtable

a shared background, and which manifests itself often subconsciously in the judgements they make. This occurs across both inspection reports for individual schools, and in cross cutting areas such as best practice guides. The former has been catalogued by blogger Andrew Old, who consistently identifies wording in Ofsted inspection reports which indicate a preference for a particular teaching style and which typically downgrade a school due to its absence.<sup>65</sup> The latter has been identified by Daisy Christodoulou in her book “7 myths about education” where she summarises 228 examples from subject reports between June 2010 and May 2012, and shows that “of the lessons that are praised, very few involve the teacher teaching facts. In the lessons that are criticised, very often the feature being criticised is the teacher talking too much, or imparting facts or teaching activities that involve factual recall”.<sup>66</sup>

Many headteachers spoke of the variability of inspection teams, and how this made them feel the need to play to the lowest common denominator, particularly when it came to lesson observations. Most we spoke to were positive about the intention behind Michael Wilshaw’s emphasis on ‘no preferred teaching style’ but were unsure if this had in fact trickled down through to the inspectors themselves. A worrying number spoke of the need to have plenaries, display progress in twenty minutes, do lots of activities and so on. One teacher discussed how their very ‘traditional’ teacher-led style of lesson had been graded badly by an inspector who wanted to see more independent learning; however, another described how their three-hour lesson with a great deal of group work was graded badly as the inspector did not understand how to judge the quality of teaching. This suggests that there is a certain lack of skill and flexibility among inspectors to adapt their judgements to different styles of teaching.

Again, there should be a distinction drawn here between Ofsted expecting to see a certain type of lesson, and pre-emptive judgements made by schools in the belief that Ofsted will, and so making changes on that basis. The latter cannot be blamed on Ofsted – at least not entirely. For instance, teachers in our roundtable drew a distinction between weaker Senior Leadership Teams they had served under, who were more likely to try and prepare exhaustively through coaching and checking for ‘Ofsted Outstanding lessons’, and stronger ones who protected staff and encouraged them to teach in their preferred way. But the variability of inspectors makes this in some way a rational decision, particularly if the school is on the cusp of a judgement below Good. As a headteacher commented, “you don’t know who’s going to walk through the door so you have to be ready for anything and prepare for the lowest common denominator.”<sup>67</sup>

## An inspector in every school?

Ofsted’s answer to much of this has been an increased drive to get more serving headteachers or senior leadership to undertake inspector training, with 40% of inspection teams now containing a serving head or senior school leader. This was supported as an idea by feedback from our headteachers roundtable, with many headteachers saying they felt it was essential for teams to include a serving head. However, when asked, only two in the roundtable were actually trained inspectors themselves! The reasons given for doing so, furthermore, were in a sense negative – both heads felt that they had bad experiences of Ofsted and therefore owed it to their colleagues to sign up and try and do better. Reasons given for not becoming an inspector included the upfront cost to the school of undergoing training; that

65 Andrew Old, *Teaching Battleground* <http://teachingbattleground.wordpress.com/>

66 Christodoulou, *7 Myths about Education*, The Curriculum Centre June 2013

67 Respondent, Policy Exchange Call for Evidence

heads didn't have time (especially from the LLEs/NLEs), and that some didn't want to be part of Ofsted.

So there remains some way to encourage serving teachers to form even the majority of inspections. Furthermore, this would solve at best half of the problem. It should mean that inspectors are indeed fully up to date with current pedagogies, curriculum etc, and more comfortable with different assessment models. It may also mean that they are more comfortable with the large quantities of data in current schools. But there is no guarantee that these serving teachers will have any different opinions as to preferred style, or any greater loyalty to Ofsted (as opposed to loyalty to their school). So again, whilst the fundamental structure of Ofsted, with external observation of lessons and inspectors making judgements as opposed to validating them, still exists, such a solution can only go part way. More fundamental reform is needed.

### **Case Study: The Health and Safety Executive**

It is useful to think about other regulatory bodies, and how they go about addressing some of the concerns raised. Although of course undertaking a very different kind of work, it is possible to draw some useful contrasts with the Health and Safety Executive.

#### **Mythbusting**

The Health and Safety Executive has faced significant challenges with the perception of what constitutes 'Health and Safety' and what constitutes sensible precautions. They describe how many people and businesses use health and safety as an excuse for not doing something, often something that they didn't want to do for other reasons. Additionally, certain people were setting themselves up as health and safety 'consultants' – suggesting work that wasn't really necessary, or went too far in terms of what the law required. There are obviously comparisons we can draw with senior leadership and consultants in the school sector which will be explored further in the next section. HSE's approach has been to draw up a register of recognised health and safety consultants, but additionally has set up a 'Myth Busters Challenge Panel' allowing anyone who has been advised – by employers, insurance agents or others – to do something potentially disproportionate or wrong in the name of health and safety to bring this before the HSE Chair and a panel of independent members and have the advice scrutinised. Publication of the panel decisions provides a reassuring message to people about proportionate precautions.

#### **Consistency**

One of the major concerns cited by teachers and headteachers we spoke to was the lack of consistency between Ofsted teams – many were unsure that two individual inspectors (sometimes even inspectors from the same team) would come to the same conclusion, particularly on lesson observations. A lack of reliability in any measurement invalidates the results, so this is a serious concern. This is also an issue HSE has had to confront. There are important differences though – all HSE inspectors are directly employed, with no out-sourcing, meaning that their training is all in house and consistency can be more easily checked. For example, the records of any inspector – the time spent on an inspection, the judgements reached on each one – can all be accessed by management. In order to ensure consistency, peer review meetings are held where teams of inspectors discuss the decisions made and are asked to come to conclusions about certain cases to cross check their reliability.

---

# 5

## Effect of Ofsted

This chapter will explore the way in which Ofsted affects schools; firstly by exploring the effect that the expectation of Ofsted inspection has on teachers' and headteachers' practice, and then moving on to examine the effect that different judgments have on a school's improvement.

### Before an inspection

It is possible to break down the pre-inspection effect into two types, short-term and long-term. The short-term impact is on school's response when 'the call' from Ofsted comes, the long-term impact concerns planning, training and decision making in the months or years before Ofsted arrives.

#### Short-term effect

When it comes to short-term impacts, the majority of the responses from the call for evidence highlighted that schools changed their practice significantly when Ofsted inspectors came to school, usually by teaching a lesson in a significantly different way to what they would describe as their normal practice. Out of the 262 classroom practitioners who responded to this question (and excluding Heads, chairs of governors, Academy chains and other respondents) 65 said they would not change their practice, while 189 said they would (the rest did not respond to this question). For those who did change practice, this was a fairly typical response:

*'Every time Ofsted come to school you need to change your practice to fit in with the new Ofsted criteria. For example you have to show progress in your lesson. In reality pupils may make progress understanding a complex new idea over a few lessons or over a unit of work.'*<sup>68</sup>

Changes to teaching practice during the time of an inspection is problematic in terms of Ofsted being able to make valid or accurate judgements (as discussed in Chapter 1). It is also harmful to pupils if, teachers amend their practice by seeking 'pace' or otherwise focussing on poor proxies for learning<sup>69</sup> – although the disruption caused by one observation (or maybe two) should not be overstated. Preparation for such lessons is often time consuming- not least because it is often a change from normal style – and is a complete waste of this time, which actively prevents more useful preparation being done. Lastly, feedback on such artificial practice is unlikely to be valuable to teachers in terms of improving teaching in the longer term.

<sup>68</sup> Kirstie Sanders, respondent to Policy Exchange Call for Evidence

<sup>69</sup> David Didau has written about the work of cognitive psychologist Robert Bjork on the need for learning to be sustained by having materials taught more slowly, and with suitable spaces, over a sustained period of time. Seeking to achieve rapid progress, or pace, will likely lead to shallower learning ultimately. Prof Rob Coe has also identified a series of visible but poor proxies for learning which hamper making judgements on overall learning, – for example, by looking at whether children are working quietly, whether everyone seems engaged (as opposed to passive), or whether large chunks of material in the curriculum are being covered. Such indicators do not necessarily show learning is being achieved or sustained but are often prioritised during observations, at a cost to children's more effective learning – much of which is covered in his inaugural lecture Coe *"Improving Education, a triumph of hope over experience"* CEM 18 June 2013

### Long-term effects

However it is the long-term effects of preparing for Ofsted that are of more concern. If schools are sufficiently driven in their own decision making by Ofsted views (or belief in Ofsted views), then that is potentially problematic if this means decisions are being made that run counter to decisions the school would have otherwise made in the interests of pupils. Conversely, such strong accountability has the potential to be positive if it improves schools' practice. The call for evidence responses were almost unanimous – only 7 of 305 respondents felt the pressure from Ofsted was positive. However this may well be to do with the self-selecting nature of a survey of this kind.

In terms of specific concerns raised, respondents commented that the work burden was significant, which was particularly noted when it came to data – with Ofsted expecting to see frequent assessment of children's progress recorded in a rigorous fashion. Many teachers felt that they knew how their children were

faring, and that collecting the data in this way was just box-ticking for Ofsted, and took time away from planning and marking. It is important to be clear here about a distinction between schools and teachers who are sceptical of 'any' monitoring, and those who feel there is a need for dual systems; one for

their internal use and one for Ofsted. This report argues that the former of these is crucial and should not be considered a burden (albeit it can be time consuming). The second is more problematic, and reflects a wider concern already raised about Ofsted's desire to see data presented in one format (eg through using National Curriculum levels).

Teachers also commented frequently on the preparatory activities which were required for Ofsted even when not immediately expecting an inspection. These ranged from frequent lesson observations, 'Mocksteds' and other consultancy approaches, to work planning, whole school initiatives such as a push on what could be seen to be the latest fad in policy, and scrutiny on marking. Again, it is worth repeating that some of this focus is not necessarily a bad thing. The distinction needs to be drawn between useful activities (defined here as likely to improve outcomes for young people), and nugatory, duplicative, or badly organised activities which take time away from opportunities to do more effective planning.

It is also worth noting the clear distinction which both teachers and heads freely drew between what might be termed effective and ineffective schools in this regard. Teachers and Heads were unanimous in their belief that higher performing Senior Leadership teams did not obsess about Ofsted, ensured that staff didn't feel the need to change practices or conduct time wasting activity, and emphasised that the Ofsted grade was secondary to the purpose of the school for ensuring young people achieve. Respondents to us commented, for example, that "Ofsted doesn't judge teachers, it's Heads who don't know how to manage Ofsted that judge teachers... it's all my Head talks about".<sup>70</sup> One other commented that "we came out of our Ofsted with Good with Outstanding features but the whole school was miserable, depressed... because the language we had been pushed to us by SLT was that nothing less than Outstanding would do".<sup>71</sup> By contrast, teachers and Heads discussed how effective schools ensured that teachers were not distracted or placed under undue pressure before or during an inspection,

“ Respondents commented that the work burden was significant, which was particularly noted when it came to data ”

<sup>70</sup> Respondent to Policy Exchange Call for Evidence

<sup>71</sup> Attendee, Policy Exchange roundtable

and were told explicitly not to change their style or methods. However, the clear consensus from the call for evidence – 262 responses – was that in all types of school, teachers felt under constant pressure to review and consider their practice for the sake of Ofsted.

The other significant impact of Ofsted on school planning, before an inspection, is on the large strategic decisions made by schools. Schools, particularly academies, have been given impressive freedoms but in many cases there is still a limited use of these freedoms<sup>72</sup> – indeed, Ofsted are about to commence a small revision to their inspection of Academies to ascertain how well they are using their freedoms. Many reasons have been suggested for this, but it is very likely that Ofsted itself may be one of the limiting factors. For example, the freedom to move away from using national curriculum levels is welcome, but the Ofsted framework still includes reference to expected levels of progress. We also heard evidence during the research phase of this report around schools delaying organisational changes. For example, a school deciding to move to a 3 year Key Stage 4 would need to make a variety of changes to the school: to timetabling, to staffing mix, to curriculum and schemes of work, and potentially even to the physical structure of the school. The organisational change literature is clear that many changes – even ones which are ultimately effective – can have teething difficulties and often performance takes a short term dip during the change process. Schools which are sceptical about Ofsted’s ability to recognise this change and interpret school’s performance during this time are likely to delay making such changes until after an inspection has happened. With a school wide change process potentially taking a number of years, this means it is possible that many schools to be delaying often beneficial strategic and pedagogical changes because of Ofsted.

Given the huge power of Ofsted (judging a school Inadequate is very likely to lead to a headteacher leaving, while an outstanding grade opens up a range of further opportunities such as Teaching School status) it is perhaps unsurprising that so much energy is spent interpreting what they want and how to achieve it. In some instances, this can be effective. However, so much of this is based on flawed assessments, second hand information, Chinese whispers, and expensive consultancy efforts into “what Ofsted want”. Secondly, when planning and preparing can be seen as unhelpful, and when it affects good schools as much as bad ones – in the way teachers teach, or Heads and SLT plan – and when much of this judgement is based on shaky foundations – then the overall impact here is disproportionate and unjustified. If the end goal is a school system with individuals using research to inform best practice and trying innovative techniques to test their effectiveness, this energy poured into the divination about what Ofsted wants seems like a great deal of wasted professional energy, time and effort.

### After the inspection

It is also important to determine what impact Ofsted has after it arrives, and ask whether it does indeed drive up standards. It is very hard to unpick this question, as it is difficult to isolate out the impact of Ofsted when it comes to improvement. Demographic changes, school improvement strategies, teacher training and recruitment will all have significant effects on a school’s trajectory as well.

72 Bassett D et al, *Plan A+:  
Unleashing the potential of  
academies*, Reform, March 2012

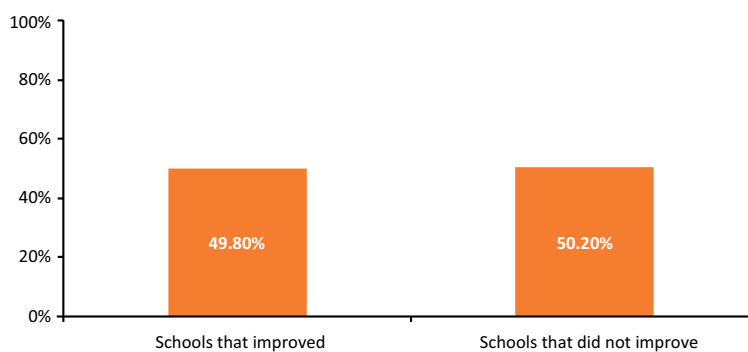
“This energy poured into the divination about what Ofsted wants seems like a great deal of wasted professional energy, time and effort”

In conversation with teachers and headteachers there were mixed opinions about whether Ofsted improved schools that were graded RI or Inadequate. There were some very positive responses, saying that when HMIs became involved clear improvement strategies were laid out and schools began to turn around. Some felt that there was a change, but this was only because senior leadership resigned or were asked to leave, and better management was put in place. However some felt it began a spiral of decline, leading to teachers and students leaving, morale collapsing and standards falling. It therefore seems necessary to explore in more detail what happens after an Ofsted inspection has occurred.

One way of looking at the impact of Ofsted is to simply judge look at the improvement of schools placed in various categories. For example, the Ofsted annual report 2011–12 shows that 48% of schools judged satisfactory previously had improved to good or better, although of course this leaves over half that had not. They also show that 90% of the schools previously judged inadequate that were re-inspected in 2011–12 showed improvement.<sup>73</sup>

However it is perhaps more useful to look at longer term trends. Starting with schools graded less than good since 2005, the education consultancy LKMco have analysed the number that have ‘turned around’, namely are now graded good or better. This shows that of those schools which have received a grade of lower than 2 since 2005, 49.8% of schools have improved, while 50.2% have not.

**Figure 4: Percentage of schools judged Satisfactory or Inadequate since 2005, that have since become Good or better**



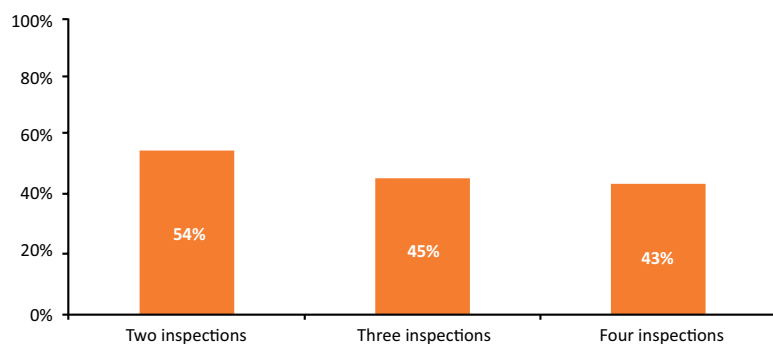
Source: LKMco analysis

On this analysis, there are 232 schools which are still open that were graded 3 in their first inspection since 2005 and have not improved since (in fact, 39 now have a grade 4 at their latest inspection). In addition there are 67 open schools that were graded 4 in their first inspection which have not achieved good or better status, with 12 still being graded 4.

73 Ofsted Annual Report 2011–12

LKMco also looked at the improvement trajectory for schools that have been inspected multiple times (particularly those who were judged inadequate). This shows that, if anything, that multiple inspections decreased the likelihood of a school improving.

**Figure 5: Percentage of schools judged Satisfactory or Inadequate who received multiple inspections and are now Good or better**



Source: LKMco analysis

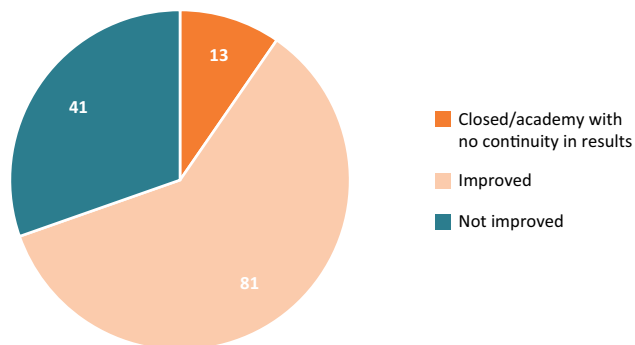
There are 59 open schools that have had four Ofsted inspections since 2005, yet are still in a category 3 or 4.<sup>74</sup> Of course, these schools that are frequently re-inspected are likely to be those that are deemed higher risk or are in a more problematic initial starting point, but this still shows that there are some schools which seem stubbornly resistant to any Ofsted improvement efforts.

It is also interesting to look at the impact of an Ofsted inspection on pupil attainment. This can be done by comparing overall attainment scores before and after an Ofsted inspection. Policy Exchange conducted new analysis for this report, which took a sample of 135 secondary schools graded 3 in 2009 (which allows for an assessment of progress since the inspection over a longer time period) and looked at their performance in 2012. 13 of these schools have either closed or been taken over by an Academy and show no continuity in their published results. Of the remaining 122, this report found 81 have improved the proportion of students achieving 5A\*–C at GCSE, relative to the national average (this specifically looks at a proportion of the average, as opposed to raw score increase, to counteract the national increases of GCSE grades over the period). On average, these schools go from having 60% of the national average of students achieving the threshold for the three years prior to inspection, to having 65% doing so.

<sup>74</sup> Data and analysis provided by Loic Menzies, LKMco

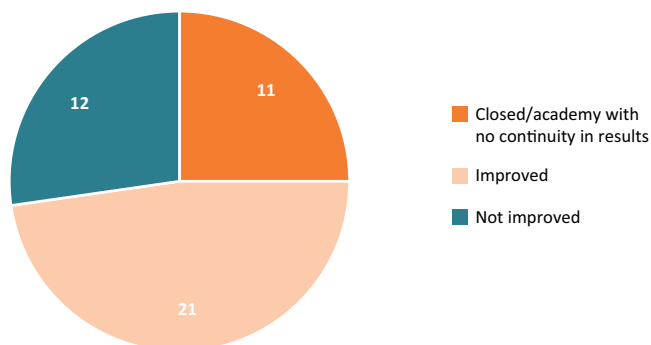


**Figure 6: Impact in 2012 of schools graded Satisfactory in 2009**



For schools graded Inadequate in 2009, 11 of the 44 either closed or were taken over by an Academy with no continuity in results. Of the remaining, 21 improved their results, while 12 did not. On average they went from schools achieving 56% of the national average of 5A\*–C to 62% of the national average.

**Figure 7: Impact in 2012 of schools graded Inadequate in 2009**



While this is of course vulnerable to changes in pupil intake ability, and simple reversion to the mean, it is still promising that things do show improvement.

A variant of this approach is to look at how schools that ‘just passed’ an inspection compare to those that ‘just failed’. Research has shown that schools that are very similar, but those that just fall into the 4 category do improve:

*‘Our results suggest that schools only just failing do see an improvement in scores over the following two to three years. The effect size is moderate to large at around 10% of a pupil-level standard deviation in test scores. We also show that this improvement occurs in core compulsory subjects, suggesting that this is not all the result of course entry gaming on the part of schools.’<sup>75</sup>*

It is also interesting to note that they found that enrolment does decline on average by about 5 students (relative to a mean cohort of around 180) in the first three years, but this estimate is not significant. They also state that while they

75 Rebecca Allen and Simon Burgess; *How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England*, March 2012

---

cannot accurately assess the timings of headteachers departures their inspection of the School Workforce Census suggests that:

*'Schools that fail their Ofsted inspection are indeed more likely to see a change of headteacher within two years than those who pass their inspection.'*<sup>76</sup>

So this analysis suggests that, at the lower end of the spectrum, Ofsted does have an important role to play as a driver of change. Given that the Allen and Burgess research includes schools that were converted to Academies, and finds that failing schools are more likely to have a change of headship, it is important to remember that Ofsted may act as a trigger of other improvement mechanisms, rather than working in and of itself. However, the idea that poorly performing schools are driven into a cycle of decline does not seem to be supported across all schools – some certainly don't improve, some do end up closing, but on average improvement is seen. A strong system of accountability can drive change in poorly performing schools.

However, much more could be done for those struggling schools that we have shown are resistant to improvement and, as discussed earlier in this chapter, there are real risks for schools that are currently doing well. In the 78% of schools that are rated good or better, there are still teachers and headteachers who feel they are 'working for Ofsted', who are expended time and professional development on understanding the Ofsted framework rather than, for example, the latest in teaching research. It is here that we need to relieve some of the intense pressure of Ofsted in order to allow teachers and heads to work, and feel that they are working, for the good of their students, and free up Ofsted to focus more on those struggling schools that are failing to improve.

76 Ibid

---

# 6

## Recommendations

So far the following conclusions can be drawn about the current model of school inspections:

1. Although a risk based and proportionate system, it is expensive – around £30m a year just on outsourced contracts for Additional Inspectors, plus a (large but unknown) proportion of the remaining £127m of Ofsted's total budget;
2. Much of that money is spent on undertaking lesson observations, which are simultaneously unreliable and invalid; harmful to schools through the behaviours they drive; and largely of little use since as far as can be understood, the data which drives the Achievement grade is also significant in driving the Quality of Teaching grade;
3. Other methods of evidence – including book scrutiny, pupil and parent conversations, challenge of head and governors – are generally thought to be more productive, but there is limited external evidence to prove this;
4. Inspections are broadly successful at identifying and turning around underperforming schools – particular the more targeted HMI support – and indeed, often welcomed as a catalyst for change. Conversely, they are often seen as unhelpful or counterproductive by schools around the middle of the performance spectrum, or higher performing schools. (Outstanding schools, of course, are exempt except on a risk assessed basis);
5. Much of the weakness stems from poor quality control amongst the inspectors themselves, which include large variance in recent school experience, levels of training, and the extent to which personal preferences and behaviours as to models of school improvement affect what ought to be impartial judgements;
6. By far the majority of the call for evidence responses and roundtable discussions were largely negative about Ofsted: the process, the consequence, or both. Some other survey evidence<sup>77</sup> also suggests widespread disillusionment with the inspection regime and the 'discouragement' caused by Ofsted. Neither of these should necessarily be taken as a representative cross section of schools or heads, however.

Conversely, the clear conclusion from the academic literature, the call for evidence, and the roundtable discussions with heads, teachers, and with other regulators, suggest some clear principles for what an effective school inspection regime should look like:

77 Garner R, Ofsted 'intimidation' risks exodus of headteachers, survey warns, Independent, May 2012, reporting NAHT survey findings

- (Even more) proportionate;
- Produces judgements which are valid and reliable, including through both methods used and quality of inspectorate undertaking them;
- Have a transparent process for third party scrutiny of these, including through appeals;
- Encourages and supports a system in which schools are the primary judges of their own progress, and primary drivers of their own improvement – albeit with Ofsted playing an external check and validation;
- Encourages – or at a minimum does not discourage – innovation amongst schools;
- Doesn't drive perverse behaviours, or other unhelpful responses, which include a disproportionate focus of school time in advance.

Although there is no empirical evidence on what the 'right' sum is to spend on external inspection, and it was not mentioned by many respondents, we would also add in the current climate "ought to cost less".

In other words, the conclusion of this report is that changes to the current model – clarifying that individual lessons should not be graded, bringing Additional Inspector training in house – are necessary but by no means sufficient. Achieving a world class school inspectorate that supports the next stage of school improvement will require a fundamental redesign of how Ofsted conducts school inspections.

## A new, two stage Ofsted process for school inspections in England

It is worth remembering that Ofsted already carries out some elements of a risk based or proportionate process – exempting Outstanding schools on a routine basis, inspecting Good schools every 3–5 years on a risk basis, and having a more tailored follow up inspection regime for Requires Improvement schools (and again for Inadequate schools). Nevertheless, there is an attraction in the more formal splitting of inspections for those schools deemed in the English scale of Good or better, and those Requiring Improvement.

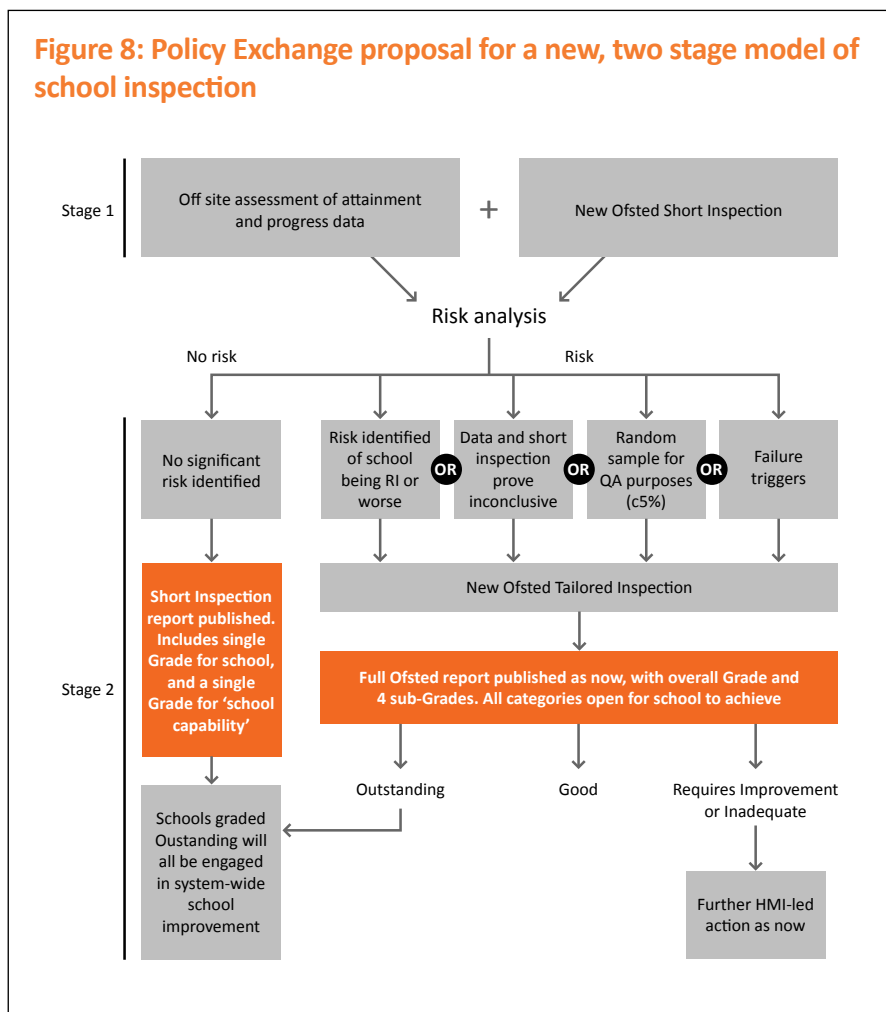
During the latter stages of research for this project, Ofsted announced that they too were considering a more tailored inspection system which had some of these features. In particular, they indicated a willingness to consider "Rather than focusing on full inspections, it would make sense to conduct more frequent, shorter monitoring inspections for good schools"<sup>78</sup>

Unsurprisingly, given this is the same conclusion that this report was reaching, this is a sensible approach. Ofsted have said that more details will follow in due course. This chapter sets out the conclusions for how Ofsted should move to a new, two stage, model of school inspections.

“Achieving a world class school inspectorate that supports the next stage of school improvement will require a fundamental redesign of how Ofsted conducts school inspections”

78 Article by Mike Cladingbowl, Guardian 7 March 2014 [www.theguardian.com/teacher-network/teacher-blog/2014/mar/07/ofsted-new-blueprint-inspect-good-schools](http://www.theguardian.com/teacher-network/teacher-blog/2014/mar/07/ofsted-new-blueprint-inspect-good-schools)

**Figure 8: Policy Exchange proposal for a new, two stage model of school inspection**



This model would operate as follows.

Every school – regardless of previous Grade – would be inspected at least every two years. This inspection would be renamed the ‘Short Inspection’ and would consist of two elements:

- There would be an off site assessment of the attainment and progress data of pupils in a school. This data would play a major role in determining the overall grade of a school – making clear and transparent what in effect happens now. Inspectors would be expected to be better trained to interrogate and analyse statistical data (discussed more below).
- Alongside that, one inspector would visit a school for a single day (for primary, secondary and special schools). **That inspector’s role would be strictly to validate the head and governing body’s own assessment of their school through its own process of self evaluation.** This would include, for example, testing their understanding and presentation of the school’s data on achievement and progress to see if it was plausible, discussing (as now) the school’s actions leadership and management, and walking around the school at key intervals to observe behaviour and safety.

- Importantly, the inspector would judge quality of teaching only by scrutinising and challenging the head's own assessment of their staff. **In a Short Inspection, there would be no routine lesson observations at all of teaching staff.** Instead, the inspector would expect to see – and challenge – the head's judgement made over the course of the year. This could include looking at copies of staff objectives and appraisals; decisions made on performance related pay and promotions; any internal notes of formative observations or programmes of lesson study done within the school; book scrutiny of pupils' work over time; discussions with staff or pupils; or other ways in which the head and SLT had decided to assess teaching quality.
- Inspectors would still be able to drop into lessons and watch teachers to help validate any of their judgements, and to get a feel of the school. But as well as not giving grades for lessons – as clarified by Ofsted recently- under this model they **would not complete any formal evidence forms as part of these brief observations.** This would remove the nuance and confusion that still exists in the system, and make clear to inspectors that their role is as a validator, not an assessor themselves.

On the basis of the data provided by the school as to pupils' achievement and progress, and this one day 'Short Inspection', the inspector would make a judgement as to whether the school presented a 'risk'.

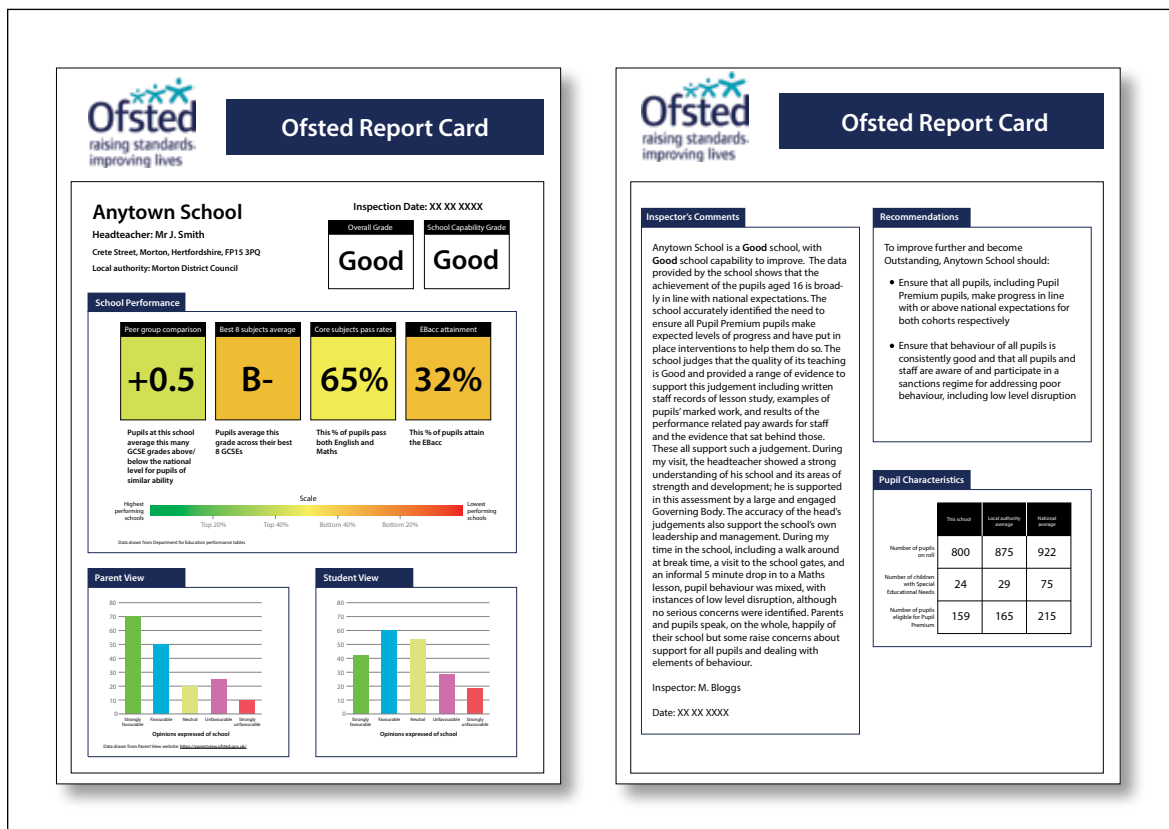
- For the majority of schools – as a rough rule of thumb, those currently graded Good or better – we would expect no risk to be identified.
- For those schools, that would be the end of the Ofsted process. A report would be published, as now, with an emphasis on visual representation of progress and achievement made by pupils, and a brief accompanying commentary from the inspector.
- In terms of grades, an overall Grade would be given as now. In addition, the four subgrades currently made would be merged into one overall assessment of 'School Capability', which would also be graded as Outstanding through to Inadequate. Importantly, a school would only qualify as no risk if both its overall Grade and its Capability was judged as Good or Outstanding<sup>79</sup> In other words, most schools would receive a combined grade of either Outstanding/Outstanding, Outstanding/Good, Good/Outstanding, or Good/Good.

Ofsted would publish a short report for these schools. We recommend that this draws heavily on the new Data Dashboard figures<sup>80</sup> that are produced for all schools and which provide an easily understandable visual representation of that school's performance. This data would be accompanied by a short written commentary by the inspector which supplemented the information – for example, that commented on the small numbers of pupils in a cohort which meant the data should be treated with care, or which particularly praised high quality leadership which meant the School Capability was higher than might be seen from the data, or so on. Schools would be assessed on a really straightforward basis – how well are the pupils doing in the school (the overall Grade) and what is the capability of the school to continue and to improve that (the School Capability Grade).

<sup>79</sup> Because of the merging of the four sub categories, it would be possible for an inspector to judge, for example, that one or maybe even two of the areas Required Improvement, but that taken as a whole, the school capability was Good. Inspectors would make a 'best fit' judgement as to the capability as a whole. The written commentary would be the opportunity to refer to specific areas which ought to be addressed within that.

<sup>80</sup> <http://dashboard.ofsted.gov.uk/>

The dummy report here shows what the a two page report may look like (full size version is in the appendix)



For some schools, the short inspection and the data assessment would identify a potential risk that the school was likely to be Requires Improvement or below. Similarly, what is termed here the 'failure triggers' may also indicate risk – which would principally be parental complaints but ought to also include concerns raised around financial sustainability or other issues (perhaps rapid leadership changes) by other agents in the system, most notably the EFA (and perhaps the Regional Schools Commissioners) for Academies and Free Schools, and Local Authorities for maintained schools. The Secretary of State would retain his power to ask Ofsted to make an immediate inspection of any school. Such risks would move the school into the new, tailored Ofsted inspection programme.

Importantly, so as to avoid the perception that the Tailored Inspection is a proxy for failure, or analogous to a monitoring visit, two other categories of schools would fall into risk. The first is where the data and the short visit prove inconclusive to make a judgement on a school. This is not to say Short Inspection is failing – indeed, the school may be hugely successful – but it is simply to say that a short visit cannot make a formal judgement. We envisage that this will particularly be the case with small schools and new schools, such as free schools, where there is not the wealth of historic data to be able to always make a reasoned judgement. Secondly, we believe that a small number of schools – maybe 5% – should randomly be sampled every year into the Tailored Inspection process, to ensure rigour in the self evaluation process from schools and quality control over the short inspection process.

The Tailored Inspection process would take elements of the current section 5 approach, as well as elements of the section 8 monitoring visits. It would operate as follows:

- It would take place over a longer time period. The current Section 5 tariff for inspection of schools and Academies allows for

Number of pupils	Total inspector days on site *
Under 50	1
51-150	2
151-250	4
251-400	5
401-600	6
801-1100	8
1101-1400	9
1401 or more	10

\* Inspector days equals number of inspectors multiplied by days on site. For example, a team of 2 inspectors who both attend two days of an inspection equals four inspector days. In addition, the tariff provides for the Lead Inspector to spend one day beforehand scrutinising the data and preparing for the inspection, and one day afterwards writing the report

The modelling for this report suggests increasing this to an average of 14 inspector days for a primary school, and 21 for a secondary school (including the two additional days for the Lead Inspector as before). As before, smaller schools will receive fewer days and larger schools receive more. **But this reform would mean approximately doubling the total amount of inspector days spent on Tailored Inspections, compared to a section 5 now.**

- The lead inspector and as many of the staff as possible would be HMIs. This is because of the additional quality control that Ofsted can have over the selection, training and management of staff who directly work for them, rather than under contract to organisations who themselves work under contract to Ofsted. This would mean a need to recruit additional HMIs nationally.
- The inspection team would be always matched by phase, and by subject specialism. As well as schools that require great scrutiny deserving to have inspectors that understand the nuance of the school that comes with phase specific understanding, subject specialism is also critical here. If a Short Inspection has flagged up a concern with the Achievement of pupils in say maths, then it seems absolutely essential that a maths specialist joins the Tailored Inspections team
- The inspection team would be trained to a high level, to allow them to carry out limited lesson observations, where necessary, to probe Quality of Teaching. This would require carrying out observations in the way identified by the MET study as being reliable, or else otherwise externally validated by an academic study commissioned in the UK to certify its rigour.<sup>81</sup> Ofsted would also have to prove that its inspectors met benchmarks of inter-rater reliability before they were able to carry out any inspections – with a higher bar for those carrying out Tailored Inspections.
- A school would be graded with an overall Grade as now, and 4 separate grades for the four current subgrades. Importantly, it would be open to a school to

81 A reminder that MET suggests two 45 minute lesson observations from within the school, one 45 minute observation by an external observer and three fifteen minute observations by different external observers. The recommendations below discuss further how Ofsted would need to certify their levels of training and frequency of observations to be allowed to carry them out during a tailored inspection



secure any Grade – moving to a Tailored Inspections would not be a limiting factor in a judgement.

- If a school received an overall Grade of Requires Improvement or Inadequate, then the monitoring process and interventions that take place as now would still continue, and these would be unaffected by this change to the overall inspection process. HW.

Finally, we recommend one last change to the framework. **In future, a school should not be able to be rated as Outstanding unless Ofsted judged that it was not just Outstanding in its own practice, but that it was engaged in a serious and meaningful way in some form of school to school improvement with other schools – as chosen by the school itself.** Such an approach, of course, is already practised by many schools through a myriad of routes – becoming a Teaching School, working within a Multi Academy Trust to support weaker schools, being part of a School Direct consortium, or being part of some of the exciting new, sector led networks for school improvement.<sup>82</sup> The intention of this change would be to systematise this model of school to school improvement across the country. Much of the future challenges in the school system will best be addressed by schools working in partnership, as set out in Policy Exchange’s previous report on school chains.<sup>83</sup> Most recently, some of the difficulties experienced by some of the larger chains has shown the need for smaller, localised partnerships.<sup>84</sup> This is likely to particularly be the case for primary schools, which is a future area of Policy Exchange work.

Such a change would give out a message that a school cannot be truly outstanding – or Outstanding – unless it is engaged in some form of beyond its own gate, school to school improvement, this would also give Good and Outstanding schools boost themselves, as they would likely benefit in some way from a partnership. It is however vital that the method it chooses to do so is at the school’s discretion. It must be able to consider the demand in its local area, and the school’s own capacity to support other schools. Placing a restrictive set of demands on a school that says only certain things will count would totally undermines the principle of school led improvement, as well as risking tipping schools too far into overstretch, at a cost to their own pupils. Rather, Ofsted’s approach should be like the Charity Commission’s approach to public benefit test ought to be for independent schools – “you tell us what you’re doing, and we assess whether it’s rigorous and extensive enough to qualify”.

### The benefits of the new Ofsted model

Moving to a two stage Ofsted inspection system as proposed above has the opportunity to transform the way in which Ofsted inspect schools and the way in which schools respond to the inspectors:

- It is right that Ofsted have now recognised the flaws in grading lesson observations. But even the current clarification only moves the issue on so far. So long as inspectors are observing lessons formally, and completing evidence forms, there will always be a risk that they will use personal preferences in making judgements either about the lesson as a whole, or in using these judgements to come to a conclusion about Quality of Teaching overall – rather than simply validating or otherwise the school’s own judgements on this issue

82 During this research report, we had discussions with both NAHT around their new model of school improvement – sometimes dubbed ‘Instead’ – and with Challenge Partners around their network. Both schemes have at their heart a model of schools inspecting other schools – called a Quality Framework or a Quality Assurance Review respectively, with Challenge Partners’ inspections led by an Ofsted accredited inspector. The idea is that schools both benefit from receiving formative assessment by their peers, and in return have to engage in other schools to scrutinise their own work. Both organisations feel that the attractiveness of the offer is that such inspection from their peers – whilst rigorous – is more helpful because it comes from an ‘intelligent accountability perspective’ and is seen as more constructive than an external inspection. We are also aware that many Multi Academy Trusts – including some whom we spoke to for this report – are also running similar systems, which go beyond CPD and into shared inspections. All of these are to be welcome and supported.

83 O’Shaughnessy J, *Competition meets collaboration: helping school chains address England’s long tail of educational failure*, Policy Exchange, 2012

84 This has been explored thoughtfully and constructively by Robert Hill in his work for ASCL and others on school chains, including most recently here ‘Quality not quantity is the litmus test for academy chain expansion’, February 2014

- Eliminating lesson observations also removes the need for many teachers to spend considerable time preparing ‘the perfect Ofsted lesson’. Despite all the current guidance, the uncertainty about the quality of inspections and the way in which lessons are observed means that this is sadly a rational response. The result is significant time wasted by teachers, a sizeable increase in stress, and the potential for children’s learning to actively be harmed, if teachers are trying to demonstrate ‘pace’ or ‘rapid and sustained progress’ in a 20 minute lesson, or trying to change behaviours which are seen as negative despite being poor proxies for learning, such as children reading quietly.
- Thirdly, eliminating routine lesson observations also saves significant time and manpower, given that the majority of inspector time is spent in lessons. **Modelling for this report calculates that the new two stager model set out here would mean almost 12,000 fewer inspector hours are needed across the system in a year, even accounting for almost doubling the inspector hours allocated for Tailored Inspections.** If these savings were realised when Ofsted comes to consider the future of the outsourced contracts in 2015, this has the potential to save Ofsted almost 6% a year.<sup>85</sup>
- A model which explicitly focusses on the school making its own self evaluation, which is then scrutinised and validated (or not) by Ofsted, will be of significant benefit to those increasing number of schools which operate using innovation in some way, or are otherwise different (because of size, or length of time opening). The clear policy direction of the government is to allow schools greater autonomy, and for them to innovate within this.<sup>86</sup> Deregulation of the national pay and conditions document for main scale teachers has devolved down to individual schools decisions on teacher appraisal, pay and progression. Forthcoming changes such as the abolition of National Curriculum levels will require all schools to design their own models of pupil tracking. Already increasing numbers of Academies are disapplying elements of the National Curriculum. New Free Schools are opening every year that do not have historic data and standardised ways of operating. Simply reminding Ofsted inspectors that they must treat all methods as valid if the results are good will not work, whilst the current framework – with lesson observations, and a handbook that in many ways does ask the inspectors to make value judgements – and the current model of the workforce exists. Only a fundamental change to the way in which inspections work will allow innovative schools not to risk being disadvantaged through unfamiliarity.
- An even more proportionate inspection system would allow Ofsted to focus more of their resources on the small number of schools where they are in need of external support. The evidence in this report suggests that Ofsted’s record at school turnaround is reasonably positive in that most schools improve from Requires Improvement (or Satisfactory) to Good as a result of Ofsted’s monitoring regime. Dedicating resources to that, and further improving it, is better value for money than scrutinising schools in depth who are mostly capable of evaluating and demonstrating their own progress and improvement mechanisms
- Finally, a more limited model, with fewer inspectors, instantly allows for an increase in quality control. We heard from the call for evidence a continuous request from regional providers to sign up teachers to become inspectors, and a suggestion (which we have not been able to verify) that this was

85 This modelling includes taking account of the fact of an average doubling of the time for tailored inspections. It also assumes that there will be some substitution between AIs and HMIs as Ofsted reinvests some of the savings in the former by employing more of the latter to lead tailored inspections. We do not make any assumptions on whether short inspections will be led by inspectors on outsourced contracts or by directly employed Ofsted inspectors, but the day rate calculation is based on current outsourced contracts so Ofsted would need to include on costs (eg NI and pensions) within this overall envelope if it wished to bring these inspectors back in house.

86 It is important to clarify here the use of the word innovation, which is chosen deliberately in opposition to what might be called ‘expected’. It does not necessarily refer to what is sometimes called ‘progressive’ teaching or is used as a synonym for consciously engaging and technology heavy education. A school which chose relatively traditional teaching methods – and in so doing, moved away from a previous model of teaching – could still, under this definition, be called innovative.

because there was a shortage of qualified people who wished to carry out school inspections – exacerbated by Michael Wilshaw’s (sensible) belief that more inspectors should be serving school leaders. In any situation whereby there is a labour shortage or near shortage, there will be quality control issues – and as discussed above, there are many concerns with the quality of some current inspectors, hence the bringing back of training in house. By reducing the demand for inspectors, this instantly allows Ofsted and/or the contractors to become more selective in who they hire, improving quality at a stroke.

In summary, the proposal here would be a radical move towards improving the quality of school inspections. It would make a crystal clear reality of the desired outcome whereby that schools should lead the improvement system and primarily hold themselves to account, and that Ofsted should act in partnership with schools as a check on that to ensure the systems that schools are using are satisfactory. This would return Ofsted to being the hygiene inspector – someone who checks empirically to see if the kitchen is clean and the food is fresh – and away from being the food critic, who passes judgement on the basis expertise but also personal preference not just on the outcome, but the way in which the outcome is put together. If a school is underperforming, then it is right that Ofsted challenges that, alongside other accountability measures. But no longer will Ofsted – by accident or design – be able to routinely pass judgements on the ways in which schools operate – simply what the outcomes are for children and young people.

### Further changes

In addition to the over-arching changes outlined above, there are some specific steps Ofsted could take to improve the quality of inspections under this new system. Some of these are to do with improving the quality and consistency of individual inspectors and inspection teams, and some are changes that can be made to the methods and systems they use.

#### Inspector quality

##### *Recommendation 1*

One common complaint that emerged was that inspectors simply didn’t understand the setting they were inspecting because of a lack of experience or knowledge of a certain phase or kind of education. This was particularly true for Special schools, although we equally had complaints that inspectors experienced in Secondary did not have sufficient understanding of the needs of Reception children, and vice versa. For example, when it comes to judging progress, the Ofsted handbook emphasises that scrutiny of children’s books should play a key role in reaching conclusions; however as discussed earlier in this report assessment is a highly skilled task, which requires experience and curriculum knowledge. When it comes to interpreting data, particularly in Special schools, expert knowledge is required to understand the tracking of progress of children with multiple needs. As already outlined, in the fuller second inspection schools would have subject and phase experts in order to really pinpoint difficulties. However this should be extended further so that:

**The Ofsted inspector person specification should be tightened so that all Inspectors should only be allowed to inspect a school when they have relevant and recent teaching experience in Special, Primary or Secondary Schools, or a high knowledge of assessment and pedagogical practice in that area.**

#### Recommendation 2

One of the themes that emerged from discussions with teachers, and from the call for evidence, was a concern around inspectors' training. The move to bring training in-house may address some of this. However, there is still more room for ensuring accuracy and consistency.

Professor Coe has argued for an exam for Ofsted inspectors, a suggestion which this report agrees with.<sup>87</sup> Inspectors should have to pass a test in order to become accredited, and update this on a regular rather than be 'signed-off' after completing the required training and shadowed inspection. For all inspectors this test would include the interpretation of progress and attainment data, to ensure they understood all the issues of small sample sizes and statistical significance. They would have to show that their judgements had a reasonable level of agreement with Ofsted's judgements. In addition for those who would be taking part in the full-scale inspections, they would need to pass a test to ensure the reliability of their judgements of teaching based on lesson observations. While it is of course still important that inspectors are able to exercise professional judgement, it is essential when it comes to interpreting key pieces of data that they show they can do so accurately.

**Inspectors should have to pass a data interpretation test in order to become accredited. Such accreditation should be time limited and regularly renewed – perhaps every five years. For inspectors who wish to deliver Tailored Inspections, they will need to be trained in lesson observations to the extent set out by MET, or as the result of a specific UK study commissioned to identify the conditions necessary for a high level of validity and reliability. All elements of this process – the data interpretation test, the extent of renewals, and the protocols around lesson observation drawn from MET or a new UK study – should be transparent and visible to anyone, including schools who are being inspected.**

#### Recommendation 3

While a more rigorous training system should go a long way to ensuring accuracy and consistency of inspection judgements, there also needs to be ongoing monitoring to make sure that Ofsted are quality assuring inspections in an appropriate way. Currently, reports will be signed off by Ofsted, and a proportion will have in depth evidence reviews in order to quality assure them. Inspection teams may be additionally be visited by HMIs to check their practice. These are sensible measures, but more work could be done.

The Health and Safety Executive (HSE) organises moderation days, where different teams share evidence from inspections and decide what judgements they would have reached. Additionally, on occasion they will have repeat inspections – where a second inspector is sent out to verify the work of the first.

**Ofsted should consider how to introduce additional methods to test the reliability and validity of their inspections on a randomised basis, not just**

<sup>87</sup> Coe R, *Critics should not have to prove Ofsted's judgements are wrong; it should be up to Ofsted to prove they are right*, TES, October 2013

**when complaints or appeals are raised by schools. This could include random sampling follow up moderation days**

**Methodology**

The main recommendations for changes in the Ofsted methodology have of course been outlined above under our new inspection regime. However there are other important steps that should be taken alongside this to ensure that both the lighter touch inspections and full-scale ones are using a more reliable evidence base.

*Recommendation 4*

Under the new system, as currently, a vital part of the evidence on a school's performance will be the progress and attainment data, so it is vital this is accurate. Some improvements have been made, with the Progress 8 measure being introduced and (if it is carefully designed and implemented) the proposed baseline assessment at age four. However, teachers admitted to us that they worried other schools inflated assessments, that they did so themselves or that they knew of others who had, in order to improve their data. And this is not necessarily done intentionally; assessment is a difficult skill to master and some teachers are better than others. In order to address this, and make sure that the data judgements are based on can really be trusted, increased moderation at a local level should be introduced. Schools should be principally responsible for deciding how their judgements are moderated, and how frequently, and this will typically be done on a risk basis and depending on the in house experience of the school. Ofsted will need to make a judgement on validating these processes. If an inspector is not sure of the validity of the data, then this would be a trigger that would typically require a Tailored Inspection.

**Schools' internal assessment procedures should be validated by Ofsted as to their rigour and frequency, to ensure moderation is reliable**

*Recommendation 5*

One of the methods of teacher evaluation that has been shown to be very accurate has been that of student surveys. Of course, Ofsted inspectors will currently speak to students about their work and experiences in the school – however this is not done in a comprehensive fashion. It is possible to design surveys with a range of questions about children's school experience which in fact give an accurate picture of the quality of education they receive. In addition, through surveys there would be opportunity to explore questions around bullying and safety, which it may be hard for an Ofsted inspector (who is, after all, a stranger in the school for only a few days) to glean from conversations. Of course, younger children may be less able to answer some of these questions, and there would be concerns from teachers about students using them in an overly negative or unfair fashion. However the evidence – including from the TELLUS surveys – suggests that if carefully designed and used, student surveys could be an important additional tool in school inspections, though likely more applicable in secondary schools and at the higher end of primary schools.

**Ofsted should pilot a survey of students' school experiences, including views on teaching, bullying and safety. This should be published where used and made available on the Short Inspection report card. This pilot should**

---

**be tested against other judgements made on these elements to explore its reliability and validity ahead of a possible wider roll out.**

*Recommendation 6*

Ofsted have been making great efforts to clarify that there is no required teaching style, but as well as difficulties in the message being conveyed, there are some inconsistencies in the language used. For example in the Ofsted handbook it states that during lesson observations inspectors should ‘assess the extent to which pupils have grown in knowledge’. It is phrases like this which lead to the situation many teachers complain of where they feel the need to show, when observed for twenty minutes, how much the children have learnt, using techniques such as ‘mini-plenaries’ where they check on progress halfway through a lesson. In reality it may not always be useful or possible to demonstrate growth in knowledge in the time of an observation. This sort of language should be removed from the handbook. Likewise, when it comes to other messages coming from Ofsted there are some problems; the Annual Report includes advice about how to pace a lesson, for example. One response from our call for evidence reflected that pace was now a buzz-word in their school. While Ofsted clearly has a wealth of knowledge about teaching, if they truly believe they should not advocate a certain style they need to be very careful in all communications, and in the handbook, to ensure that they are not influencing practice in a particular way.

**Ofsted should be exercise more caution in publications which seem to endorse certain teaching methods.**

*Recommendation 7*

The views of parents’ about their child’s school are important. Ofsted recognises this and has introduced the ParentView site where parents can log-in and answer a survey. However, it is very easy for anyone to log-in to provide feedback as no verification is required (albeit users have to tick an honesty box at the end of the survey). Conversely, many schools have low levels of parental feedback.

**Building on other areas of public policy – such as tax letters and organ donor registrations – Ofsted should work with the Behavioural Insights Team (now mutualised) to trial different models of ensuring high level of parental sign up to the survey, combined with low levels of fraudulent feedback.**

**System**

These changes to could go a long way to improving the accuracy of individual school inspections. However changes could also be made at a system-wide level to improve the efficiency of Ofsted.

*Recommendation 8*

In any inspection the leadership and management of a school will be judged. When it comes to schools in Academy chains, these judgements will necessarily include opinions of the chain itself, however they are not currently inspected. On one level, this is simply inefficient; one Academy chain CEO explained how he was interviewed twice in a day, with the same questions, by two inspection teams inspecting two of the schools in his chain. This creates excessive extra work for each Ofsted team, as they each individually assess the competency of the chain.

Additionally, having several inspection teams asking very similar questions is likely to gather less information than one inspection team really interrogating the evidence a chain can produce. They would also gain evidence about schools the chain were more concerned about, and be able to target future inspections more carefully. As more and more schools enter Academy chains it is essential that they have they are held to account. When important decisions are made about whether chains should be allowed to expand or not, it would be a useful extra resource to have the Ofsted evaluation of their competency.

**Ofsted should design a system for inspecting Academy chains.**

*Recommendation 9*

This report has discussed extensively the difficulties with the current model of some outsourced Additional Inspectors, and a ‘twice removed’ system of accountability. The new 2 stage model of Tailored Inspections recommends that such inspections are always led by an HMNI, with as many staff as possible also being HMI; this will mean Ofsted needing to reduce the number of AIs they contract for and increase their in house recruitment of HMIs. Some organisations – including ASCL – have called for Ofsted to completely abolish the contracts with AIs and bring all inspections – including what this report calls short inspections – back in house. This report is neutral on whether, in principle, using AIs for short inspections could work. Should Ofsted continue to use AIs, however, when it comes to renegotiate its contracts in 2015, it should place a new requirement on contractors that their AIs work for them full time, as so many other outsourced contracts require. This would ensure that the contractors would feel some organisational loyalty to the contractor, and ideally to Ofsted. Having AIs working under employment to the contractor would also mean that information would be able to cascade more effectively, and training, development and if needed addressing individual weaknesses can be addressed more effectively because of the presence of organisational wide techniques for this.

**Ofsted should consider carefully whether it retenders its contracts for Additional Inspectors when the contracts are re-let in 2015. At a minimum, this new 2 stage model will require fewer inspectors to be recruited, with a strong preference for tailored inspections to be largely or entirely HMI staffed. Should Ofsted decide to let a reduced contract for AIs to carry out short inspections, it should place a condition that AIs work full time for the contractor so as to ensure organisational loyalty and mechanisms for development and information flow.**

---

# Appendix 1:

## Summary of Call for Evidence Responses

During the course of our research we sent out a ‘call for evidence’ to learn more about teachers experiences of Ofsted. This was hosted on our website for a month and shared on social media. Of course, there is likely to be a self-selecting bias in respondents to the survey, so caution was exercised in interpreting the results.

The questions asked were:

- Do you feel Ofsted inspectors make accurate assessments of teaching in your school, and the progress of students, through the use of data, lesson observation and other techniques?
- Do you feel under pressure to adapt the way you teach, mark or any other practices to meet Ofsted requirements, or encourage others to do so? Do you feel this is positive or negative?
- Have you ever felt under pressure, at your current school or previous schools, to ‘game’ the system, either by presenting data in a certain way, or changing your normal practice during an inspection?

We received 305 responses to this questionnaire:

---

Teachers (including heads of year and departments, and one retired teacher)	169
Headteachers, Assistant Headteachers and Senior Leadership (including two retired headteachers)	94
Governors	23
Education Consultants	4
Teacher Trainers	2
Lecturers	4
Academy Chains	2
Other	7

---



The following is the breakdown of respondents working in various categories of school:

Nursery	5
Infant	5
Primary with Nursery	7
Primary	89
Middle	1
Primary and Secondary	2
Secondary	118
All-through	2
Special School	4

The following is the number of respondents working in schools with Ofsted scores 1 to 4 (percentages do not equal 100 due to rounding).

Outstanding	29	14%
Good	84	41%
Requires Improvement	79	39%
Inadequate	11	5%

The main themes that emerged from responses were:

- The variability of quality between different inspection teams
- Inspection teams making decisions based on the data before they had come into the school
- Short lesson observations not allowing for accurate judgements
- Problems with the reliability and interpretation of data
- Some leadership teams very guided by ‘what Ofsted wants’
- Pressure felt to adapt teaching style when Ofsted are present
- Some teachers felt pressure to inflate levels to show better progress
- Positive attitudes towards HMIs compared to Additional Inspectors

## Roundtables

In addition we held two roundtables, one for headteachers and one for classroom teachers. The breakdown of attendees is listed below. We are very grateful to Edapt, NAHT and ASCL for their help in organising these events, and all the headteachers and teachers for attending:

Headteachers in Primary Schools	5
Headteachers in Secondary Schools	6
Headteachers in Special Schools	1 (secondary)
Teachers in Secondary Schools	9
Teachers in Primary Schools	2

---

The main themes that emerged from the headteachers roundtable were:

- More experience as a head helps you to manage Ofsted inspectors
- Serious problems around data systems when it comes to Special Schools
- Lesson observations are too short to be meaningful
- Appeals process is lengthy and difficult
- Time and expense reasons for not wanting to become inspectors
- Better headteachers protect staff from the pressures of Ofsted
- Huge variability in inspection teams
- Inspectors need experience in the kind of school they are inspecting
- Undergoing inspection a very stressful experience

The main themes that emerged from the teachers roundtable:

- Inspectors struggle to understand teaching styles they are not used to, and grade them down
- A bad teaching grade can destroy a teacher's confidence
- Lesson observations too short
- Stronger leadership teams don't focus on Ofsted, weaker ones do
- HMIs were more helpful
- Ofsted a useful driver of change when a school is really struggling

## Interviews

In addition conversations with a number of individuals and organisations during the course of our research; in particular we are very grateful to those listed below for their time and insights.

- Sir David Carter, CEO Cabot Learning Foundation
- Professor Rob Coe, School of Education and Director of the Centre for Evaluation and Monitoring (CEM), Durham University
- David Didau, Author and Education Consultant
- Russell Hobby, General Secretary of NAHT
- Dame Sue John, Headteacher and Senior Partner at Challenge Partners
- Mary Myatt, School improvement adviser who leads inspections
- David Weston, CEO Teacher Development Trust
- Stephen Williams, The Health and Safety Executive
- CBI
- Ofsted

# Appendix 2: A Dummy 2 Page Report Card for a Short Inspection



## Ofsted Report Card

### Anytown School

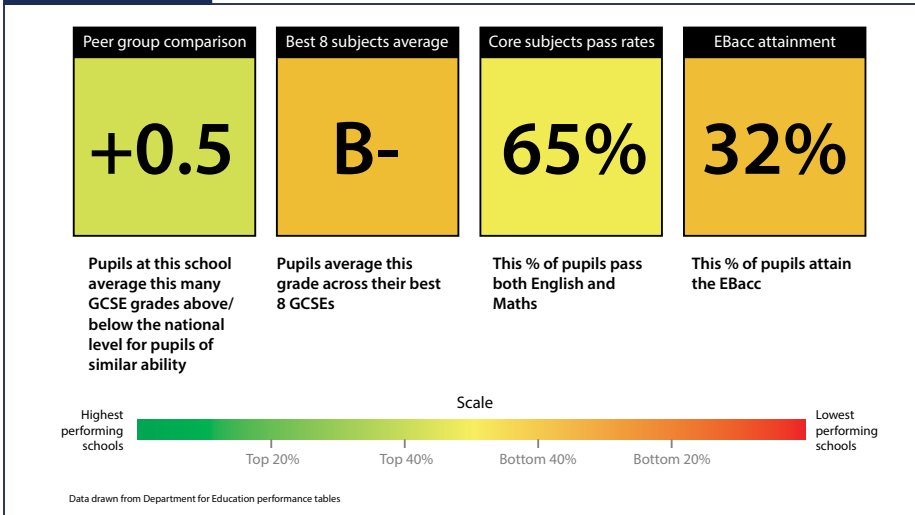
Headteacher: Mr J. Smith  
Crete Street, Morton, Hertfordshire, FP15 3PQ  
Local authority: Morton District Council

Inspection Date: XX XX XXXX

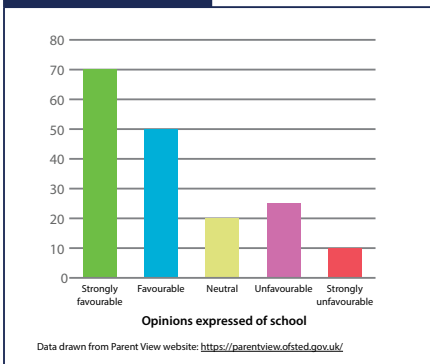
Overall Grade  
**Good**

School Capability Grade  
**Good**

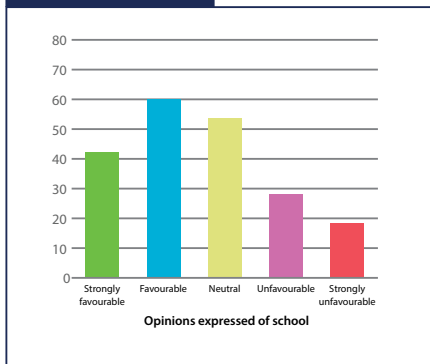
#### School Performance



#### Parent View



#### Student View



## Ofsted Report Card

### Inspector's Comments

Anytown School is a **Good** school, with **Good** school capability to improve. The data provided by the school shows that the achievement of the pupils aged 16 is broadly in line with national expectations. The school accurately identified the need to ensure all Pupil Premium pupils make expected levels of progress and have put in place interventions to help them do so. The school judges that the quality of its teaching is Good and provided a range of evidence to support this judgement including written staff records of lesson study, examples of pupils' marked work, and results of the performance related pay awards for staff and the evidence that sat behind those. These all support such a judgement. During my visit, the headteacher showed a strong understanding of his school and its areas of strength and development; he is supported in this assessment by a large and engaged Governing Body. The accuracy of the head's judgements also support the school's own leadership and management. During my time in the school, including a walk around at break time, a visit to the school gates, and an informal 5 minute drop in to a Maths lesson, pupil behaviour was mixed, with instances of low level disruption, although no serious concerns were identified. Parents and pupils speak, on the whole, happily of their school but some raise concerns about support for all pupils and dealing with elements of behaviour.

Inspector: M. Bloggs

Date: XX XX XXXX

### Recommendations

To improve further and become Outstanding, Anytown School should:

- Ensure that all pupils, including Pupil Premium pupils, make progress in line with or above national expectations for both cohorts respectively
- Ensure that behaviour of all pupils is consistently good and that all pupils and staff are aware of and participate in a sanctions regime for addressing poor behaviour, including low level disruption

### Pupil Characteristics

	This school	Local authority average	National average
Number of pupils on roll	800	875	922
Number of children with Special Educational Needs	24	29	75
Number of pupils eligible for Pupil Premium	159	165	215



£10.00  
ISBN: 978-1-907689-73-4

Policy Exchange  
Clutha House  
10 Storey's Gate  
London SW1P 3AY

[www.policyexchange.org.uk](http://www.policyexchange.org.uk)