

Requires Improvement

**A new role for Ofsted and
school inspections**

Tom Richmond



About the author

Tom Richmond is the Director of EDSK.

He has spent over 15 years in the world of education. He began his career teaching A-level Psychology at one of the country's leading state schools, having gained a BSc in Psychology from the University of Birmingham and an MSc in Child Development from the Institute of Education in London.

After three years in teaching, he moved into politics to work on policy development and research across the education, skills and welfare sector. This included roles at think tanks such as Policy Exchange and the Social Market Foundation, Pearson, G4S, a leading professional body and working for an MP.

He subsequently spent two years as an advisor to ministers at the Department for Education, first under Michael Gove and then Nicky Morgan, where he helped to design and deliver new policies as well as improve existing ones. After leaving the Department for Education, he spent two years teaching at a Sixth Form College and then moved back into education policy and research, first at the Reform think tank and then at Policy Exchange.

He has also written extensively for publications such as the TES and has appeared on numerous media outlets including the BBC News Channel, BBC Radio 4 Today, BBC Radio 5 Live and TalkRADIO.

Acknowledgements

The author would like to thank the four external peer reviewers of this paper who kindly provided their feedback and comments during the drafting process.

Thanks also to those who commented on specific sections of the report or provided expert input for one or more of the recommendations.

The views expressed in the report are attributable to the author and do not necessarily reflect those of EDSK's advisory board or supporters. Any errors remain the responsibility of the author.

About EDSK

EDSK is an independent, impartial think tank.

Our mission is to design new and better ways of helping learners of all ages succeed.

We aim to achieve this by conducting research on topics across the education and skills landscape and then producing evidence-based recommendations for where the system needs to go next, both now and in future.

Our research covers primary schools, secondary schools, colleges, apprenticeships and universities. Whichever topic or policy area we are investigating, our priority is always to produce better outcomes for learners while also ensuring that the greatest support is given to the most disadvantaged members of society.

We work closely with a range of individuals and organisations to ensure that our research reflects a wide range of evidence on any given issue. EDSK retains copyright and full editorial control over our research, irrespective of how it is funded.

EDSK is a not-for-profit think tank registered in England and Wales.
(Company Registration No. 11795717)

Find out more at www.edsk.org

Contents

	Executive Summary	1
1	Introduction	7
2	Are school inspections valid and reliable?	11
3	Does Ofsted look at the right things?	19
4	What impact does Ofsted have on teachers and leaders?	23
5	Do Ofsted inspections lead to school improvement?	27
6	Recommendations	29
7	Areas for further consideration	45
	Conclusion	47
	Appendix	49
	References	50

Executive Summary

Ofsted's ratings of each state school as 'Outstanding', 'Good', 'Requires Improvement' or 'Inadequate' are frequently used by parents. The giant PVC banners that adorn many school gates are testament to how much weight is placed on the judgements made by inspectors. Ofsted's own research has shown that their ratings are the second most important factor for parents when choosing a school (behind 'proximity') and two-thirds of parents believe that Ofsted is a valuable source of information about education in their local area.

Ofsted occupies an important position in our education landscape and the process of inspecting schools can, if done in an appropriate and proportionate way, have a positive effect on the quality of education being delivered. Even so, inspecting over 20,000 state schools is a considerable undertaking and our school system has also changed dramatically over the past quarter of a century. Consequently, this report seeks to answer four questions: first, can parents trust the grades given to schools by Ofsted; second, is Ofsted measuring the right things; third, what impact does Ofsted have on teachers and school leaders; and fourth, do Ofsted inspections lead to better schools?

Are Ofsted judgements reliable (consistent) and valid (accurate)?

Since it was created over 25 years ago, Ofsted has not published any research to support the notion that their judgements on schools accurately reflect the quality of education that a school provides. As far back as 1996, Ofsted's research had shown that pairs of inspectors awarded different grades after observing the same lesson in 33% of cases. A parliamentary committee in 1999 heard evidence from academics who declared that the consistency in inspectors' judgements was simply "not there" and that the judgements were "not very reliable". It was almost two decades before Ofsted revisited the issue of reliability in a 2017 study that showed inspectors largely agreeing on simple binary judgements about each school, but the findings were criticised in terms of the design of the study and the failure to secure greater agreement between even the most experienced inspectors.

Independent research on the ability of observers, including highly-experienced school leaders, to accurately judge what is happening in a classroom is equally concerning. Numerous studies have shown that you need multiple observers watching the same teacher on multiple occasions before their opinions start to converge (even then, observers tend to disagree about 20-30% of the time). This compares to Ofsted's stated approach of an inspector "spending a few minutes in each [lesson]". Research has also found that overall judgements on schools made by Ofsted inspectors were no better at predicting a secondary school's future performance than just looking at the past two years of GCSE results. In addition, of those primary schools performing below the government's minimum 'floor

standard', 69% were rated 'Good' or 'Outstanding' by Ofsted. For secondary schools below the 'floor standard', 35% were rated 'Good' or 'Outstanding'. Other research has shown that of those schools that had received 'Outstanding', 'Good' or 'Requires Improvement' grades at their previous inspection but whose performance had since deteriorated substantially, 47% of primaries and 33% of secondaries actually saw their Ofsted grade *improve* at their next inspection. This shows that many schools performing very poorly in national assessments are still being awarded high ratings by Ofsted.

Ofsted's survey of teachers in 2018 found that 72% agreed and only 5% disagreed with the statement that 'Inspection judgements lack consistency and are too prone to be influenced by the personal views and prejudices of individual inspectors'. Just 62% of teachers felt their school's overall grade was a 'fair and accurate assessment' and only 35% of teachers agreed that Ofsted is a 'trusted arbiter' of school standards. Furthermore, the National Audit Office (NAO) found that 1,620 schools had not been inspected for six years or more, including 296 schools that had not been inspected for at least 10 years. In response to the NAO report, Ofsted accepted that they couldn't tell if those schools had since "become middling, or mediocre or coasting."

The new 'inspection framework' released for consultation by Ofsted in January 2019 contains some positive developments such as paying more attention to how a school designs and implements its curriculum. However, Ofsted has not published any studies showing that inspectors will be able to make consistent judgements in these areas across thousands of schools, even though the new framework begins operation in September this year. Ofsted has also stated that lesson observations will still carry considerable importance in future inspections, despite the wealth of evidence suggesting that they are not valid and reliable tools.

What's more, Ofsted has signalled that 'book scrutiny' will play a much greater role in inspections under the new framework. Inspectors will spend more time comparing samples of books to "evaluate pupils' progression through the curriculum". The prospect of exercise books being scrutinised in this way raises crucial questions about reliability and validity yet again. Ofsted has not published any research to demonstrate that inspectors can make accurate and consistent judgements on pupils' work across the full range of National Curriculum subjects, GCSEs and A-levels, which is compounded by the fact that inspectors may not even be experts in the subjects or year groups that they are inspecting.

Does Ofsted look at the right things?

Vague terminology from Ofsted is a cause of considerable frustration for teachers and school leaders. For a school to be judged 'Outstanding' for teaching, learning and assessment under the current inspection regime, teachers must "provide pupils with incisive feedback". At no

point in the school inspection handbook does Ofsted explain what ‘incisive feedback’ means in practice. The new inspection handbook for September 2019 seems to fare little better. For example, when judging the ‘Quality of education’, teachers will now have to “provide clear, direct feedback” without any explanation of what will be considered unclear or indirect feedback. To be judged ‘Good’ or ‘Outstanding’ for the ‘Quality of education’, school leaders must not use pupil assessments “in a way that creates unnecessary burdens on staff or pupils”, yet there is no clarity on how the ‘necessity’ of these burdens will be assessed. When judging the ‘Personal development’ of pupils in future, a school can only be rated ‘Outstanding’ if it “consistently goes ‘the extra mile’ to promote the personal development of pupils, so that they have access to a rich set of experiences”. Definitions of ‘going the extra mile’ and ‘rich experiences’ were notable by their absence.

The fact that different Ofsted judgements appear to overlap significantly with each other also raises questions about their value to parents and other stakeholders. The overall grade awarded to a school is the same as the subgrade for ‘Outcomes for children and learners’ in 97% of cases as well as the subgrade for ‘Quality of teaching, learning and assessment’ in 96% of cases. This suggests that process-based judgements made by inspectors, and all the associated subjectivity and confusion, may be largely redundant if the data on outcomes can capture the effectiveness of a school and communicate this to stakeholders.

There are also many aspects of school life that have previously not received much attention from Ofsted. As mentioned above, the increased emphasis on a school’s curriculum in the new inspection framework is a welcome development in principle. Moreover, extra-curricular activities and careers advice are barely mentioned in the current inspection handbook so it is encouraging to see the new handbook go some way to addressing these historical omissions.

The impact of Ofsted on teachers and leaders

For several years, Ofsted has published what they call ‘myth busters’ that seek to highlight specific practices that are not required by Ofsted but are still found in some schools. For example, one myth buster states that “Ofsted does not expect to see any specific frequency, type or volume of marking and feedback”. Nevertheless, inspectors continue to make comments about precisely these areas, as demonstrated by recently published Ofsted reports. A survey by the Association of School and College Leaders showed that schools are still being asked for evidence by inspectors that Ofsted says they do not have to provide.

Inevitably, the influence of Ofsted can affect staff workload. Ofsted’s own surveys of teachers have shown that 76% thought school inspections would be “highly stressful for everyone” and over half of teachers felt that an inspection would “mean a huge amount of unnecessary extra work”. Most concerningly, 83% of teachers agreed that “Ofsted

inspections introduce unacceptable levels of burden into the system". An academic study in 2017 showed that "policy decisions [in schools] were often being made to conform to Ofsted's expectations" and that "schools to some extent performed 'the good school' for inspections". A survey of school leaders found that they "regularly felt incentivised to prioritise the interests of the school over the interests of particular groups of, usually more vulnerable, children." The quality of education available at any given school could easily be diminished if teachers and leaders are more worried about what is best for Ofsted than what is best for their pupils. This has become even more pertinent in the context of on-going issues with teacher recruitment and retention.

Ofsted is keen to emphasise its commitment to reducing workload. For instance, under the new inspection framework inspectors will no longer look at internal progress and attainment data of current pupils or students in the hope that staff will spend less time on data collection and analysis. That said, even if Ofsted no longer pays attention to internal performance data, the emphasis on book scrutiny in the new inspection handbook could easily replace one significant source of workload with another equally absorbing and burdensome set of tasks. Surprisingly, Ofsted has also decided that inspectors will now formally judge the workload faced by teachers and potentially mark a school down if staff workload is deemed too high. For an inspectorate that is widely recognised as being one of the root causes of many workload pressures to be judging school leaders on the workload they generate for their staff is counterintuitive, to say the least.

Do Ofsted inspections result in schools improving?

Although many schools improve each year, recent data from Ofsted showed that 33% of primary schools and 58% of secondary schools rated as 'Requires Improvement' did not subsequently improve and a higher proportion of schools rated 'Requires Improvement' (11%) declined to 'Inadequate' compared to previous years. The NAO found that "Ofsted does not know whether its school inspections ...raise the standards of education and improve the quality of children's and young people's lives", although Ofsted's new 2017-2022 strategy has attempted to address this by establishing several key performance indicators and targets.

The views of headteachers and teachers on the impact of Ofsted inspections on school performance are decidedly mixed. In the NAO survey, only 44% of headteachers said that inspections had led to improvements in the school and Ofsted's survey of teachers in 2018 found that just 31% shared this opinion. Academic research has concluded that a causal link between inspections and school improvement "cannot be clearly supported from the literature." Looking across the full range of evidence, there does not appear to be a compelling case for concluding that Ofsted inspections necessarily lead to better schools.

Conclusion

Despite their desire to be seen as a 'force for improvement', Ofsted's own survey showed that just 24% of teachers agree that they perform this function. This report describes how Ofsted's approach to judging teachers and schools is not based on research evidence from this country or abroad, nor have the reliability and validity of their judgements ever been satisfactorily tested and assured. While these issues could be tackled in isolation, the widespread effects of Ofsted's approach to inspecting schools show that merely modifying existing processes and procedures will not be sufficient to bring about the changes that are now required.

There is still a crucial place in our education system for an independent inspectorate that visits schools to provide parents of current and future pupils with valuable insights. On that basis, this report seeks to craft a new, more constructive role for Ofsted in which their goal is to help parents make more informed choices instead of trying to come up with all the answers themselves. By moving away from the notion of 'grading' schools and towards empowering parents with better information, the education community as a whole will reap the benefits of having a self-improving school system that includes Ofsted, but without teachers and leaders having to experience the burdens that inspections generate at present. If this can be delivered, our education system will be a much better, happier and healthier place.

RECOMMENDATIONS

Part 1: Creating a new role for Ofsted

- **RECOMMENDATION 1:** Ofsted should not provide an overall rating or grade for schools due to a lack of reliability and validity in their judgements.
- **RECOMMENDATION 2:** Ofsted should not conduct any observations of lessons or scrutinise pupils' work until their processes and procedures have been rigorously tested in terms of both reliability and validity.
- **RECOMMENDATION 3:** Ofsted inspections should be entirely focused on observable aspects of school life that are not captured by performance data. This includes pupil behaviour and safety, the curriculum, careers advice and extra-curricular activities.

Part 2: A better way of holding schools to account

- **RECOMMENDATION 4:** A new 1-page School Information Card (SIC) should be published on an annual basis for every school (see example on page 41). The SIC will contain 12 indicators, covering four performance measures (e.g. pupil progress), four judgements made by Ofsted inspectors (e.g. curriculum quality) and four measures of wider school life (e.g. student and staff survey results). The measures on the SIC will be presented in both absolute and relative terms to provide a comprehensive view of the school.
- **RECOMMENDATION 5:** Ofsted inspections should be carried out every 2-3 years, with inspectors spending a day in each school to update their judgements. Ofsted should aim to cover two of their four new judgements as part of each visit (e.g. a 1-day inspection covering behaviour and the curriculum). The revised judgements will be added to the SIC as soon as possible after the inspection has taken place.
- **RECOMMENDATION 6:** In future, Ofsted reports should be much shorter than at present – ideally no more than two pages. Using the information published on the SIC, they should provide commentary for parents on two main areas: the current strengths of a school, and its areas for improvement. These reports will be regularly updated by Ofsted.
- **RECOMMENDATION 7:** The proposals in this report should be introduced by Ofsted as their new inspection framework in September 2020 – a year later than currently planned. The intervening period should be used to trial the proposals outlined in this report including the new SIC, the updated stakeholder surveys and Ofsted's revised role and responsibilities (e.g. assessing the validity and reliability of scrutinising pupils' work).

1. Introduction

While schools are well used to policies and initiatives coming and going over the years, the late 1980's and early 1990's witnessed several substantial changes that have stood the test of time. The introduction of GCSEs, a National Curriculum and SATs remain core features of our education system today, yet the creation of the Office for Standards in Education (Ofsted) in 1992 has arguably proved the most controversial of all the reforms brought in during this period. Ofsted emerged from the *Education (Schools) Act 1992*, which required the Chief Inspector for England to keep the Secretary of State informed about:

- The quality of education provided by schools in England;
- The educational standards achieved in those schools;
- Whether the financial resources made available to those schools are managed efficiently;
- The spiritual, moral, social and cultural development of pupils at those schools.¹

The Chief Inspector also had a duty to ensure that all schools in England were inspected, with the Ofsted-appointed inspectors reporting on each school in relation to the areas outlined above.²

In the same month that Ofsted was created in 1992, the then Conservative Government released its manifesto for the upcoming General Election. Their pledge to “introduce, for the first time, regular independent inspection of all schools” was motivated by their desire to provide information on the performance of all local schools to parents, “enabling them to exercise choice more effectively”.³ Moreover, “independent inspection of schools will provide parents with straightforward reports on their child's school, together with an action plan from governors to remedy any weaknesses.”⁴ Promoting informed choices by parents was part of a wider drive within the manifesto to “extend competition and accountability in public services.”⁵

Ofsted has changed considerably as an organisation since it first began inspections in September 1993. The length of inspections has shrunk from over a week to just one day in some cases while the notice given to schools before an inspection takes place has dropped from several terms in advance to just 24 hours.⁶ The cycle of inspections changed from every four years to every six soon after Ofsted was set up. Schools that are now thought to be underperforming can have Ofsted return soon after a previous inspection, whereas schools that are judged to be performing well may not see Ofsted return for over ten years.⁷ The

days of teams of 15 inspectors descending on a large school to inspect every National Curriculum subject as well as cross-curricular aspects of provision are also long gone.⁸

Similarly, the grading system used by Ofsted has evolved over the years. When Ofsted first began inspections it used a five-point grading scale for lessons (it did not assign an overall judgement to schools), but it soon switched to a seven-point grading scale ranging from 'Excellent' to 'Very Poor'.⁹ 2005 saw a major overhaul of several aspects of the inspection process, part of which was a shift to a four-grade system for judging a school:

- 'Outstanding' – Grade 1
- 'Good' – Grade 2
- 'Satisfactory' – Grade 3
- 'Inadequate' – Grade 4

This grading system remains in place today, albeit with a change of language at Grade 3 from 'Satisfactory' to 'Requires Improvement'.¹⁰ Since 2011, in addition to the four-point scale for the overall judgement of a school, inspectors have also made graded judgements (using the same four-point scale) on the:

- Effectiveness of leadership and management
- Quality of teaching, learning and assessment
- Personal development, behaviour and welfare
- Outcomes for children and learners

To disseminate their judgements, Ofsted inspectors produce a publicly-available report following each school inspection. The reports are approximately 6 to 10 pages in length depending on the type of inspection, and they attempt to capture the evidence collected by inspectors during their visit as well as displaying the overall grade for the school and the four graded sub-judgements.

During its 25 years of inspecting not just schools but colleges, apprenticeship providers, local authorities, early years provision, children's services and social care, Ofsted has inevitably generated considerable debate over its methods, procedures and impact. Its seven full-time Chief Inspectors of schools have also been met with varying degrees of acceptance by the teaching profession. Even so, rather than focusing unduly on the past, the more pertinent question is what role Ofsted should perform under the current Chief Inspector Amanda Spielman.

There are three reasons why now is the right time to ask fundamental questions about the future of Ofsted. First and foremost, Ofsted is currently rethinking its approach to inspecting

schools because the next iteration of its 'education inspection framework' (EIF) is due to be implemented from September 2019.¹¹ In January of this year, Ofsted launched a consultation on its new EIF alongside the draft documents that are likely to underpin future inspections such as the 'inspection handbook' that is being updated. This consultation signalled a change in approach from Ofsted, which is most visibly reflected in the newly proposed set of four graded sub-judgements to accompany a school's overall grade:

- Quality of education
- Behaviour and attitudes
- Personal development
- Leadership and management

Before the consultation on their new framework, Ofsted had already published its 5-year strategy that set out how it intends to deliver its mission of improving the lives of children and young people. The strategy described its fundamental guiding principle as Ofsted being "a force for improvement", which it aims to achieve by ensuring that its inspection and regulation is:

- Intelligent: "our work will be evidence-led and our evaluation tools and frameworks will be valid and reliable"
- Responsible: "our frameworks will be fair. We will seek to reduce inspection burdens and make our expectations and findings clear"
- Focused: "we will target our time and resources where they can lead directly to improvement"¹²

The second reason why the future of Ofsted is worthy of discussion is that the education system in England now looks very different from the one in which Ofsted began its first inspections in 1993. A 'quasi-market' in education has emerged in which parents have been given more opportunity to choose state schools for their children.¹³ Ofsted has acknowledged this development by stating publicly that it wants to make the results of its inspections more accessible to parents because they are its "most important education stakeholders".¹⁴ At the same time, schools have been given more autonomy, particularly in recent years, through a large proportion of schools becoming 'academies' that are free from local authority control. There has also been an increasingly strong emphasis on building a 'self-improving' school system in which teachers and leadership teams are supposedly tasked with driving school improvement rather than relying on external agencies. An essential ingredient of this self-improving system is that school leaders and teachers must feel empowered and continuously innovate to improve their practices, be it learning from each other or setting their own course towards providing a high-quality education. The

accountability system – of which Ofsted is a central tenet – needs to reflect this goal by providing accurate and timely data to all stakeholders.

The third reason to assess the future of Ofsted is that the grades they give to state schools are frequently consulted by parents. It is essential that parents are given meaningful and accessible information on schools to support their decisions about where to send their children, particularly in an increasingly autonomous education system. This means that there is considerable pressure on inspectors to deliver fair and accurate judgements. Professor Julian Le Grand from the London School of Economics has argued that for any choice-based system to function effectively, users must be properly informed about whatever choices they make.¹⁵ The new EIF accepts that parents (as well as employers and learners) “should be able to make informed choices based on the information published in inspection reports”.¹⁶

Ofsted’s survey of parents in 2017 showed that their reports were second only to ‘word of mouth from other parents’ as their main source of information,¹⁷ while a YouGov survey of parents in 2013 found that they are more likely to use Ofsted reports to help choose a school than examination results.¹⁸ In addition, a provider’s current Ofsted rating is the second most important factor for parents when choosing a school (behind ‘proximity’).¹⁹ 9 out of 10 parents knew the rating that their child’s school received at their last inspection and 67% agree that ‘Ofsted is a valuable source of information about education in your local area’.²⁰ Collectively, these figures suggest that the grades given by Ofsted remain hugely influential. As a result, the implications for schools and local communities of these grades being incorrect – either too high or low – could be very significant.

This new report starts from the presumption that it is right and proper to have an independent inspectorate for schools. The primary goal of inspections, which have been taking place in some form since 1839, is to ensure that public money is spent on the education of children and young people in the most effective way. This is precisely why an inspectorate detached from the whims of both politicians and the teaching profession is so important. Nonetheless, it is still appropriate and necessary to question how Ofsted should use its role as a regulator and inspectorate in a way that commands the confidence of the education sector, the general public and policymakers.

This report begins with an analysis of the evidence pertaining to the functions and impact of Ofsted, both in terms of its current inspection regime and the draft new EIF. After this evidence has been presented, the report will outline a new model for school inspections that gives parents the information they need to make informed decisions while retaining a relentless focus on high standards.

2. Are school inspections valid and reliable?

Ofsted's 5-year strategy stated their ambition of ensuring that their inspections are both valid and reliable.²¹ In the context of school inspections, 'reliability' refers to the degree of consistency between the judgements made by different inspectors – both of lessons and schools – whereas 'validity' refers to whether the result of an inspection or observation accurately reflects the true situation.

Evidence from Ofsted on reliability and validity

A few years after Ofsted began inspecting schools, a report by the House of Commons Education and Employment Select Committee in 1999 stated that "the difficulty of ensuring the validity, reliability and consistency of the judgements made by a large number of individual observers across a national system should not be underestimated."²² In 1996, Ofsted's own research had shown that pairs of inspectors awarded different grades after observing the same lesson in 33% of cases. The judgements were typically one grade apart (e.g. one graded a lesson '3', another '4') but in 3% of cases the difference was two grades.²³

The statistical correlation between the pairs of inspectors' judgements was 0.81,²⁴ which Ofsted described as "reassuringly high". Other witnesses were far from convinced, with one claiming that the reliability was simply "not there" while another concluded that Ofsted's judgements were "not very reliable".²⁵ Professor Carol Fitz-Gibbon argued that the sample of inspectors was poor as they were experienced inspectors, many of whom had previously worked together. She also felt that the results did not indicate a high level of agreement between inspectors as the "generally accepted" standard of reliability is 0.90 or better.²⁶ Subsequent academic research in 1998 found that two trained inspectors, independently observing the same lesson, "are likely to identify the same strengths and weaknesses in the teaching and to arrive at similar conclusions about its overall quality",²⁷ although this did not produce comparable reliability statistics.

It was almost two decades before Ofsted revisited the issue of reliability in a 2017 study that involved 24 short inspections by Her Majesty's Inspectors (HMIs) of primary schools that were rated 'Good'.²⁸ In 22 of the 24 cases, the inspectors agreed on whether the school remained 'Good' or the inspection should convert to a longer inspection to gather more evidence.²⁹ Amanda Spielman described this as "a welcome positive view of inspector consistency",³⁰ although the findings still attracted criticism. Colin Richards, a former teacher and HMI, noted that the inspectors were HMIs rather than Ofsted inspectors³¹ even though the latter outnumber the former by approximately 9 to 1 and it is increasingly

common for them to lead inspections. In addition, inspectors only had a binary choice between two options ('Good' or convert the inspection), which is not representative of most Ofsted visits.³² Russell Hobby, then general secretary of the National Association of Headteachers, pointed out that if similar rates of disagreement (8%) were found across the whole system then around 2,000 schools might have got different judgements if they had been inspected by a different team.³³ Given the considerable limitations of this research, useful as it was, it is not possible to conclude with any certainty that Ofsted inspections produce reliable results.

In terms of the validity of inspections, Ofsted accepts that "there remains little empirical evidence about the validity of inspection judgements"³⁴ and even acknowledged that the findings from their own recent reliability study "will be largely irrelevant if the components of current inspection processes are found to have little association in determining school quality".³⁵ Ofsted has never published any research on the validity of its judgements, meaning that there is no empirical support for the idea that the grades they give to schools are an accurate reflection of the quality of education being delivered.

Ofsted has occasionally referred to inspection validity in its wider publications. For example, Ofsted's annual report for 2016/17 showed that there was "a clear relationship between median Progress 8 scores³⁶ and inspection grades for the [secondary] schools inspected this year", as schools that were awarded a better Ofsted inspection rating typically had higher Progress 8 scores.³⁷ This suggests that, at least for secondary schools, Ofsted judgements may correlate with a school's performance in examinations. However, examination results cannot be used to test the validity of Ofsted judgements because these results are available to inspectors and are included as part of Ofsted's judgement process – if a school gets very good results then this will, at least in part, influence Ofsted's judgment of the school. That said, analysis carried out in 2012 identified a strong correlation between Ofsted judgements and the ratings that pupils and parents gave to a school,³⁸ suggesting that Ofsted could to some extent detect changes in teaching practice.

In November 2017, Ofsted hosted an international seminar on lesson observations. When deliberating on the criteria for conducting a lesson observation, it was reported that:

It was generally agreed among the international experts that learning is invisible and happens over a long period of time. It is not something that can be directly observed. At the very least, this is something they felt could not be measured in a valid way through observation alone, hence its exclusion in [their inspection] models.³⁹

Even though Ofsted remarked that 'learning' is just one of a range of factors they consider when assessing the quality of teaching, it is clear they are an outlier in this regard. To

illustrate the point, the word 'learning' appears well over 100 times in their current inspection handbook, including statements such as "Inspectors will consider how written and oral feedback is used to promote learning",⁴⁰ "inspectors will scrutinise pupils' work, talk to pupils about their work, gauging both their understanding and their engagement in learning"⁴¹ and "inspectors will give specific attention to the quality of learning within mainstream lessons".⁴² To compound the problem, one of the four main judgements made by Ofsted on each school under its current inspection framework is titled 'Quality of Teaching, Learning and Assessment'. For other countries to discount the entire notion of trying to observe learning at the same time as Ofsted prioritised it within their inspection regime suggests that even some of Ofsted's most high-profile functions have not been providing pupils, parents or school leaders with valid and reliable information.

When this evidence from Ofsted is taken as a whole, it is perhaps no surprise that in their latest survey of teachers, 72% agreed and only 5% disagreed with the statement that 'Inspection judgements lack consistency and are too prone to be influenced by the personal views and prejudices of individual inspectors'.⁴³

Independent evidence on reliability and validity

A key source of information on the reliability of lesson observations is the Measures of Effective Teaching (MET) project in America. This involved 3,000 teachers volunteering to open their classrooms to investigate better ways of identifying and developing effective teaching, which included assessing different methods of conducting lesson observations. When one school leader observed a 45-minute lesson, the reliability was just 0.51, and this only rose to 0.58 after they saw a second lesson from the same teacher.⁴⁴ The highest reliability score recorded anywhere in the MET project – 0.72 – could only be achieved through two 45-minute observations by a school leader, plus a 45-minute observation by a peer observer, plus three 15-minute lesson observations by three additional peer observers.⁴⁵ This compares to Ofsted's stated approach of an inspector "spending a few minutes in each [lesson]".⁴⁶ A score of 0.72 is well below the reliability benchmark of 0.90 set by Select Committee witnesses in 1999 and also fails to meet the benchmark for 'high agreement' of 0.75-0.80 cited by Ofsted in their own reliability study.⁴⁷

A separate analysis of teacher observation systems in 2012 recorded a similar pattern of reliability to the MET project. If only one observer watched 30 minutes of any given lesson, the reliability scores were around 0.3-0.5 across the three measures being recorded: the richness of the lesson material; teacher errors and imprecision; and student participation and reasoning. If two observers watched 30 minutes of a lesson, the reliability improved but only to around 0.5-0.65 for the three measures. In fact, it took four raters watching four lessons

from the same teacher to consistently reach a reliability score of approximately 0.8-0.9⁴⁸ - something that is simply not feasible with the resources available to Ofsted.

In terms of inspection validity, Education Datalab, a group of education data experts, looked at the performance data from over 800 secondary schools and found that current Ofsted grades could not identify whether a school's results were about to improve or deteriorate. In fact, judgements made by Ofsted inspectors were no better for predicting a school's future performance than just looking the past two years of GCSE performance data.⁴⁹ In addition, Education Datalab has found that, of those primary schools that were performing below the 'floor standard' for performance⁵⁰ set by the Department for Education (DfE), 69% were rated 'Good' or 'Outstanding' by Ofsted.⁵¹ For secondary schools below the 'floor standard',⁵² 35% were rated 'Good' or 'Outstanding'.⁵³ This shows that many schools performing very poorly in national assessments were still being awarded high ratings by Ofsted, raising doubts over whether inspectors are able to accurately assess what is happening within a school.

On a similar note, research by the Education Policy Institute showed that of those schools which had received 'Outstanding', 'Good' or 'Requires Improvement' at their previous inspection but whose performance had since deteriorated substantially, 47% of primaries and 33% of secondaries actually saw their Ofsted grade *improve* at their next inspection.⁵⁴ The same report noted that schools with high levels of disadvantaged children were being disproportionately awarded low ratings by Ofsted, as they identified "a systematic negative correlation between school intakes with more disadvantaged children, or more children with low prior attainment, and with favourable Ofsted judgements."⁵⁵ They concluded that "if schools were rated according to levels of pupil progress, we would expect many fewer 'outstanding' schools with very low proportions of pupils eligible for free school meals, or low prior attainment".⁵⁶ Ofsted has since disputed this, claiming that differences between progress scores for high and low disadvantage schools apply to all their judgements.⁵⁷

Another way of judging the validity of inspections is to ask those involved whether they felt that the judgements genuinely represented the situation in a school. A report by the National Audit Office (NAO) found that 84% of headteachers thought the outcome of their school's most recent inspection was fair, with headteachers who received better inspection grades being more likely to perceive the result as fair.⁵⁸ However, Ofsted's survey of teachers in 2018 found that just 62% felt their final judgement was a 'fair and accurate assessment'. 23% of teachers thought Ofsted was too harsh while 10% of teachers thought Ofsted was too lenient and missed some problems during the inspections.⁵⁹ Only 35% of teachers agreed that Ofsted is a 'trusted arbiter' of school standards.⁶⁰

The considerable time gap between Ofsted's inspections at some schools is also noteworthy because the validity of inspectors' judgements will inevitably decrease as more time passes since the last inspection. The NAO found that 1,620 schools had not been inspected for six years or more, including 296 schools that had not been inspected for 10 years or more.⁶¹ In response to the NAO report, Ofsted accepted that they couldn't tell if those schools had since "become middling, or mediocre or coasting."⁶² Amanda Spielman recently stated that out-of-date inspection reports were "the biggest bugbear for parents that came out of our focus groups"⁶³ and Ofsted are now speaking to the DfE about whether the exemption from inspections afforded to 'Outstanding' schools should be altered or potentially removed. There is little doubt that these large time gaps between inspections make it much harder for parents to gain accurate insights from Ofsted grades. This was further emphasised by Amanda Spielman telling the Public Accounts Committee in January this year that three quarters of schools previously rated as 'Outstanding' lost that rating at their next inspection and a third dropped by two grades to 'Requires Improvement'.⁶⁴

Aside from overall judgements about a school, the validity of lesson observations has been questioned. Analysis by Professor Robert Coe that drew on the data from the MET project concluded that, when comparing a judgement of teacher quality from a lesson observation against data on the 'Value Added' progress made by pupils in that class, there was approximately a 50% chance that the two would not match – a probability that rose to around 70-80% when looking at lessons judged either 'Outstanding' or 'Inadequate'.⁶⁵ Other research supports these concerns, such as the overview of teacher evaluation methods carried out by the Sutton Trust that found "even when conducted by well-trained independent evaluators, classroom observations are the least predictive method of assessing teacher effectiveness".⁶⁶

Some sources of evidence support the validity of lesson observations. The Effective Pre-school and Primary Education Project 3–11 (EPPE) found that overall the scores given in their observations for teaching quality in primary schools were a significant predictor of greater cognitive progress between the ages of 6 and 10, although these observations were carried out according to internationally developed teacher observation frameworks and by observers who had already shown a relatively high level of reliability in their judgements.⁶⁷ Meanwhile, a study that found headteachers were able to identify teachers at the extremes of effectiveness (i.e. very good or very poor) was based on headteachers evaluating teachers' work ethic, parent satisfaction and a range of other factors, which is much more rigorous than what would happen during an Ofsted inspection.⁶⁸

A separate study in 2011 ran three experiments to test whether raters could accurately judge the quality of teachers from lesson observations. The results showed that, in every case, the raters "were absolutely inaccurate, leading us to question whether educators can identify

effective teachers when they see them.”⁶⁹ The authors even went as far as concluding that “judges, no matter how experienced, are unable to identify successful teachers.”⁷⁰ That said, the research identified some aspects of particular observation tools that seemed to predict teacher effectiveness even when the raters could not identify the same trends. This led the authors to assert that “it is desirable and possible to develop a new measure that does produce accurate predictions of a teacher’s ability to raise student achievement test scores.”⁷¹ Unfortunately, the current observation protocols used by Ofsted mean that we cannot be confident their inspectors will correctly distinguish between more and less effective teachers.

In summary, the evidence on school inspections and lesson observations strongly suggests that the reliability and validity of judgements made by Ofsted are, at best, questionable.

Reliability and validity under the new inspection framework

The draft EIF released by Ofsted in January 2019 contains some constructive developments. For example, the movement away from a separate judgement for ‘teaching, learning and assessment’ should remove the unwarranted focus on trying to inspect invisible learning processes. The separate judgement for pupil behaviour is also a clear indication that Ofsted wishes to prioritise this vital area. However, even a positive change in direction of the overall inspection regime does not detract from the need to ensure that valid and reliable judgements are being produced.

In their new inspection handbook, Ofsted note that “through the use of evidence, research and inspector training, we ensure that our judgements are as valid and reliable as they can be.”⁷² Such non-committal phrasing suggests Ofsted recognises that they cannot necessarily achieve a high level of validity and reliability in their inspections. Furthermore, it is a notable retreat from the claim made by Ofsted in their 5-year strategy (published in 2017) that “our evaluation tools and frameworks will be valid and reliable”.⁷³ Alongside the draft new framework and handbook, Ofsted published an overview of the research evidence that underpins their proposals and “draws on a range of sources, including both our own research programme and a review of existing evidence bases”.⁷⁴ The term ‘reliability’ does not appear at all in the entire 69-page document, and the term validity only appears once (in reference to accountability measures).

In December 2018, Ofsted published Phase 3 of their research into how they will judge the ‘Quality of education’ in a school, with a particular focus on curriculum implementation and impact.⁷⁵ This involved trialling 25 potential curriculum indicators across 64 schools, with each indicator being scored on a 1–5 scale (with 5 equating to a ‘strong curriculum design’

and 1 a 'weak design'). The study found that, when inspectors followed the methodology in the research for evaluating a school's curriculum, there was often a strong relationship between the first-hand evidence they collected (e.g. examples of pupil's work) and the quantitative scores that were assigned to each indicator.⁷⁶ However, inspectors sometimes had a tendency to record what school leaders 'said' or what processes they had in place rather than looking at the depth, coverage and impact of curriculum.⁷⁷ The evidence collected for different curriculum indicators also varied depending on how much a subject leader in a school had considered it when designing their curriculum.⁷⁸ This demonstrates how easy it is for subjective assessments to creep into these new inspection tools.

Arguably, such issues are to be expected when creating a new approach to assessing a school's curriculum, which is why Ofsted acknowledged that "inspector training is paramount [and] quality assurance is required to ensure consistency".⁷⁹ Nevertheless, Ofsted was confident that "the findings provide a degree of confidence that our plans to look beyond data and assess the broader quality of education are achievable and necessary."⁸⁰ A reasonable level of confidence in the validity (accuracy) of their curriculum judgements is encouraging, but this crucial research study did not even test the reliability of judgements made by different inspectors. Although Ofsted claimed that their study suggested the 1-5 scoring system was being applied "in a reliable way",⁸¹ they admitted that they "were not able to test for interrater reliability between inspectors".⁸² This means that Ofsted does not know whether their new curriculum indicators will lead to consistent judgements between inspectors, even though the new EIF is scheduled to begin in just a few months.

Although inspecting 'learning' has been de-emphasised with the removal of the separate judgement on 'teaching, learning and assessment', Ofsted intends to continue using lesson observations to provide evidence for their judgement on the 'Quality of education'. The observations will then be triangulated with evidence from other sources (e.g. scrutinising pupils' work).⁸³ Although Ofsted's recent curriculum research did not include lesson observations within the triangulation process, they were clear that observations will "carry considerable importance for helping to assess the quality of education in the new framework".⁸⁴ The evidence presented throughout this chapter on the low reliability and validity of lesson observations applies just as much to curriculum-based observations as they do to previous judgements on teaching quality. It is therefore hard to have confidence in these plans when there is no evidence from Ofsted or elsewhere to suggest that any reformed observation protocols will be either reliable or valid.

The draft EIF also places a much greater emphasis on 'work scrutiny' than the previous framework. In future, inspectors "will scrutinise pupils' books and other work across a faculty, department, subject, key stage or year group and aggregate insights to provide part

of the evidence for an overall view of the quality of education”.⁸⁵ The aim of this exercise is “to evaluate pupils’ progression through the curriculum [as] work scrutiny will show whether pupils know more and can do more, and whether the knowledge and skills they have learned are well sequenced and have developed incrementally.”⁸⁶

The prospect of exercise books being scrutinised in this manner across thousands of schools raises crucial questions about reliability and validity yet again. Ofsted has not published any evidence to show that inspectors can make consistent judgements on pupils’ work across the full range of National Curriculum subjects, GCSEs and A-levels and much more besides (particularly when inspectors may not be experts in the subjects or key stages that they are inspecting). It remains unclear how the development of pupils’ skills and knowledge can be satisfactorily demonstrated by teachers and leaders or over what timescale this will be judged for schemes of work that may cover two or three years of teaching. Furthermore, when assessing the overall quality of education at a school, Ofsted will include “discussions with pupils about what they have remembered about the content they have studied” alongside lesson observations and book scrutiny,⁸⁷ which is likely to generate even more problems in terms of making consistent judgements.

The new inspection handbook says that inspectors will “ensure that the samples of pupils they choose [for observations, work scrutiny and discussions] are sufficient to allow them to reach a valid and reliable judgement on the quality of education offered by the school overall”.⁸⁸ When this relates to schools containing hundreds of pupils spread across perhaps 10-15 subjects, the imprecision of this goal becomes apparent. In addition, it serves to highlight the lack of research into the minimum amount of information required by inspectors to form valid and reliable judgements, which one might have expected to be a priority for any new inspection framework.

For all the benefits that an increased focus on areas such as the school curriculum may bring, the absence of evidence that these new inspection areas will be judged in an accurate and consistent manner means that the new EIF does not appear to represent a step forward in terms of providing parents with fair and accurate information. Ofsted has publicly committed to operating as an evidence-led organisation that makes valid and reliable judgements, but this commitment is not sufficiently evident in their new plans for inspecting schools.

3. Does Ofsted look at the right things?

Both Ofsted's current inspection framework and draft new EIF are clear that, when coming to their judgements, inspectors will "evaluate objectively", "be impartial" and "base all evaluations on clear and robust [strong] evidence".^{89 90} It is therefore noteworthy that several important areas within the inspection process do not appear to promote an objective and evidence-based outcome.

The current inspection handbook says that "it is up to schools themselves to determine their practices and for leadership teams to justify these on their own merits".⁹¹ This suggests that, rather than looking for evidence-based practices, Ofsted are judging the level of compliance shown by staff to whatever processes or procedures that a leadership team dictates. For example, Ofsted state that the marking and feedback given to pupils is "for the school to decide through its assessment policy" and that "marking and feedback should be consistent with that policy".⁹² However, later in the handbook it says that inspectors will pay "...particular attention to ...how well teachers' feedback, written and oral, is used by pupils to improve their knowledge, understanding and skills".⁹³ For Ofsted to state that on the one hand it is for schools to decide how to assess their pupils, while on the other hand stating that inspectors will be the arbiters of whether an assessment policy is correct, neatly illustrates how subjective judgements from inspectors can still overrule school leadership teams even in the absence of an evidence-based case for change.

Vague terminology does not help matters either. For a school to be judged 'Outstanding' for teaching, learning and assessment, teachers must "provide pupils with incisive feedback".⁹⁴ At no point in the current handbook does Ofsted explain what 'incisive feedback' means in practice or what differentiates it from apparently non-incisive feedback. Similarly, numerous statements elsewhere in the handbook lack supporting evidence. For an 'Outstanding' judgement the handbook requires that "pupils love the challenge of learning and are resilient to failure"⁹⁵ yet this too is a purely subjective judgement by inspectors that will be based on very limited observations of, and interactions with, pupils.

The new inspection handbook seems to fare little better than its predecessor. For example, when judging the 'Quality of education', teachers will now have to "provide clear, direct feedback"⁹⁶ without any explanation of what will be considered unclear or indirect feedback. What's more, the new handbook states that the most important factors in how the curriculum is assessed include "teachers use assessment to check pupils' understanding in order to inform teaching" and "teachers use assessment to help pupils embed and use knowledge fluently". Needless to say, the scrutiny of pupils' work discussed in the previous

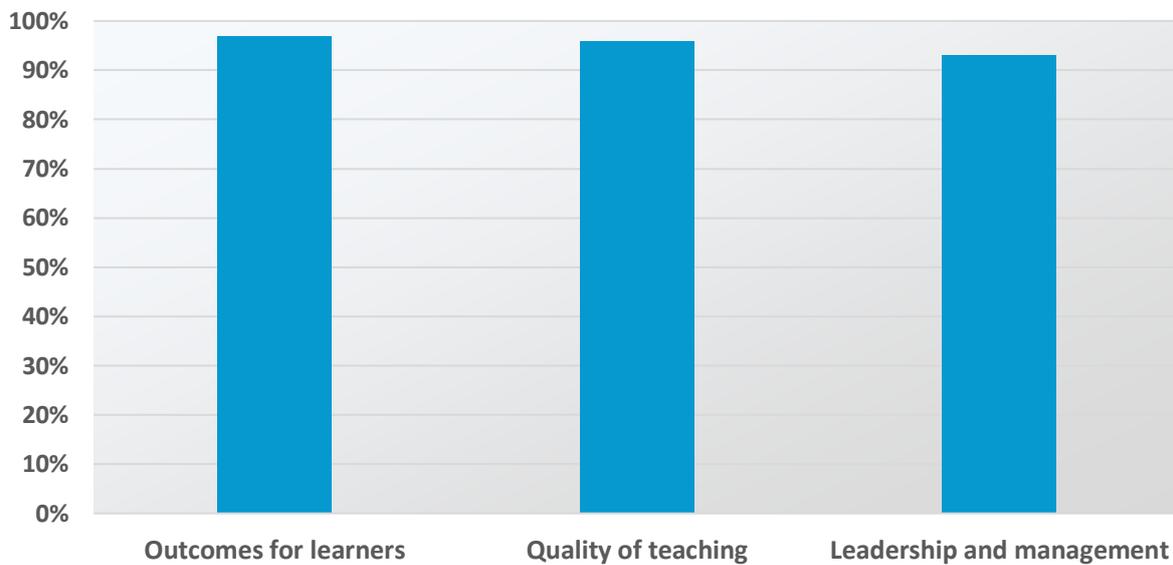
chapter will be a key source of evidence on these matters, yet the scope for subjective judgements from inspectors remains significant.

The use of vague terminology is evident in the new handbook as well. To be judged 'Good' or 'Outstanding' for the 'Quality of education', school leaders must not use pupil assessments "in a way that creates unnecessary burdens on staff or pupils",⁹⁷ yet there is no clarity on how the 'necessity' of these burdens will be assessed. Moreover, when judging the 'Personal development' of pupils, a school can only be rated 'Outstanding' if it "consistently goes 'the extra mile' to promote the personal development of pupils, so that they have access to a rich set of experiences" and "the way the school goes about developing pupils' character is exemplary and is worthy of being shared with others."⁹⁸ Definitions of 'going the extra mile', 'rich experiences' and 'pupil's character' were notable by their absence. Expecting a school's curriculum and wider work to "support pupils to develop resilience, confidence and independence and lead a healthy and active lifestyle, helping them to know how to keep physically and mentally healthy"⁹⁹ is another worthy aspiration but a problematic area for inspections, especially when mental health is a very broad and complicated issue.

The fact that different Ofsted judgements appear to significantly overlap with each other raises further questions about their value to parents and other stakeholders. Figure 1 (overleaf) shows the correlations between the three judgements currently made by Ofsted related to academic performance at a school.¹⁰⁰ The overall grade awarded to a school is the same as the subgrade for 'Outcomes for children and learners' in 97% of cases as well as the subgrade for 'Quality of teaching, learning and assessment' in 96% of cases. Similarly, and perhaps unsurprisingly, the subgrades for 'Quality of teaching, learning and assessment' and 'Outcomes for children and learners' match for 97% of schools.

On the one hand, it could be argued that the high correlations between these judgements show that the inspection process is able to identify good and bad practice. On the other hand, it demonstrates that a judgement based purely on 'outcomes' would produce remarkably similar judgements to Ofsted's entire inspection system but without the adverse consequences on teachers and leaders (see next chapter). It also suggests that process-based judgements made by inspectors, and all the associated confusion and misunderstandings, may be largely redundant if the data on outcomes can capture the effectiveness of a school and communicate this to stakeholders.

Figure 1: level of agreement between schools' overall grades and their subgrades



Not only is Ofsted seemingly paying too much attention to some areas, there are other aspects of school life that have previously not received much attention at all. To promote a broader view of what constitutes an effective education, Ofsted launched a curriculum review in 2017 to investigate what is happening inside hundreds of schools and assess whether inspections should place more emphasis on curriculum design and implementation. The initial findings of the review showed a great deal of ambiguity about what 'curriculum' meant, and this lack of shared understanding had contributed to primary schools narrowing their curriculum to focus on preparing for Key Stage 2 tests, schools entering some secondary pupils for inappropriate qualifications and the Key Stage 3 curriculum being reduced in some instances to give more time for GCSE courses.¹⁰¹ The greatly increased focus on the curriculum under the new EIF is therefore a welcome shift in emphasis from Ofsted.

There are other areas of school life that have not attracted much attention from Ofsted in the past. Extra-curricular activities are mentioned as part of one criterion within the 'Quality of teaching, learning and assessment' judgement in the current inspection handbook¹⁰² while careers advice and guidance only appears once within the 'Personal development, behaviour and welfare' judgement.¹⁰³ There have been numerous calls for these issues to be given greater priority, exemplified by the Social Mobility Commission recommending that no school should be judged 'Outstanding' unless it offers excellent extracurricular activities, careers advice and work experience.¹⁰⁴ It is encouraging to see the new inspection handbook go some way to addressing these concerns. Schools will now be required to ensure that "participation in extra-curricular activities is consistently very high, including among those

from disadvantaged backgrounds”¹⁰⁵ for them to be judged as ‘Outstanding’. Likewise, careers guidance is in a marginally stronger position within the new handbook because a judgement of ‘Good’ or ‘Outstanding’ will require a secondary school to provide “unbiased information to all about potential next steps and high-quality careers guidance and opportunities for encounters with the world of work.”¹⁰⁶

Ofsted also conducts surveys as part of their inspections and will continue to do so under the new EIF. Student and staff surveys are intended to complement the information collected by the inspection team. If designed correctly, survey data can provide a valuable insight into school life, particularly when research (including the MET project in America) has demonstrated that student surveys can be a better predictor of student achievement than lesson observations, teacher or headteacher effectiveness ratings.¹⁰⁷ Nevertheless, current Ofsted surveys appear to spread the potential for new insight too thinly. The student survey includes questions on student motivation, the behaviour of peers, safeguarding, pupils’ physical health, equality and diversity and independent learning in the space of just 20 questions. Likewise, the staff survey asks about pupil behaviour, bullying policies, professional development, workload concerns, pupil progress and support from senior management across 23 questions.¹⁰⁸ By including such a wide range of issues in a short survey of this nature, Ofsted could be missing out on a powerful driver of behaviour among school leaders, teachers and pupils.

4. What impact does Ofsted have on teachers and leaders?

For several years, Ofsted has published what they call ‘myth busters’ that seek to highlight specific practices that are not required by Ofsted but are still found in some schools.¹⁰⁹ For example, the myth buster states (as does the current and draft new inspection handbook) that “Ofsted does not expect to see any specific frequency, type or volume of marking and feedback”.¹¹⁰ However, inspectors have continued to make comments about precisely these areas, as demonstrated by the following remarks from inspectors in recently published Ofsted reports:

- “...teachers do not always give detailed enough written feedback to help pupils reach the highest levels in their work, or make sure that pupils respond, in writing, to their comments.”
- “Marking across the school is excellent in all lessons and subjects. The very best examples refer to pupils’ individual targets”
- “...teachers do not always give pupils written pointers as to how to improve their work.”
- “Students respond energetically to teachers’ very precise feedback in the vast majority of circumstances. However this is not always insisted upon by teachers.”
- “Teachers mark work regularly. Marking is at its best when teachers tell students how to further improve their work and then provide the opportunity for students to act on their advice.”¹¹¹

A survey by the Association of School and College Leaders showed that schools are still being asked for evidence by inspectors that Ofsted says they do not have to provide. For example, despite Ofsted not requiring schools to predict the attainment or progress scores for their pupils, 62% of leaders were asked to predict pupil attainment and 47% were asked for predicted progress scores.¹¹² These mixed messages about marking, pupil data and other issues are evidently causing problems for leaders and teachers without adding any value to our shared understanding of the quality of any given school.

Inevitably, the influence of Ofsted can affect staff workload. In 2014, the DfE ran a ‘Workload Challenge’ that generated almost 44,000 responses.¹¹³ The single most commonly cited cause of the workload burden faced by teachers was ‘accountability / perceived pressures of Ofsted’.¹¹⁴ A similar phenomenon was documented in a 2013 survey, which found that over a third of headteachers and deputy headteachers thought ‘unnecessary or bureaucratic tasks’ had increased. For headteachers, ‘Ofsted changes’ and ‘preparation for an Ofsted inspection’ were two of the main reasons quoted.¹¹⁵ An international comparison

in 2013 showed that the top issue named by headteachers in England as creating a barrier to their effectiveness was ‘government regulation and policy’ (cited by 85% of academy heads and 95% of maintained school heads),¹¹⁶ although this broad topic may have encompassed matters outside of Ofsted’s remit.

Aside from workload pressures, Ofsted’s 2017 teacher survey showed that 76% thought school inspections would be “highly stressful for everyone”.¹¹⁷ Curiously, Ofsted chose not to ask this same question again in 2018. Instead, the most recent version of the survey found that over half of teachers felt an inspection would “mean a huge amount of unnecessary extra work” with only 16% of teachers agreeing that they would “simply carry on doing your job as normal”.¹¹⁸ Most concerningly, 83% of teachers agreed that “Ofsted inspections introduce unacceptable levels of burden into the system”,¹¹⁹ although the NAO survey found that 54% of headteachers thought that the burdens were proportionate in order for Ofsted to form a reliable judgement.¹²⁰

A report in 2014 found that over 70% of the teachers admitted to significantly changing their practice when Ofsted inspectors came to their school, usually by teaching a lesson in a different way to what they would describe as their normal practice. The authors recognised that this was “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.”¹²¹ Inevitably, teachers changing their practice to please Ofsted reduces the validity and reliability of any observations by inspectors and means that any feedback received during or after the inspection is unlikely to be meaningful.

Academic evidence has borne out the impact of Ofsted on school leaders. A 2017 study found that “the influence of the inspection agenda was strong [...as] policy decisions were often being made to conform to Ofsted’s expectations and the influence on leadership and management was clearly apparent”. The researchers also found that “schools to some extent performed ‘the good school’ for inspections” to gain as high a rating as possible.¹²² A separate study in the same year, which analysed the consequences of school inspections across seven countries, found “a clear association between increasing pressure in a school inspection system and an increase in the narrowing of the curriculum and instructional strategies in the school”.¹²³ Other consequences of inspections highlighted by this research, albeit on a smaller scale, included the “discouragement of experimenting with new teaching methods”.¹²⁴

A study by Professor Toby Greany and Rob Higham demonstrated the practical implications of this perceived need to conform to Ofsted expectations. Their survey of school leaders showed that the vast majority of respondents (77% of primary and 83% of secondary schools) agreed or strongly agreed that ‘making sure my school does well in Ofsted

inspections is one of my top priorities'.¹²⁵ As a result, "school leaders regularly felt incentivised to prioritise the interests of the school over the interests of particular groups of, usually more vulnerable, children."¹²⁶ In addition, "all of the schools we visited recognised the need to perform for Ofsted and it was notable how language and concepts from Ofsted had been widely internalised into school practices and ways of thinking".¹²⁷

At the launch of the new EIF, Amanda Spielman was keen to emphasise that "while we know that any kind of accountability necessarily involves some irreducible workload, we have to do what we can to make sure that inspection adds no more burden than it must."¹²⁸ In line with this sentiment, Ofsted has made some notable changes. For instance, inspectors will no longer look at internal progress and attainment data of current pupils or students in the hope that, as Amanda herself said, "we will once and for all bust the myth that data should be created for Ofsted".¹²⁹ This policy shift will force leadership teams to evaluate the frequency and density of the data they collect on their students, which in turn could feasibly result in a reduction of data collection and analysis for staff.

More surprisingly, though, Ofsted has decided that inspectors will formally judge the workload faced by teachers and potentially mark a school down if staff workload is deemed too high. To deliver this, the 'Leadership and management' judgement in the new inspection handbook includes an assessment of "the extent to which leaders take into account the workload and well-being of their staff in order to deliver a high-quality education".¹³⁰ Under this new system, for a school to be judged as 'Good' on 'Leadership and management', "leaders must engage with their staff and [...] take account of the main pressures on them", and they must also be "realistic and constructive in the way they manage staff including their workload."¹³¹ To be judged 'Outstanding', leaders must further ensure that "highly effective and meaningful engagement takes place with staff at all levels and that issues are identified [and] when issues are identified – in particular about workload – they are consistently dealt with appropriately and quickly."¹³²

For an inspectorate that is widely recognised by teachers and leaders as being one of the root causes of numerous workload pressures to now be judging school leaders on the workload they generate for their staff does not seem logical. It is not obvious how an inspector is supposed to judge staff workload in a valid and reliable way when the context in which schools operate (in a geographical, educational and financial sense) will vary enormously between inspections. Even if Ofsted no longer pays attention to internal performance data, the emphasis on book scrutiny in the new inspection handbook could easily replace one significant source of workload with another equally absorbing and burdensome set of tasks. Michael Tidd, headteacher at Medmerry Primary School and a regular commentator on school assessment, noted that "it's a big leap to say that one Ofsted inspector's interpretation of progress from books is more valid than any school's own data. I'm not so sure they're

right.”¹³³ Stephen Tierney, Chief Executive of a Multi-Academy Trust and chair of the Headteachers’ Roundtable, has said that “the audit culture that drives the workload monster will switch from data to exercise books, as people feel the pressure to provide evidence of learning that occurred when an inspector was not present [...which is] at least 99 per cent of an academic year.”¹³⁴ Such a scenario does not appear to represent definitive progress in tackling teacher workload.

Ultimately, the considerable pressures generated by Ofsted inspections plus the questionable behaviours that the inspection system incentivises do little to help parents and politicians understand which schools are performing well and which are not. At the same time, the quality of education available at any given school could easily be diminished if teachers and leaders are more worried about what is best for Ofsted than what is best for their pupils. The new EIF has not done enough to allay these concerns.

5. Do Ofsted inspections lead to school improvement?

Despite its prominent role in judging school standards, Ofsted does not directly intervene to improve schools, nor does it decide what action should be taken after it has completed an inspection. This makes it hard to empirically assess whether any improvements seen in a school after an inspection were caused by the inspection itself.

Although some schools do improve each year, recent data from Ofsted showed that 33% of primary schools and 58% of secondary schools rated 'Requires Improvement' did not improve and a higher proportion of schools rated 'Requires Improvement' (11%) declined to 'Inadequate' compared to previous years.¹³⁵ Of those 58% secondary schools rated 'Requires Improvement' that did not improve, over a third of them have been judged 'Requires Improvement' / 'Satisfactory' or 'Inadequate' in every inspection they have had since 2005 (despite some converting to become academies).¹³⁶ This highlights how some schools are stubbornly resistant to improvement. In addition, of the 3,500 short (1-day) inspections of schools judged to be 'Good' in their previous inspection, many schools stayed at 'Good' or improved to 'Outstanding' but around 17% of primary schools and 23% of secondary schools dropped to 'Requires Improvement' or even 'Inadequate'.¹³⁷

On the matter of Ofsted and school improvement, the NAO found that:

*"Ofsted does not know whether its school inspections are having the intended impact: to raise the standards of education and improve the quality of children's and young people's lives. It has not had clear performance indicators or targets to track progress towards these high-level aims. Its performance measures have instead focused mainly on activity and processes."*¹³⁸

The report noted that "Ofsted set few targets to measure performance against its 2016 strategic plan, and has provided limited information to allow others to assess its progress."¹³⁹ The NAO also discovered that seven of the nine measures created by Ofsted to capture its quality, efficiency and effectiveness did not even have any associated targets (and three of them were not publicly reported in any case). Ofsted's new 2017-2022 strategy has addressed some of these points by establishing several key performance indicators and targets such as the percentage of parents who consider that Ofsted is a valuable source of information and the percentage of teachers who see Ofsted as a force for improvement.

The views of headteachers and teachers on the impact of Ofsted inspections on school performance are decidedly mixed. In the NAO survey, only 44% of headteachers said that

inspections had led to improvements in the school¹⁴⁰ and Ofsted's survey of teachers in 2018 found that just 31% shared this opinion.¹⁴¹ There were slightly more positive views about Ofsted's feedback to schools, as 71% of headteachers agreed that inspectors provided useful feedback both during and at the end of the inspection.¹⁴²

A review in 2014 by Karen Jones and Peter Tymms at Durham University identified several factors that could potentially generate improvement through inspections, such as giving schools feedback about strengths and weaknesses, the desire among schools to avoid sanctions for poor performance and the pressure generated by publishing inspection reports.¹⁴³ Their review of previous studies on the relationship between inspections and school improvement found that "the findings have been mixed."¹⁴⁴ This echoes a review for the Dutch Inspectorate of Education, which commented that "much of the research is based on data collected many years ago and shows slight or no effect".¹⁴⁵

With regard to other evidence from academics, Hussain (2012) found that a school failing an inspection in England led to increased gains in Key Stage 2 tests of around 0.1 of a standard deviation while Allen and Burgess (2012) found that schools which just fail their inspection saw an improvement in test scores two and three years after inspection (mostly driven by middle and higher-ability pupils). Luginbuhl et al. (2009) assessed the impact of Dutch inspections and found small positive effects and no effects using two different designs respectively.¹⁴⁶ The review by Jones and Tymms concluded that "there is a lack of evidence from strong research designs to assess the impact of inspections and the assumption that there is a causal link between inspections and school improvement cannot be clearly supported from the literature."¹⁴⁷

Looking across the range of available evidence, there does not appear to be a compelling case for concluding that Ofsted inspections necessarily lead to improvements. The data suggest that some schools seem to improve following a visit from inspectors while others stay the same or get worse. Comparisons over time are also potentially distorted by poorly-performing schools being closed (meaning they are no longer captured in the data) or by the data not accounting for changes in the way that Ofsted inspects schools. In short, Ofsted's role in driving school improvement appears to be a tenuous one and there is no reason to think that the new EIF will alter this situation.

6. Recommendations

In an increasingly autonomous education system, the provision of accurate and meaningful information to parents becomes even more paramount. We should therefore be concerned if parents, teachers and school leaders are being given – and potentially acting upon – misleading information because an education system can only become ‘self-improving’ if it is built on trusted and timely evidence. School leaders and teachers must also feel empowered and strive to continuously innovate and improve their practice, be it learning from each other or setting their own course towards providing a high-quality education. At the same time, we must avoid placing unnecessary burdens on teachers and leaders, particularly in the context of on-going issues with workload pressures as well as teacher recruitment and retention in general.

The evidence presented in this report has shown that, to achieve these goals, a new approach is needed in order to create a more appropriate way of holding schools to account that draws on the benefits of having an independent inspectorate but without generating as many drawbacks.

Parents as the most important stakeholder

The theme of parents from less advantaged backgrounds making different school choices is common in the academic literature. A 2015 paper by Professor Simon Burgess and colleagues found that “[parents] from lower socio-economic quintiles are more likely to choose a less academic school.”¹⁴⁸ Another research paper in the same year showed, using data from the Millennium Cohort Study, that parents with lower levels of income and qualifications are significantly less likely to give weight to academic performance than more affluent and educated parents.¹⁴⁹

The reasons behind this variation in school choices have been the subject of debate. A review by the OECD observed that parents from less advantaged backgrounds may “have access to less information, or lower quality information” when selecting a school.¹⁵⁰ This could be because “information acquisition has very high costs, especially for parents who lack the needed social capital, the resources, the time, the connections or the cultural resources to effectively choose.”¹⁵¹ To address this, the OECD recommended that “public institutions must take into account the limitation that certain parents have in making choices, by minimizing the cost of information acquisition.”¹⁵² It was proposed that “targeted and active parent information programmes” may be required to get more families to actively choose schools.¹⁵³

Researchers in America ran an experiment in a school district in North Carolina to examine whether transparent and easily accessible information on school-level academic performance could change parents' choices. The results were striking, as "providing parents with direct information on school test scores resulted in significantly more parents choosing higher scoring schools for their children".¹⁵⁴ The authors concluded that "simplified information on school academic achievement may have a significant impact on the efficacy of school choice plans for disadvantaged families."¹⁵⁵ Meanwhile, a study of school choice in Pakistan found that not only did the provision of performance data mean that the "perception of quality [among parents] became better aligned with school test scores", the actual test scores themselves rose by 0.11 standard deviations.¹⁵⁶

Whether these same benefits would appear in England is hard to ascertain, given that school performance data are available online. The fact that the DfE publish all their data on schools might, ironically, make it more difficult for many families to digest this information in the absence of any guidance or filtering. The notion of simplified and relevant information being given to parents is something that Ofsted themselves have encountered. When Ofsted asked parents what they felt would make their reports more useful, a '1-page summary of key findings / traffic light system' was the most commonly mentioned improvement, with 'details about how a school compares to other schools locally' and 'including views from parents of children who attend the school' also proving highly popular.¹⁵⁷ Furthermore, Ofsted's parent survey identified several topics that parents felt were not covered at all, or in enough detail in inspections. These included the happiness and mental health of pupils, pastoral care, parental experiences and staff wellbeing and engagement.¹⁵⁸

With this feedback from parents in mind, plus Ofsted's recognition that parents are its "most important education stakeholders",¹⁵⁹ the need to provide better support to parents when making choices about schools should form the basis of Ofsted's new role in our accountability system. What's more, the evidence suggests that providing more effective support to parents could, if done well, support efforts to improve social mobility by helping families from less advantaged backgrounds access better schools.

PART 1: Creating a new role for Ofsted

RECOMMENDATION 1

Ofsted should not provide an overall rating or grade for schools due to a lack of reliability and validity in their judgements.

This report has explained in detail why schools are spending too much time and energy worrying about the subjective judgements of inspectors, when they should be focused on delivering a high standard of education to their pupils. That is not to say Ofsted is no longer needed, but rather that their existing role is not tenable in a self-improving school system.

The evidence presented in this report has highlighted a wide range of negative consequences that Ofsted grades have on headteachers and their staff. Ofsted inspections distort the behaviour of leadership teams and classroom teachers, who understandably seek to impress inspectors instead of relentlessly focusing on improving standards. These two goals do not necessarily have to be mutually exclusive, but the evidence suggests that they are too often misaligned. The lack of reliability and validity in much of Ofsted's work makes it extremely hard to have confidence in any overall judgement that they make. Moreover, the lengthy gaps between inspections mean that the overall grades assigned by Ofsted can be years out of date and may therefore be of little use to parents as well as potentially being inaccurate. These uncertainties are compounded when large numbers of poorly-performing schools are still being awarded the top ratings by inspectors.

The most common rebuttal against scrapping Ofsted grades comes from the claim that parents use the grades to help them choose schools so they cannot be removed. Amanda Spielman said last year that “[parents] like the clarity of four grades in helping them to make informed choices, and as a marker of how well their child’s school is performing”.¹⁶⁰ There are two reasons to reject this position. First, parents will undoubtedly have got used to having access to Ofsted grades in recent years, which is reflected in the survey data on how many parents consult them. Even so, if they are presented with the stark evidence described in this report about the potential lack of accuracy and consistency of these grades then it is likely that parents’ enthusiasm for them will wane.

Second, some natural experiments in the US suggest that parents are in fact not keen on the notion of grading schools when given the choice. In the District of Columbia in Washington DC, the new school report cards that were introduced in December 2018 include a ‘STAR (School Transparency and Reporting) Rating’. This STAR Rating, displayed at the very top of each school’s card, claims to measure each school’s performance and then allocates a rating from one to five stars, with five being the highest.¹⁶¹ During the consultation phase for

these new report cards, hundreds of respondents were asked to identify the three most important pieces of information about a school that they wanted to see on the new report cards. In the online consultation survey, the 1-5 STAR Rating was ranked 20th out of the 26 potential pieces of information to be included on the report cards, while in the face-to-face community engagement phase of the consultation the STAR Rating was ranked 12th out of 13 pieces of information.¹⁶²

This research from the US indicates that parents do not consider overall rankings or grades to be important when they are given the choice of whether or not to include them in an accountability system. On the contrary, many parents in Washington were actively critical of the ratings during the consultation. One said that “inevitably ...some of the poorest schools that have large numbers of students from poor families are going to get a ‘one’ rating” and that parents will not want to send their kids to a school with a one-star rating, creating a “vicious cycle of failure for high-poverty schools”. Another parent said that proponents of the ratings “would argue that it makes it easy for parents to say which is a good school and which is not [but] unfortunately, it’s not accurate”, while another added that if the report cards “gave you a wide palette of what to look at about a school, that makes a lot of sense, but I’m particularly concerned about the dominance of this star rating.”¹⁶³ Such observations from parents lend further weight to the removal of overall grades because parents are well aware of the consequences for the children and communities who end up with poorly-rated schools (irrespective of whether the grades are accurate or not). This illustrates how grading schools promotes almost the complete opposite of what a self-improving system should be trying to accomplish.

As noted in the introduction to this report, inspectors have been required to judge “the quality of the education provided by the school” ever since the 1992 Education Act created Ofsted. Even though the grading scale has changed several times in the intervening period, this requirement remains in place today. The *Education Act 2011* replaced the wording from Section 5 of the *Education Act 2005* to create the current setup of an overall grade with four sub-judgements:

“(5) It is the general duty of the Chief Inspector, when conducting an inspection under this section, to report on the quality of education provided in the school.

(5A) The Chief Inspector’s report under subsection (5) must in particular cover —

- (a) the achievement of pupils at the school;*
- (b) the quality of teaching in the school;*
- (c) the quality of the leadership in and management of the school;*
- (d) the behaviour and safety of pupils at the school.* ¹⁶⁴

This report therefore recommends that the clause describing the general duty to report on the quality of education in a school should be removed from the *Education Act 2005*, as this would relieve Ofsted of having to provide an overall grade.

RECOMMENDATION 2

Ofsted should not conduct any observations of lessons or scrutinise pupils' work until their processes and procedures have been rigorously tested in terms of both reliability and validity.

When Ofsted was created over 25 years ago, league tables were a recent innovation, so school inspectors were in effect the only independent source of information about the quality of teaching and learning within schools. However, the absence of reliability and validity in Ofsted inspections, plus the considerable advancements in the depth and breadth of school performance data, means that the notion of inspectors walking into classrooms up and down the country to pass judgement on what is happening in lessons has outlived its useful purpose.

Inspections place enormous pressure on teachers and leaders across a whole host of areas such as how and when to provide feedback, how to structure lessons, what lessons must include and how curricula and lesson content should be delivered. While the planned removal of the separate grade for 'Quality of teaching, learning and assessment' in the new EIF is a positive step, the new framework is set to draw heavily on lesson observations, work scrutiny and other limited interactions with pupils and staff. As described earlier in this report, the absence of research evidence on the reliability and validity of these new processes is alarming, particularly when they will be rolled out in schools across the country in just a few months' time.

Ofsted cannot expect to gain the trust and respect of teachers or leaders when they demand that schools provide evidence of the impact they are having yet at the same time do not submit themselves to the same level of scrutiny. Irrespective of exactly how the new EIF is implemented, it will almost certainly require inspectors to continue judging what is happening in classrooms within the space of just a few minutes in each lesson – an approach that clearly lacks empirical support. Until a suitable range of well-designed research studies in support of their inspection processes are conducted and published, Ofsted inspections should not include any such classroom-based data collection.

It is perfectly possible that the new EIF will promote high-quality educational practices in many instances, but no assumptions should be made on this matter. Instead, the proposed

inspection procedures should undergo a sufficient period of testing and refinement in full view of the education profession, academics and government ministers. This would be the most responsible course of action for an organisation that claims to be evidence-led and driven by valid and reliable inspection tools.

RECOMMENDATION 3

Ofsted inspections should be entirely focused on observable aspects of school life that are not captured by performance data. This includes pupil behaviour and safety, the curriculum, careers advice and extra-curricular activities.

Many aspects of school life are best understood when they are witnessed first-hand. Performance data in terms of examination results is a key indicator of how well a school is educating its children, yet examinations only represent a single aspect of school life. A team of trained inspectors can assess other important facets of a rounded education, even if they are no longer required to judge areas such as the quality of teaching.

Last year, Amanda Spielman said she wants to make sure that Ofsted focuses on “the essence of what performance tables cannot capture”, adding that:

“The bottom line is that we must make sure that we, as an inspectorate, complement rather than intensify performance data, because our curriculum research and a vast amount of sector feedback have told us that a focus on performance data is coming at the expense of what is taught in schools.” ¹⁶⁵

This report has highlighted the low reliability and validity of inspecting less tangible aspects of education such as ‘learning’, which is why other countries do not include such matters in their inspection models. That said, Ofsted’s recent international seminar showed that more tangible elements of school life typically form a core part of the inspection process elsewhere. For example, it was noted that “aspects of classroom management [...] and student behaviour and attitudes were routinely included in the models presented at the seminar.”¹⁶⁶ In terms of specific models, CLASS (one of the protocols included in the MET project in America) uses ‘behaviour management’ as one of its main dimensions of classroom practice¹⁶⁷ and ICALT (used in the national inspection system in the Netherlands) includes ‘safe learning climate’ and ‘classroom management’ among its list of six observable domains.¹⁶⁸ This switch away from unobservable characteristics (e.g. learning) towards observable ones (e.g. behaviour) is vital for Ofsted, not least because ‘visible’ aspects of a school will be more likely to lead to reliable and valid judgements by inspectors as they lend themselves much better to objective observations.

As a result, this report endorses several changes found in the new EIF. This includes the decision to provide a separate inspection judgement on 'behaviour and attitudes' as this will give it greater prominence among school leadership teams. That said, the EIF proposes that safeguarding is judged within the 'Leadership and management' section of an inspection, whereas it would be more logical to include it alongside pupil behaviour. It is therefore recommended that a new judgement is made on 'Behaviour and pupil safety' (echoing the *Education Act 2005*). This is designed to prevent a situation occurring where, for example, pupil behaviour within lessons appears constructive (which would lead to a good judgement on 'Behaviour and attitudes') but school leaders are failing to investigate safeguarding allegations against staff (which would only affect the 'Leadership and management' judgement). For the sake of simplicity, parents should be given a single rating of how well pupils are behaving and how well a school is protecting them from harm, be it bullying, safeguarding concerns or related issues.

As discussed throughout this report, the greater focus on the school curriculum is a welcome development in the new EIF. However, the on-going concerns around the reliability and validity of such judgements as well as the potential impact on workload means that a different approach to this topic is required. The indicators recently trialled by Ofsted regarding the intent, implementation and impact of a school's curriculum represent a useful starting point, but the way that the final set of indicators are used should be revised. Of the 25 indicators tested by Ofsted, only two relate to the 'impact' of the curriculum. This is not a problem in itself because the impact of a high-quality curriculum will be felt in the quality of work produced by learners. Ofsted seem to tacitly agree with this, as their new inspection handbook says that to receive 'Good' or 'Outstanding' on the 'Quality of education' judgement, pupils must "develop detailed knowledge and skills across the curriculum [...and] this is reflected in results from national tests and examinations ...or in the qualifications obtained."¹⁶⁹ In effect, Ofsted recognise that exam results and tests will largely (if not entirely) capture the outcome of a high-quality curriculum, which surely negates the need for Ofsted to inspect the 'impact' of the curriculum as well.

Consequently, this report envisages that Ofsted's new judgement on the curriculum should be restricted to its intent and implementation. This will allow Ofsted to assess whether school leaders and teachers have constructed a well-designed, properly sequenced programme for each subject as well as allowing them to monitor signs of the curriculum being narrowed or pupils otherwise being denied access to a broad and balanced curriculum.

It is also encouraging to see more emphasis being placed on careers guidance and extra-curricular activities in the new EIF but these areas are still combined within the proposed 'Personal development' judgement. Such is the importance of these two topics for the wider

development of pupils, and because leadership teams need a stronger incentive to prioritise such matters given the pressure on examination results, this report recommends that separate judgements are given to careers guidance and extra-curricular activities. In both cases, the focus must be on measurable and tangible aspects of the respective activities to avoid inspectors making subjective judgements about, for example, the ‘character’ of pupils and whether schools are ‘going the extra mile’. Instead, the inspection should assess the breadth of provision in terms of how many opportunities are being offered, the take-up among pupils (including which groups of pupils engage with them), the frequency of contact with employers and so forth.

As noted earlier, the driving force behind the four sub-judgements currently given to schools is the *Education Act 2011*. The Act should therefore be amended as follows to incorporate both Recommendation 1 (scrapping overall grades for schools) and Recommendation 3:

CURRENT LEGISLATION – EDUCATION ACT 2011¹⁷⁰

“(5) It is the general duty of the Chief Inspector, when conducting an inspection under this section, to report on the quality of education provided in the school.

(5A) The Chief Inspector’s report under subsection (5) must in particular cover —

- (a) the achievement of pupils at the school;*
- (b) the quality of teaching in the school;*
- (c) the quality of the leadership in and management of the school;*
- (d) the behaviour and safety of pupils at the school.*

PROPOSED LEGISLATION

“(5) It is the duty of the Chief Inspector, when conducting an inspection under this section, to report on —

- (a) the behaviour and safety of pupils at the school;*
- (b) the quality of the curriculum in the school;*
- (c) the quality of careers information, advice, guidance and experiences at the school;*
- (d) the range, quality and take-up of extra-curricular activities at the school.*

PART 2: A better way of holding schools to account

RECOMMENDATION 4

A new 1-page School Information Card (SIC) should be published on an annual basis for every school. The SIC will contain 12 indicators, covering four performance measures (e.g. pupil progress), four judgements made by Ofsted inspectors (e.g. curriculum quality) and four measures of wider school life (e.g. student and staff survey results). The measures on the SIC will be presented in both absolute and relative terms to provide a comprehensive view of the school.

Following a consultation, the Labour Government in 2009 confirmed their plans to introduce a 'school report card' (SRC; see Appendix for an example). This SRC, which was to be introduced from 2011, would report on a wide range of different outcomes achieved by schools in England rather than simply focusing on attainment and league tables.¹⁷¹ The final list of measures on the SRC included: Pupil Progress; Pupil Attainment; Pupil Wellbeing; Pupils' Perceptions; Parents' Perceptions; and Narrowing Gaps in Pupil Performance. Along with basic factual information about the school, these measures would be displayed alongside the current Ofsted ratings. The measures would be updated on an annual basis. In addition, each of the measures would be accompanied by a letter-based grading system (from A to F) and a description of how each score had changed over time.

There would also be an 'overall score' from A to F for the whole school, although this was included against the wishes of the majority of consultation respondents.¹⁷² The government even accepted that "parents and other stakeholders will rightly have different views about what – for them – constitutes good outcomes for a school [and] different parents will be looking for different strengths, reflecting the specific interests, aptitudes and needs of their children",¹⁷³ yet they insisted on an overall score regardless.

A report by the House of Commons Children, Schools and Families Committee in 2010 on school accountability "[welcomed] in principle the introduction of the school report card as a rationalisation of current accountability mechanisms and an attempt at providing a broader evidence base for assessing schools' performance".¹⁷⁴ Even so, they were "struck by the weight of evidence we have received which argues against an overall score" and stated that "a school report card is not, and in our view never can be, a full account of a school's performance, yet the inclusion of an overall score suggests that it is."¹⁷⁵

Recent research in America analysed the sixteen states that have adopted school report cards that assign overall A-F letter grades to schools, and reached the same conclusion:

*“A single composite score as an index of school qualities is a dubious proposition. It is by no means clear what a single grade can mean across such a diverse array of criteria - achievement, attendance rates, dropout rates, advanced class offerings, and so on ...Little, if any, attention is paid to how to justify combining the diverse components of each grade to render a value on with letter grades as a measure of school quality.”*¹⁷⁶

The Select Committee felt that, ultimately, “parents and others should be able to decide for themselves those measures of performance most important to them.”¹⁷⁷ The Committee was also concerned that Ofsted ratings would be put alongside scores on the SRC, as this would lead to “potential for substantial confusion” and “a perception of incoherence in the accountability system.”¹⁷⁸ More broadly, the Committee noted that the accountability system “implies that schools welcome the opportunity to take ‘ownership of their own improvement’ but then provides the perfect example of how they have been prevented from doing just that.”¹⁷⁹ This was illustrated by the fact that the Committee “received strong evidence that schools feel coerced and constrained by the outcomes of Ofsted inspection”.¹⁸⁰

The simplicity offered by the SRC’s 1-page summary of a school using a broad range of metrics would deliver what parents have requested to help them choose schools. Nonetheless, there are good reasons to move beyond the original SRC model. First, the potential confusion and incoherence caused by overlapping roles and responsibilities for Ofsted and the SRC data detracted from the usefulness and accessibility of the SRC as it was previously proposed, so greater clarity is required. Second, the quality and focus of school performance data has changed considerably over the past decade so an SRC would need to reflect this. Third, the case for an overall score was not adequately made by the then government and there is considerable evidence to reject this aspect of the SRC. By building these alterations into the initial concept of an SRC, an improved model can be constructed that delivers a broad range of information to parents in a simple and accessible manner.

This report proposes that a new School Information Card (SIC) is compiled each year by the DfE. The example discussed in this report (see Figure 2 on page 41) will describe a SIC for a secondary school, although the same principles would apply to primary schools and colleges, albeit with a slightly adjusted set of metrics.

The first section of the SIC will be ‘Examination Results’, which covers the main performance measures that the DfE publishes for each school:

- Pupil progress – a description of the Progress 8 score (‘well above average’, ‘above average’, ‘average’, ‘below average’ or ‘well below average’)
- Pupil achievement - a description of the Attainment 8 score¹⁸¹ (‘well above average’, ‘above average’, ‘average’, ‘below average’ or ‘well below average’)

- English and Maths – the percentage of pupils achieving Grade 5 or above for GCSE English and Maths
- EBacc – the percentage of pupils attaining Grade 5 in English Baccalaureate subjects

All this information will be calculated as a three-year rolling average instead of a single-year metric. This mirrors a similar proposal from the NAHT Accountability Commission, as they too felt that presenting performance data in this way “will help to ensure that annual data is not skewed by simple cohort differences, year-on-year.”¹⁸²

The second section of the SIC will be ‘Ofsted inspections’, which includes the four separate judgements made by inspectors under the proposed inspection model in this report. These four judgements highlight the value of having inspectors visiting schools to assess the observable aspects of school life not captured by examination data. These will all be scored on a scale of Very Good-Good-Average-Poor-Very Poor:

- Behaviour and safety – the behaviour of pupils and the extent to which they feel safe and are kept safe
- Curriculum quality – the rigour, depth and breadth of curriculum being provided
- Careers advice and guidance – the quality of careers information, advice, guidance and experiences for pupils
- Extra-curricular activities – the range, quality and take-up of activities by pupils

The third section will be ‘School life’, which captures the views and perspectives of the main stakeholders in a school community as well as additional information about pupils:

- Student well-being – the average score from students on a survey measuring satisfaction with academic and pastoral support
- Staff well-being – the average score from teaching staff on a survey measuring satisfaction and well-being
- Parent satisfaction – the average score on Ofsted’s Parent View survey
- Pupil absence – unauthorised absence rates for pupils at the school

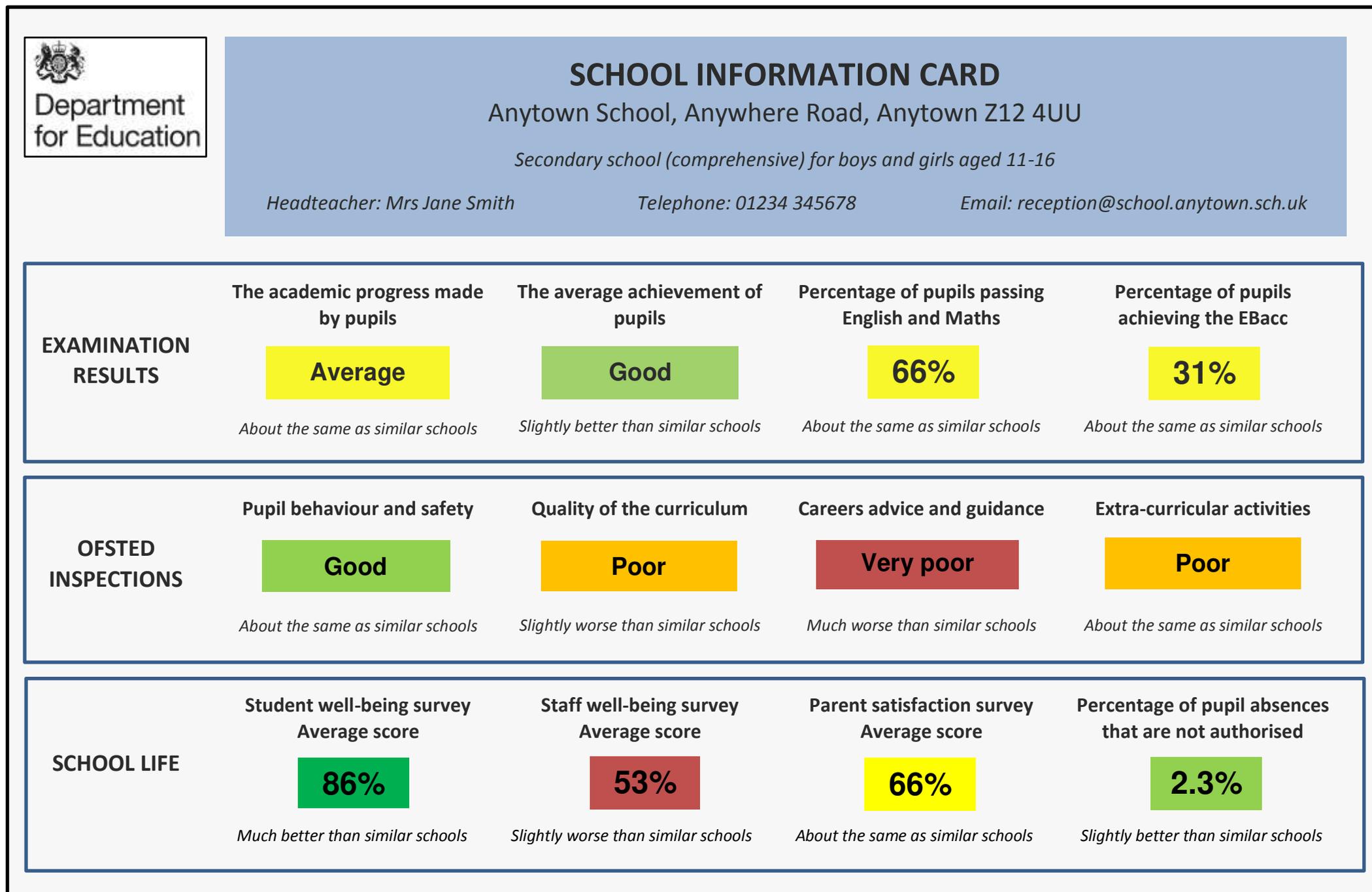
In terms of how the 12 indicators are presented on the SIC, this report proposes providing two pieces of information to parents. First, it will show the score on each measure in absolute terms, with a colour-coding system to indicate the school’s position nationally. Second, it will also offer a short verbal comment that indicates whether the school is doing better or worse than other schools with similar characteristics. Moreover, it is envisaged that the reverse side of the SIC will contain a brief description of all 12 measures, should parents wish to understand more about what each score represents and how it is calculated.

Ofsted already conducts pupil and staff surveys during inspections, but the results are not available to parents or even the school leadership team unless they request it. If schools are to be truly accountable to parents in an autonomous school system, they should be given access to this information. What's more, the existing surveys for staff and pupils cover a plethora of topics, which reduces their value because it is possible that respondents may be positive about some aspects of school life but less so about others.

As discussed earlier in this chapter, parents have told Ofsted that they would like more information about topics such as the happiness and mental health of pupils, pastoral care, parental experiences and staff wellbeing. This report therefore recommends that not only should the results of student and staff surveys be made public, the surveys themselves should be tightly focused on the overall well-being and treatment of staff and pupils. For staff, this should explicitly focus on their perspectives on the effectiveness of the leadership team, the breadth and depth of support they receive from senior leaders, workload and planning issues, cohesion and collaboration among teachers and leaders and the quality and quantity of communication from senior leaders. For pupils, this should focus on their views on pupil safety, the extent of bullying and how effectively it is tackled, support for (and promotion of) physical and mental well-being, the quality and availability of academic support from teachers (including whether pupils know how to improve) and the quality and availability of pastoral support from teachers, tutors and senior leaders. Some of the above issues are already covered in the existing surveys, but a more targeted set of questions will help improve their usefulness to parents.

Ofsted collects the views of parents through its 'ParentView' portal, which is designed to give parents "the chance to tell Ofsted what you think about your child's school, from the quality of teaching to dealing with bullying and poor behaviour."¹⁸³ Unlike the staff and student surveys that are completed during the inspection, the parent survey can be completed at any time and Ofsted will then use the survey results when making decisions about which schools to inspect and when.¹⁸⁴ Given that parents have already told Ofsted that they want to hear more from parents of children who attend the school,¹⁸⁵ it is proposed that the results of the ParentView survey for each school should be included on the SIC. As with the staff and student surveys, Ofsted should focus their questions for parents on a targeted set of issues rather than asking questions that are too generic (e.g. 'my child is taught well at this school'). The current survey also asks whether parents agree that 'my child makes good progress at this school' and 'I receive valuable information from the school about my child's progress' within its set of 12 questions,¹⁸⁶ which does not appear to maximise the value of such an exercise.

FIGURE 2: EXAMPLE OF A NEW SCHOOL INFORMATION CARD



For the SIC, an average score (in percentage terms) should be calculated from the responses to each of the three surveys, which are then added onto the SIC. Needless to say, like any other source of data, survey evidence has its imperfections. It will take time to test and confirm the reliability and validity of an expanded set of survey questions for all three stakeholder groups. Ofsted has struggled to encourage parents to fill in ParentView, with an average of just 18 responses being received from parents at maintained schools.¹⁸⁷ It is hard to know for certain what is driving these low response rates, although the lack of visibility of the results from these surveys in Ofsted reports and the inspection process as a whole will not have encouraged parents to become involved. A concerted effort is required to improve the quality and coverage of questions in the surveys as well as ensure that sufficient responses are received from pupils, staff and parents. In addition to the results of the surveys, it is proposed that the SIC should show the rate of unauthorised pupil absences. This is already collected and published by the DfE for all schools, so it would be logical to include this information for parents.

RECOMMENDATION 5

Ofsted inspections should be carried out every 2-3 years, with inspectors spending a day in each school to update their judgements. Ofsted should aim to cover two of their four new judgements as part of each visit (e.g. a 1-day inspection covering behaviour and the curriculum). The revised judgements will be added to the SIC as soon as possible after the inspection has taken place.

With hundreds of schools not being inspected at any point within the last decade, the value of the judgements made by inspectors is necessarily diminished. The SIC should therefore be updated as regularly as possible. For examination results, this should be completed as soon as the results are confirmed by the DfE, ensuring that this information is available to parents before they choose their preferred school(s). That said, it would not be possible to repeat Ofsted inspections and collect large samples of survey responses on an annual basis without a significant injection of financial resources. These inspections should therefore be carried out every 2-3 years.

With a more targeted remit, Ofsted will be able to use their resources to greater effect. Ofsted inspections should typically last for one day, and during this time they will cover two or three of their new areas for inspection. For example, behaviour and the curriculum could be combined into a 1-day inspection, whereas on another 1-day inspection Ofsted would cover careers advice and extra-curricular activities. The purpose of these visits will be for inspectors to construct a fair and objective judgement on these areas throughout the course of the day. Once the inspection is complete, the updated rating(s) will be immediately added to the SIC.

RECOMMENDATION 6

In future, Ofsted reports should be much shorter than at present – ideally no more than two pages. Using the information published on the SIC, they should provide commentary for parents on two main areas: the current strengths of a school, and its areas for improvement. These reports will be regularly updated by Ofsted.

Another key issue for discussion is how this new role for Ofsted affects the reports they produce following an inspection. The SIC is designed to give parents easy access to a broad range of indicators about how a school is performing. Ofsted’s survey of parents in 2017 found that, of the parents who had read an Ofsted report, around 40% either just read the front page with the overall judgements or only read the sections that they were interested in. Only 19% of parents read the whole report.¹⁸⁸ On that basis, Ofsted should in future produce a much shorter and more accessible report for parents – ideally no more than two pages. This report should include two themes: the current strengths of a school, and the areas where the school needs to improve. This would represent a considerable simplification on existing Ofsted reports that are typically 6-10 pages long and attempt to describe almost every aspect of a school (largely due to the existing inspection framework).

These shorter reports should be regularly updated (ideally, after each inspection visit) and they are intended to provide commentary on the full range of the 12 indicators on the SIC. This commentary should be written primarily for parents, although it will of course be available to everyone once published. Ofsted should aim to be objective and impartial when producing this commentary on the SIC, although they should not artificially seek to balance the strengths and areas for improvement at any school if there is a good reason to place more emphasis on one or the other.

RECOMMENDATION 7

The proposals in this report should be introduced by Ofsted as their new inspection framework in September 2020 – a year later than currently planned. The intervening period should be used to trial the proposals outlined in this report including the new SIC, the updated stakeholder surveys and Ofsted’s revised role and responsibilities (e.g. assessing the validity and reliability of scrutinising pupils’ work).

At the time of writing, Ofsted is planning to implement its new inspection framework in September 2019. Given the depth and breadth of the changes proposed in this report, it would present enormous logistical challenges to both Ofsted and school leaders to introduce the required changes with only a few months’ notice. For example, several of the proposals will require their own period of testing for validity, reliability and efficacy, including the

new judgements on the curriculum and the updated versions of the staff, student and parent surveys. Ofsted inspectors may also require additional training in areas such as behavioural observations and there will need to be extensive consultations (particularly with parents) on the new SIC. On this basis, we recommend that the next 15-18 months should be set aside to give Ofsted (in partnership with the DfE, teaching unions and other key stakeholders) the time and space required to design and rigorously test the proposals in this report before they are fully implemented.

7. Areas for further consideration

The 12 measures put forward for the SIC will inevitably generate some discussion. The four performance measures chosen for secondary schools are merely a reflection of the four main measures used by the DfE rather than necessarily representing the ideal choices for the SIC. Likewise, some of the proposed areas for Ofsted judgements could potentially be combined (e.g. careers advice and extra-curricular activities) to free up space for them to inspect other aspects of school life instead, so long as it is not already captured by performance data. The set of questions that should be included on surveys for students, staff and parents will also require careful consideration (along with appropriate testing for their reliability and validity) before they represent the finished article, although existing surveys could be used in the meantime.

An important discussion would need to take place about the nature of the measures on the SIC. For example, it is debatable whether comparisons between schools should be done against all schools nationally, locally or only those schools with similar characteristics (e.g. pupil demographics, urban versus rural location). Some of the information on the SIC, particularly survey results and data on pupil absence, could potentially be skewed by such factors, as could a school's ability to offer certain extra-curricular activities and services to pupils, parents and their local community. That said, Ofsted's survey of parents in 2017 suggested that they were keen to receive 'details about how a school compares to other schools locally'.¹⁸⁹ An element of contextualisation to the information of the SIC may therefore be prudent, although no decisions should be made until stakeholders and accountability experts have been widely consulted.

The production of the SIC on an annual basis would be a step forward for school accountability but it would still require additional investment. A global review of school report cards in 2016 found that several issues can arise during their implementation, such as overcoming the barriers created by poor literacy rates among some parents as well as striking the right balance between presenting large amounts of information to parents while maintaining simplicity in the report card design.¹⁹⁰ Furthermore, there would need to be discussions on which organisations or agencies should take responsibility for disseminating and displaying the SICs (even if the DfE produces them), particularly when schools will not always have an incentive to do so.

In the absence of overall ratings for schools, Ofsted will no longer be required to re-inspect some schools more regularly than others. Instead, their role will be to provide independent judgements on non-examined areas of school life at regular intervals to ensure all stakeholders have a constant flow of relevant information about each school. At present,

Regional Schools Commissioners (RSCs) draw on Ofsted's inspection reports but undertake their own assessments as well when monitoring academies. Ofsted reports are also used by the DfE to monitor educational standards and to trigger intervention where it is needed. For example, the *Education and Inspections Act 2006* states that an Ofsted judgement of 'Inadequate' is one of the main triggers for intervention or possible school closure.¹⁹¹ Given that Ofsted reports will be much shorter in future and no overall judgements for schools will be provided, consideration will need to be given as to how the DfE, RSCs and local authorities can best monitor the performance of schools using SICs instead of Ofsted ratings. In theory, the new SIC could provide a more sensible basis for discussing interventions because it covers a broader range of activities and could potentially identify weaknesses that are not being picked up by relying solely on Ofsted grades. Using the SIC to identify failing schools could therefore take a more nuanced approach e.g. intervention could be triggered if a school's progress scores fall for three years in a row, or if the quality of their curriculum is found to be consistently low.

The notice period given to schools before Ofsted inspectors arrive is another area that may need to be revisited. Ofsted's survey of parents in 2017 found 61% agreeing that schools shouldn't have any notice of inspections.¹⁹² This contrasts with the existing arrangements where schools are typically given 1-2 days' notice, although Ofsted retain the right to inspect without notice if there are serious concerns about issues such as declining standards or safeguarding.¹⁹³ Under the new EIF, Ofsted are proposing to phone a school at 10am the morning before an inspection begins and then arrive on-site that afternoon to begin preparations for the inspection that commences the following day.¹⁹⁴ Given that the recommendations in this report mean Ofsted's responsibilities will have been redirected towards only those elements of school life that require direct observation, it is essential that they observe the reality of how a school community functions. This supports the idea that as little notice as possible should be given in order to prevent leadership teams and teachers from altering the way that a school operates to change the views of inspectors. Nevertheless, Amanda Spielman acknowledged last year that Ofsted would not want schools being caught out without the relevant representatives being present.¹⁹⁵ In advance of the new inspection framework being finalised, it is difficult to determine the best balance between these two priorities.

Conclusion

“We are clear [...] that in our concern to ensure that inspection processes are consistent, reliable and intelligible, we have not unwittingly filtered out the less measurable or more creative aspects of education. If inspection is perceived as imposing uniformity, discouraging initiative or stifling imagination, its capacity to bring about improvement will be undermined.”¹⁹⁶

This quote, from the then Chief Inspector Chris Woodhead in Ofsted’s annual report for the year 1993/94 (its first full year of inspections) shows the inspectorate has always recognised that it treads a fine line. More than 25 years on, the evidence strongly suggests that Ofsted has failed to meet the challenge it set for itself. A distinct lack of reliability and validity in their observations and judgements, coupled with vague instructions and countless mixed messages, has resulted in the education sector losing confidence in the value of Ofsted. The effect of inspections on leaders and teachers is impossible to ignore, especially the detrimental impact on staff workload. This becomes even more concerning in the context of the serious on-going issues with teacher recruitment and retention.

Despite their desire to be seen as a ‘force for improvement’,¹⁹⁷ Ofsted’s own survey of teachers in 2018 showed that just 24% of them agree that they perform this function.¹⁹⁸ Although it is encouraging to hear Ofsted talk about wanting its inspections and regulations to be intelligent, responsible and focused, this does not address more fundamental questions of why the inspectorate exists and what purpose (and whom) it should serve. This report has described how Ofsted’s approach to judging teachers and schools is not based on research evidence from this country or abroad, nor have the accuracy and consistency of their judgements ever been satisfactorily tested and assured. While these issues could be tackled in isolation, merely modifying existing processes and procedures will not be sufficient to bring about the changes that are now required.

The proposals in this report, particularly the new School Information Card, are designed to shift the attention of schools away from Ofsted and towards allowing parents to hold schools to account for a wide range of outcomes. Some of the recommendations in this report are intended to provide a basis for further discussion rather than representing definitive answers to the questions that the report has posed. Indeed, this report very much aims to start a debate rather than finish it (particularly as, for the sake of brevity, the proposals have only focused on secondary schools).

There will also be legislative hurdles that would need to be cleared to enact several of these proposals as Ofsted's functions are bound up in numerous Acts of Parliament spanning many years. That said, the clear conclusion from this report is that Ofsted needs to change course if it is to gain the respect of policymakers and the teaching profession as well as produce fair and accurate information to support parental choices.

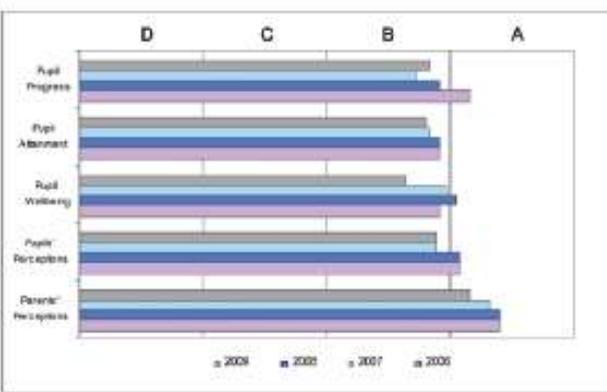
There is still a crucial place in our education system for an independent inspectorate that visits schools to provide parents of current and future pupils with valuable insights. On that basis, this report seeks to craft a new, more constructive role for Ofsted in which their goal is to help parents make more informed choices instead of trying to come up with all the answers themselves. By moving away from the notion of 'grading' schools and towards empowering parents with better information, the education community as a whole will reap the benefits of having a self-improving school system that includes Ofsted, but without teachers and leaders having to experience the burdens that inspections generate at present. If this can be delivered, our education system will be a much better, happier and healthier place.

APPENDIX

Example of a 'School Report Card' proposed in 2009

THIS EXAMPLE IS FOR ILLUSTRATIVE PURPOSES ONLY

	<h3>School Report Card 2009</h3>		
<p>Anytown School General Borough Council</p> <p>Head teacher: Mrs Jane Smith The School's website: http://www.anytownschool.co.uk</p>	<p>Pupil Progress A </p> <p>Pupil Attainment B </p> <p>Pupil Wellbeing B </p> <p>Pupils' Perceptions A </p> <p>Parents' Perceptions A </p> <p>Narrowing Gaps in Pupil Performance </p>	<p>↑ Up since last year</p> <p>↔ Same as last year</p> <p>↓ Down since last year</p> <p>↔ Same as last year</p> <p>↔ Same as last year</p>	<div style="border: 1px solid black; padding: 10px; width: 100px; margin: 0 auto;"> <p style="text-align: center; font-size: 24px; font-weight: bold;">A</p> <p style="text-align: center;">Overall Score</p> </div>
<p>Outcome of the School's Last Inspection by Ofsted This school was last inspected on Day/Month/Year.</p> <p>The key judgement on Overall effectiveness: GOOD (grade 2) The full inspection report can be seen at http://www.ofsted.gov.uk</p>		<p>Early Years Foundation Stage or Sixth Form Provision <i>A measure of the effectiveness of the school's outcomes for Early Years Foundation Stage or its sixth form.</i></p>	
<p>Behaviour The Ofsted judgement on pupils' behaviour: GOOD (grade 2) This judgement is reflected in the Ofsted overall effectiveness grade.</p>		<p>Partnership Working This school is a member of the xxx partnership. The Report Card score for this partnership is A Details can be found at www.anytownpartnership.co.uk</p>	
<p>Safeguarding The Ofsted judgement on the effectiveness of safeguarding procedures: GOOD (grade 2) This judgement is reflected in the Ofsted overall effectiveness grade.</p>		<p>The Ofsted judgement on the effectiveness of partnerships in promoting learning and well-being: GOOD (grade 2) This judgement is reflected in the Ofsted overall effectiveness grade.</p>	

	<h3>School Report Card 2009</h3>		
<p>Year On Year Comparison</p>			
			
<p>This Section Has Been Completed By The School</p>			
<p>"Our school ethos centres on the spirit of scholarship and opportunity. This is reflected in our dedication, application and the will to succeed, as well as the care and help offered to all members of the Anytown community."</p> <p>We also have a twenty place MLD Unit within our school and the children there play an active and integrated part in the life of the whole school.</p>			
<p>Background Information</p> <p>Address: Anywhere Road, Anytown, Z12 4UU</p> <p>Telephone: 01240 567001</p> <p>Age range: 11-19</p> <p>Gender: MIXED</p> <p>Type: Comprehensive</p>			
<p>Total number of pupils (all ages) 918</p>		<p>Total number of pupils with SEN, with statements or on School Action Plus 85</p>	
<p>Percentage of pupils with SEN, with statements or on School Action Plus 9.3%</p>		<p>Total number of pupils with SEN, supported at School Action 100</p>	
<p>Percentage of pupils with SEN, supported at School Action 10.9%</p>			
			

REFERENCES

-
- ¹ HM Government, *Education (Schools) Act 1992*, Chapter 32.
- ² Ibid.
- ³ Conservative Party, '1992 General Election Manifesto: The Best Future for Britain', Webpage, 2018.
- ⁴ Ibid.
- ⁵ Ibid.
- ⁶ Dr Adrian Elliott, *Twenty Years Inspecting English Schools – Ofsted 1992-2012* (London: Research and Information on State Education, 2012).
- ⁷ National Audit Office, *Ofsted's Inspection of Schools* (London: NAO, 2018).
- ⁸ Elliott, *Twenty Years Inspecting English Schools – Ofsted 1992-2012*.
- ⁹ Ibid.
- ¹⁰ Ibid.
- ¹¹ Ofsted, 'Ofsted: A Force for Improvement', Press release, 29 September 2017.
- ¹² Ibid.
- ¹³ John Micklewright et al., *Teachers in England's Secondary Schools: Evidence from TALIS 2013* (London: Department for Education, 2014), 22.
- ¹⁴ Alix Robertson, 'Ofsted Attempts Parental Charm Offensive', *Schools Week*, 25 May 2018.
- ¹⁵ Kaire Pöder and Triin Lauri, 'When Public Acts like Private: The Failure of Estonia's School Choice Mechanism', *European Educational Research Journal* 13, no. 2 (1 January 2014): 220–34.
- ¹⁶ Ofsted, *The Education Inspection Framework: Draft for Consultation* (Manchester: Ofsted, 2019), 4.
- ¹⁷ Ofsted, *Annual Parents Survey 2017: Parents' Awareness and Perceptions of Ofsted* (London: Ofsted, 2018).
- ¹⁸ Becky Francis and Merryn Hutchings, *Parent Power? Using Money and Information to Boost Children's Chances of Educational Success* (London: Sutton Trust, 2013), 15.
- ¹⁹ Ofsted, *Annual Parents Survey 2017: Parents' Awareness and Perceptions of Ofsted*.
- ²⁰ Ibid.
- ²¹ Ofsted, 'Ofsted: A Force for Improvement'.
- ²² House of Commons Education and Employment Committee, *The Work of OFSTED*, HC 62-I (London: The Stationery Office, 1999).
- ²³ Ibid.
- ²⁴ In this instance, correlations are measured on a scale of 0 to 1, with 0 indicating that there is no agreement between the judgements made by the different inspectors and 1 indicating that there is perfect (constant) agreement between the inspectors' judgements.
- ²⁵ House of Commons Education and Employment Committee, *The Work of OFSTED*.
- ²⁶ Ibid.
- ²⁷ Peter Matthews et al., 'Aspects of the Reliability and Validity of School Inspection Judgements of Teaching Quality', *Educational Research and Evaluation* 4, no. 2 (June 1998): 167–88.
- ²⁸ Ofsted, *'Do Two Inspectors Inspecting the Same School Make Consistent Decisions?': A Study of the Reliability of Ofsted's New Short Inspections* (Manchester: Ofsted, 2017).
- ²⁹ Ibid., 4.

-
- ³⁰ Adi Bloom, 'Ofsted Chief: Trials Show "Inspectors Make Consistent Judgements"', *Times Educational Supplement*, 7 March 2017.
- ³¹ National Audit Office, *Ofsted's Inspection of Schools*.
- ³² Colin Richards, 'The Reliability Genie Is out of the Bottle at Ofsted – and It Will Only Lead to More Pressure on Watchdog', *Times Educational Supplement*, 13 March 2017.
- ³³ Russell Hobby, 'Ofsted Reliability Study: "We Need to Reduce the Consequences of Inspectors" Judgements"', *Times Educational Supplement*, 7 March 2017.
- ³⁴ Ofsted, 'Do Two Inspectors Inspecting the Same School Make Consistent Decisions?': *A Study of the Reliability of Ofsted's New Short Inspections*, 9.
- ³⁵ *Ibid.*, 41.
- ³⁶ Progress 8 is the headline indicator of school performance. It aims to capture the progress a pupil makes from the end of primary school to the end of Key Stage 4 (GCSEs). It gives an indication of whether, as a group, pupils in the school made above or below average progress compared to similar pupils in other schools. Progress 8 is based on students' progress across eight subjects: English; Mathematics; three other English Baccalaureate (EBacc) subjects (sciences, computer science, geography, history and languages); and three further EBacc or non-EBacc subjects.
- ³⁷ Ofsted, *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2016/17*, 2017, 36.
- ³⁸ Iftikhar Hussain, *Subjective Performance Evaluation in the Public Sector: Evidence from School Inspections?*, Discussion Paper 135 (London: Centre for the Economics of Education, 2012).
- ³⁹ Ofsted, *Six Models of Lesson Observation: An International Perspective* (Manchester: Ofsted, 2018), 22.
- ⁴⁰ Ofsted, *School Inspection Handbook: Handbook for Inspecting Schools in England under Section 5 of the Education Act 2005* (Manchester: Ofsted, 2018), 13.
- ⁴¹ *Ibid.*, 24.
- ⁴² *Ibid.*
- ⁴³ Ofsted, *Annual Teachers Survey 2018: Teachers' Awareness and Perceptions of Ofsted* (London: Ofsted, 2018).
- ⁴⁴ Steven Cantrell and Thomas J. Kane, *Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study* (Seattle: Bill & Melinda Gates Foundation, 2013), 18.
- ⁴⁵ *Ibid.*
- ⁴⁶ Ofsted, *School Inspection Handbook: Handbook for Inspecting Schools in England under Section 5 of the Education Act 2005*, 25.
- ⁴⁷ Ofsted, 'Do Two Inspectors Inspecting the Same School Make Consistent Decisions?': *A Study of the Reliability of Ofsted's New Short Inspections*, 12.
- ⁴⁸ Heather Hill, Charalambos Charalambous, and Matthew Kraft, 'When Rater Reliability Is Not Enough: Teacher Observation Systems and a Case for the Generalizability Study', *Educational Researcher* 41, no. 2 (2012): 56–64.
- ⁴⁹ Rebecca Allen, 'Is an Ofsted Judgement a Lagging or Leading Indicator of School Performance?', *Education Datalab*, 30 November 2015.
- ⁵⁰ A primary school is below the floor standard if two criteria are met in their Key Stage 2 results: (i) the percentage of pupils reaching the expected standard in reading, writing and maths was below 65%; and (ii) any one of three progress measures was below a certain standard (reading progress was below -5; or writing progress was below -7; or maths progress was below -5).
- ⁵¹ Philip Nye and Dave Thomson, 'KS2 Performance Tables 2017: Ofsted Ratings and Floor Standards', *Education Datalab*, 14 December 2017.
- ⁵² Secondary schools with a Progress 8 score of below -0.5 are below the government's floor standard.

-
- ⁵³ Philip Nye and Dave Thomson, 'KS4 Performance Tables 2017: Everything We Know about Coasting Schools and Floor Standard Schools', *Education Datalab*, 25 January 2018.
- ⁵⁴ Jo Hutchinson, *School Inspection in England: Is There Room to Improve?* (London: Education Policy Institute, 2016), 5.
- ⁵⁵ *Ibid.*, 6.
- ⁵⁶ *Ibid.*, 7.
- ⁵⁷ Freddie Whittaker, 'Spielman: Ofsted Inspectors Know Progress 8 "Isn't Perfect"', *Schools Week*, 3 July 2018.
- ⁵⁸ National Audit Office, *Ofsted's Inspection of Schools*, 9.
- ⁵⁹ Ofsted, *Annual Teachers Survey 2018: Teachers' Awareness and Perceptions of Ofsted*.
- ⁶⁰ *Ibid.*
- ⁶¹ National Audit Office, *Ofsted's Inspection of Schools*, 6.
- ⁶² Hannah Richardson, 'Ofsted Admits Some "Outstanding Schools Aren't That Good"', *BBC News Online*, 24 May 2018.
- ⁶³ Ofsted, 'Amanda Spielman at the Bryanston Education Summit', Webpage, 6 June 2018.
- ⁶⁴ Rosemary Bennett, 'Top Ratings from Ofsted Lowered on Return Visit', *The Times*, 24 January 2019.
- ⁶⁵ Harriet Waldegrave and Jonathan Simons, *Watching the Watchmen: The Future of School Inspections in England* (London: Policy Exchange, 2012), 18.
- ⁶⁶ *Ibid.*
- ⁶⁷ *Ibid.*, 19.
- ⁶⁸ *Ibid.*
- ⁶⁹ Michael Strong, John Gargani, and Özge Hacifazlioglu, 'Do We Know a Successful Teacher When We See One? Experiments in the Identification of Effective Teachers', *Journal of Teacher Education* 62, no. 4 (2011): 378.
- ⁷⁰ *Ibid.*, 367.
- ⁷¹ *Ibid.*
- ⁷² Ofsted, *School Inspection Handbook: Draft for Consultation* (Manchester: Ofsted, 2019), 4.
- ⁷³ Ofsted, 'Ofsted: A Force for Improvement'.
- ⁷⁴ Ofsted, *Education Inspection Framework: Overview of Research* (Manchester: Ofsted, 2019), 1.
- ⁷⁵ Ofsted, *An Investigation into How to Assess the Quality of Education through Curriculum Intent, Implementation and Impact: Phase 3 Findings of Curriculum Research* (Manchester: Ofsted, 2018).
- ⁷⁶ *Ibid.*, 31.
- ⁷⁷ *Ibid.*
- ⁷⁸ *Ibid.*, 32.
- ⁷⁹ *Ibid.*
- ⁸⁰ *Ibid.*, 38.
- ⁸¹ *Ibid.*, 32.
- ⁸² *Ibid.*, 12.
- ⁸³ Ofsted, *The Education Inspection Framework: Draft for Consultation*, 26.
- ⁸⁴ Ofsted, *An Investigation into How to Assess the Quality of Education through Curriculum Intent, Implementation and Impact: Phase 3 Findings of Curriculum Research*, 11.
- ⁸⁵ Ofsted, *School Inspection Handbook: Draft for Consultation*, 27.

-
- ⁸⁶ Ibid.
- ⁸⁷ Ibid., 47.
- ⁸⁸ Ibid., 45.
- ⁸⁹ Ofsted, *Common Inspection Framework: Education, Skills and Early Years from September 2015* (Manchester: Ofsted, 2015).
- ⁹⁰ Ofsted, *The Education Inspection Framework: Draft for Consultation*, 6.
- ⁹¹ Ofsted, *School Inspection Handbook: Handbook for Inspecting Schools in England under Section 5 of the Education Act 2005*, 12.
- ⁹² Ibid., 13.
- ⁹³ Ibid., 50.
- ⁹⁴ Ibid., 53–54.
- ⁹⁵ Ibid., 53.
- ⁹⁶ Ofsted, *School Inspection Handbook: Draft for Consultation*, 49.
- ⁹⁷ Ibid., 50.
- ⁹⁸ Ibid., 59.
- ⁹⁹ Ibid., 60.
- ¹⁰⁰ Ofsted, 'State-Funded Schools Inspections and Outcomes as at 31 March 2018', Webpage, 26 June 2018.
- ¹⁰¹ Ofsted, 'HMCI's Commentary: Recent Primary and Secondary Curriculum Research', Webpage, 11 October 2017.
- ¹⁰² Ofsted, *School Inspection Handbook: Handbook for Inspecting Schools in England under Section 5 of the Education Act 2005*, 54.
- ¹⁰³ Ibid., 57.
- ¹⁰⁴ Irena Barker, 'Schools Must Go the Extracurricular Mile, Report Argues', *Times Educational Supplement*, 24 October 2014.
- ¹⁰⁵ Ofsted, *School Inspection Handbook: Draft for Consultation*, 60.
- ¹⁰⁶ Ibid.
- ¹⁰⁷ Waldegrave and Simons, *Watching the Watchmen: The Future of School Inspections in England*, 32.
- ¹⁰⁸ Ofsted, 'School Inspections: Staff and Pupil Questionnaire Text', Webpage, 16 January 2018.
- ¹⁰⁹ Ofsted, 'Ofsted Inspections: Myths', Webpage, 12 April 2018.
- ¹¹⁰ Ibid.
- ¹¹¹ These comments were taken from a random search of 40 of the latest Ofsted inspection reports available for secondary schools in the London area as of June 2018.
- ¹¹² John Roberts, Ofsted inspectors accused of creating 'unnecessary workload', TES, 18 April 2018.
- ¹¹³ CooperGibson Research, *Workload Challenge: Analysis of Teacher Consultation Responses* (London: Department for Education, 2015).
- ¹¹⁴ Ibid., 8.
- ¹¹⁵ John Higton et al., *Teacher Workload Survey 2016: Research Report* (London: Department for Education, 2017).
- ¹¹⁶ Micklewright et al., *Teachers in England's Secondary Schools: Evidence from TALIS 2013*, 68.
- ¹¹⁷ Ofsted, *Annual Teachers Survey 2017: Teachers' Awareness and Perceptions of Ofsted* (London: Ofsted, 2017).
- ¹¹⁸ Ofsted, *Annual Teachers Survey 2018: Teachers' Awareness and Perceptions of Ofsted*.

-
- ¹¹⁹ Ibid.
- ¹²⁰ National Audit Office, *Ofsted's Inspection of Schools*, 45.
- ¹²¹ Waldegrave and Simons, *Watching the Watchmen: The Future of School Inspections in England*, 43.
- ¹²² Jane Perryman et al., 'Surveillance, Governmentality and Moving the Goalposts: The Influence of Ofsted on the Work of Schools in a Post-Panoptic Era', *British Journal of Educational Studies* 66, no. 2 (14 September 2017): 145–63.
- ¹²³ Karen L. Jones et al., 'The Unintended Consequences of School Inspection: The Prevalence of Inspection Side-Effects in Austria, the Czech Republic, England, Ireland, the Netherlands, Sweden, and Switzerland', *Oxford Review of Education* 43, no. 6 (2017): 1.
- ¹²⁴ Ibid., 14.
- ¹²⁵ Toby Greany and Rob Higham, *Hierarchy, Markets and Networks: Analysing the 'Self-Improving School-Led System' Agenda in England and the Implications for Schools* (London: UCL Institute of Education, 2018), 30.
- ¹²⁶ Ibid., 12.
- ¹²⁷ Ibid., 31.
- ¹²⁸ Ofsted, 'Amanda Spielman Launches EIF Consultation Proposals', Webpage, 16 January 2019.
- ¹²⁹ Ibid.
- ¹³⁰ Ofsted, *School Inspection Handbook: Draft for Consultation*, 62.
- ¹³¹ Ibid., 71.
- ¹³² Ibid., 70.
- ¹³³ Michael Tidd, 'Why We May Live to Regret Ofsted Ignoring Internal Data', *Times Educational Supplement*, 21 January 2019.
- ¹³⁴ Stephen Tierney, 'Ofsted Should Stop Judging Schools and Focus on Improvement', *Schools Week*, 16 January 2019.
- ¹³⁵ Ofsted, *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2016/17*, 30.
- ¹³⁶ Ibid., 32.
- ¹³⁷ Ibid., 35.
- ¹³⁸ National Audit Office, *Ofsted's Inspection of Schools*, 42.
- ¹³⁹ Ibid.
- ¹⁴⁰ Ibid., 44.
- ¹⁴¹ Ofsted, *Annual Teachers Survey 2018: Teachers' Awareness and Perceptions of Ofsted*.
- ¹⁴² National Audit Office, *Ofsted's Inspection of Schools*, 44.
- ¹⁴³ Karen L. Jones and Peter Tymms, 'Ofsted's Role in Promoting School Improvement: The Mechanisms of the School Inspection System in England', *Oxford Review of Education* 40, no. 3 (2014): 324.
- ¹⁴⁴ Ibid., 327.
- ¹⁴⁵ R. Nelson and Melanie Ehren, *Review and Synthesis of Evidence on the (Mechanisms of) Impact of School Inspections* (Netherlands: Dutch Inspectorate of Education, 2014), 6.
- ¹⁴⁶ Jones and Tymms, 'Ofsted's Role in Promoting School Improvement: The Mechanisms of the School Inspection System in England', 327.
- ¹⁴⁷ Ibid., 328.
- ¹⁴⁸ Simon Burgess et al., 'What Parents Want: School Preferences and School Choice', *The Economic Journal* 125, no. 587 (September 2015): 1262–89.

-
- ¹⁴⁹ Gabrielle Leroux, *Choosing to Succeed: Do Parents Pick the Right Schools?* (London: Social Market Foundation, 2015).
- ¹⁵⁰ P. Musset, *School Choice and Equity: Current Policies in OECD Countries and a Literature Review* (OECD Education Working Paper No. 66) (Paris: OECD, 2012), 33.
- ¹⁵¹ Ibid.
- ¹⁵² Ibid., 38.
- ¹⁵³ Ibid., 40.
- ¹⁵⁴ Justine S. Hastings and Jeffrey M. Weinstein, 'Information, School Choice, and Academic Achievement: Evidence from Two Experiments', *The Quarterly Journal of Economics* 123, no. 4 (2008): 4.
- ¹⁵⁵ Ibid., 24–25.
- ¹⁵⁶ Tahir Andrabi, Asim Ijaz Khwaja, and Jishnu Das, *Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets* (Boston: Harvard University Center for International Development, 2014).
- ¹⁵⁷ Ofsted, *Annual Parents Survey 2017: Parents' Awareness and Perceptions of Ofsted*.
- ¹⁵⁸ Ibid.
- ¹⁵⁹ Robertson, 'Ofsted Attempts Parental Charm Offensive'.
- ¹⁶⁰ Charlotte Santry, 'Ofsted to Keep "outstanding" Grades', *Times Educational Supplement*, 21 June 2018.
- ¹⁶¹ Office of the State Superintendent of Education, *DC School Report Card At A Glance* (Washington: District of Columbia, 2018).
- ¹⁶² Office of the State Superintendent of Education, *Preliminary Report Card Content & Public Engagement Responses: December 5, 2017* (Washington: District of Columbia, 2017).
- ¹⁶³ Taylor Mulcahey, 'DC Officials Look to Hone New School Report Cards for December Release', *The DC Line*, 19 August 2018.
- ¹⁶⁴ HM Government, *Education Act 2011*, Section 41.
- ¹⁶⁵ Ofsted, 'Amanda Spielman Speech to the SCHOOLS NorthEast Summit', Webpage, 11 October 2018.
- ¹⁶⁶ Ofsted, *Six Models of Lesson Observation: An International Perspective*, 22.
- ¹⁶⁷ Ibid., 12.
- ¹⁶⁸ Ibid., 15.
- ¹⁶⁹ Ofsted, *School Inspection Handbook: Draft for Consultation*, 50.
- ¹⁷⁰ HM Government, *Education Act 2011*.
- ¹⁷¹ Department for Children, Schools and Families, *A School Report Card : Prospectus* (London: Department for Children, Schools and Families, 2009), 3.
- ¹⁷² Ibid., 6.
- ¹⁷³ Ibid., 8.
- ¹⁷⁴ House of Commons Children, Schools and Families Committee, *School Accountability: First Report of Session 2009–10*, HC 88-I (London: Her Majesty's Stationery Office, 2010), 7.
- ¹⁷⁵ Ibid., 8.
- ¹⁷⁶ Kenneth Howe and Kevin Murray, *Why School Report Cards Merit a Failing Grade* (University of Colorado: National Education Policy Center, 2015), 5.
- ¹⁷⁷ House of Commons Children, Schools and Families Committee, *School Accountability: First Report of Session 2009–10*, 8.
- ¹⁷⁸ Ibid.

-
- ¹⁷⁹ Ibid., 10.
- ¹⁸⁰ Ibid., 9.
- ¹⁸¹ Attainment 8 measures students' average achievement across the same eight subjects as Progress 8: English; Mathematics; three other English Baccalaureate (EBacc) subjects (sciences, computer science, geography, history and languages); and three further EBacc or non-EBacc subjects.
- ¹⁸² National Association of Headteachers, *Improving School Accountability: Report of the NAHT Accountability Commission* (West Sussex: National Association of Headteachers, 2018), 11.
- ¹⁸³ Ofsted, 'Welcome to Parent View', Webpage, 10 July 2018.
- ¹⁸⁴ Ibid.
- ¹⁸⁵ Ofsted, *Annual Parents Survey 2017: Parents' Awareness and Perceptions of Ofsted*.
- ¹⁸⁶ Ofsted, 'Welcome to Parent View'.
- ¹⁸⁷ Ofsted, 'Ofsted Parent View: Management Information', Webpage, 31 May 2018.
- ¹⁸⁸ Ofsted, *Annual Parents Survey 2017: Parents' Awareness and Perceptions of Ofsted*.
- ¹⁸⁹ Ibid.
- ¹⁹⁰ Xuejiao Joy Cheng and Kurt Moses, *Promoting Transparency through Information: A Global Review of School Report Cards* (Paris: UNESCO, 2016).
- ¹⁹¹ HM Government, *Education and Inspections Act 2006*, Part 4.
- ¹⁹² Ofsted, *Annual Parents Survey 2017: Parents' Awareness and Perceptions of Ofsted*.
- ¹⁹³ Ofsted, *School Inspection Handbook: Handbook for Inspecting Schools in England under Section 5 of the Education Act 2005*, 20.
- ¹⁹⁴ Ofsted, *School Inspection Handbook: Draft for Consultation*, 17.
- ¹⁹⁵ John Roberts and Martin George, 'Exclusive: Ofsted Looking at No-Notice School Inspections... Again', *Times Educational Supplement*, 16 April 2018.
- ¹⁹⁶ Ofsted, *The Annual Report of Her Majesty's Chief Inspector of Schools - Part 2: Quality and Service in Inspection* (London: Her Majesty's Stationery Office, 1995).
- ¹⁹⁷ Ofsted, 'Ofsted: A Force for Improvement'.
- ¹⁹⁸ Ofsted, *Annual Teachers Survey 2018: Teachers' Awareness and Perceptions of Ofsted*.