
A Tale of Two Interpretations: Ofsted's expectations re-examined

COLIN RICHARDS

ABSTRACT Since September 2015 the Office for Standards in Education (Ofsted) has introduced major changes to its inspection procedures and expectations. These are embodied in its handbook for school inspection. Ofsted claims more than it can deliver; in particular it makes impossible demands on its inspectors – in terms of applying both evaluation criteria and grade descriptors. It also raises unrealistic expectations of and demands on schools.

Possibility or Impossibility?

Ofsted claims with some justification that as an organisation it is now more transparent than it was. It not only makes its common inspection framework publicly available but also the handbook its inspectors are expected to use when inspecting schools. This handbook is very detailed; the August 2015 version runs to 74 pages and is periodically updated with supplementary guidance – again publicly available. This is all to the good provided that the detailed guidance makes reasonable demands on its inspectors and that schools can be confident that inspectors are meeting those expectations. But does it and can they?

Closer scrutiny of the inspection documentation raises very serious doubts about the possibility of conducting school inspections in line with Ofsted requirements. I argue here that school inspection *is* impossible under the current inspection regime because of the sheer, bewildering number of highly complex and interrelated judgements that inspectors are required to make in a very limited period of time. It simply cannot be done. Anyone who believes it is possible is like the Queen of Hearts in Alice in Wonderland who could believe six impossible things before breakfast. I find it puzzling that so many inspectors, school leaders and politicians continue to believe, or at least pay lip service to, the impossible.

The Ofsted Inspection Handbook: outline of its purpose and content

The introduction to the 2015 inspection handbook begins:

This handbook describes the main activities undertaken during inspections of schools in England under section 5 of the Education Act 2005. It sets out the evaluation criteria that inspectors use to make their judgements and on which they report.

The handbook has two parts:

– Part 1. How schools will be inspected

This contains information about the processes before, during and after the inspection.

– Part 2. The evaluation schedule

This contains the evaluation criteria inspectors use to make the graded judgements about schools and includes exemplification of the kinds of evidence and activities used by inspectors to make their judgements.

This handbook is primarily a guide for inspectors on how to carry out school inspections. However, it is made available to schools and other organisations to ensure that they are informed about the process and procedures of inspection. It seeks to balance the need for consistency in inspections with the flexibility required to respond to the individual circumstances of each school. This handbook should not be regarded as a set of inflexible rules, but as an account of the procedures that govern inspection. Inspectors will exercise their professional judgement when using it. (p. 4)

Note that the handbook ‘sets out the evaluation criteria that inspectors *use*’ – not ‘sets out the criteria that inspectors *may* use’ nor ‘sets out *some* of the criteria inspectors *can* use’. Taken literally (and how else should it be taken?), these are the criteria Ofsted expects its inspectors to use when making judgements.

Elsewhere in the handbook there are equally definitive statements as to the factors ‘inspectors *will* consider’ when judging the effectiveness of leadership and management or how ‘inspectors *will* make a judgement on the effectiveness of teaching, learning and assessment’. The common inspection framework states that ‘inspectors *will* make a judgement on the personal development, behaviour and welfare of children and learners’ by how far the provision promotes a series of outcomes. Again, note the use of the word ‘*will*’, not ‘*may*’ or ‘*can*’.

One Possible but ‘Impossible’ Interpretation of the Handbook’s Requirements

To explore the possibility of meeting all the evaluation criteria listed in the handbook let us take the case of a typical one-form-entry primary school ‘requiring improvement’ and therefore needing inspection by two inspectors

over a two-day period. What do these two inspectors have to come up with by the conclusion of the inspection?

Ofsted requires its two inspectors to make five overall judgements plus two major contributory ones. They need to make judgments of:

- (a) overall effectiveness;
- (b) effectiveness of leadership and management;
- (c) quality of teaching, learning and assessment;
- (d) personal development, behaviour and welfare;
- (e) outcomes for pupils.

In addition the inspectors must evaluate:

- (f) the effectiveness and impact of the provision for pupils' spiritual, moral, social and cultural development;
- (g) the extent to which the education provided by the school meets the needs of the range of pupils.

At first sight that does not seem unreasonable but the devil lies in the detail of the contributory judgements the inspectors need to make to support and justify the overall ones. The inspection handbook spells out in detail a multitude of factors that the 'inspectors will consider' when coming to an overall judgement.

Take leadership and management, for example. Ofsted requires our two inspectors to gather evidence and carefully consider it in relation to a large number of interrelated but separate judgements detailed as follows:

In making this judgement in schools, inspectors will consider:

- the leaders' and governors' vision and ambition for the school and how these are communicated to staff, parents and pupils
- whether leaders and governors have created a culture of high expectations, aspirations and scholastic excellence in which the highest achievement in academic and vocational work is recognised as vitally important
- whether leaders have the highest expectations for social behaviour among pupils and staff, so that respect and courtesy are the norm
- the rigour and accuracy of self-evaluation and how well it leads to planning that secures continual improvement
- the design, implementation and evaluation of the curriculum, ensuring breadth and balance and its impact on pupils' outcomes and their personal development, behaviour and welfare
- how well the school supports the formal curriculum with extra-curricular opportunities for pupils to extend their knowledge and understanding and to improve their skills in a range of artistic, creative and sporting activities
- how effectively leaders use the primary PE and sport premium and measure its impact on outcomes for pupils, and how effectively governors hold them to account for this

- how well the school prepares pupils positively for life in modern Britain and promotes the fundamental British values of democracy, the rule of law, individual liberty and mutual respect for and tolerance of those with different faiths and beliefs and for those without faith
- the effectiveness of the actions leaders take to secure and sustain improvements to teaching, learning and assessment and how effectively governors hold them to account for this
- how well leaders ensure that the school has a motivated, respected and effective teaching staff to deliver a high quality education for all pupils and how effectively governors hold them to account for this
- the quality of continuing professional development for teachers at the start and middle of their careers and later, including to develop leadership capacity and how leaders and governors use performance management to promote effective practice across the school
- how effectively leaders and governors monitor the progress of groups of pupils to ensure that none falls behind and underachieve, and how effectively governors hold them to account for this
- how well leaders and governors engage with parents, carers and other stakeholders and agencies to support all pupils
- how effectively leaders use additional funding, including the pupil premium, and measure its impact on outcomes for pupils, and how effectively governors hold them to account for this
- the effectiveness of governors in discharging their core statutory functions
- how well leaders and governors promote all forms of equality and foster greater understanding of and respect for people of all faiths (and those of no faith), races, genders, ages, disability and sexual orientations (and other groups with protected characteristics), through their words, actions and influence within the school and more widely in the community
- the effectiveness of safeguarding
- the effectiveness of leaders' and governors' work to raise awareness and keep pupils safe from the dangers of abuse, sexual exploitation, radicalisation and extremism and what the staff do when they suspect that pupils are vulnerable to these issues. (pp. 36-37)

It is important to note that though there are 18 bullet points, our two inspectors need to make more than 18 judgments since many of the bullet points are composite. As an example, 'How well leaders and governors engage with parents, carers and other stakeholders and agencies to support all pupils' involves consideration of (a) how leaders engage with parents and carers; (b) how leaders engage with other stakeholders and agencies; (c) how governors engage with parents and carers; and (d) how governors engage with other

stakeholders and agencies. For each of (a) to (d) the inspectors need to collect evidence and make judgements.

To take a second example, 'Inspectors will consider the design, implementation and evaluation of the curriculum, ensuring breadth and balance and its impact on pupils' outcomes and their personal development, behaviour and welfare'. To render just this one sub-judgement credible, the two inspectors would need to engage in all or most of the following: scrutiny of the school's curriculum policy, its timetables, its self-evaluation, its development plan; pupils' outcomes, discussion with senior and middle leaders, teachers, students and, ideally governors and parents, including eliciting their views on what constitutes 'breadth' and 'balance'. All of this takes time to collect and time to consider.

On my count the total number of judgements for leadership and management come to at least 29, though it's a matter of interpretation as to the exact number of judgements implied by each bullet point. Readers may wish to compare their overall tally with mine. Their count may not come to 29 but it will undoubtedly constitute a large number. That's a considerable (impossible?) call on our two inspectors' time and expertise, all within the constraints of a two-day inspection.

Similarly, on my count, 33 sub-judgements (most of them multifaceted) have to be made by the two inspectors and credible evidence considered in relation to the section on 'teaching, learning and assessment'. As just one example from many, they are required to evaluate 'the extent to which the teachers' standards are being met'. Remember that there are eight class teachers in our hypothetical one-form entry school, each with their 42 standards to be met. To offer a credible, well-founded sub-judgement the evidence that the two inspectors will need to collect should include some or most of the following: classroom observation of teachers (but no grading of course); discussion with teachers and possibly teaching assistants, pupils and parents; reports from, and discussion with, senior and middle leaders; anonymised appraisal documentation; school and department self-reviews; lesson observation documentation; school policy statements on teaching, learning and performance management. Again, how much time would be needed to collect and consider this evidence? That alone could constitute more than two days' inspection activity. Yet there are 32 other factors that the two inspectors need to consider and gain credible evidence for! How possible is that?

The same impossible demands on the inspectors' time and expertise are made in relation to 'personal development, behaviour and welfare' where 14 sub-judgements (many also multifaceted) have to be made after collection and consideration of evidence. To take one example, how much time and patience (let alone tact) would they need to spend to arrive at a credible judgement of how well the school promotes and supports children's 'understanding of how to keep themselves safe from relevant risks such as abuse, sexual exploitation and extremism, including when using the internet and social media'? But there are at

least 13 other sub-judgements, some of them complex and sensitive, which they need to get evidence for, and to make, in that two-day period!

The same points could be made and similar examples quoted in relation to 'outcomes for pupils' where 19 complex sub-judgements (most of them multifaceted and concerned with data interpretation) would need to be made by our two primary school inspectors.

On my count (and that's disputable given the vague nature of many of the inspection criteria, which can be variously interpreted) this adds up to over 100 judgements to be made by our two inspectors, each requiring the collection, collation and consideration of evidence – all in no more than two days. *That is an impossible undertaking.*

A Second Possible but Still Impossible Interpretation of the Handbook's Requirements

There is a second possible interpretation of the handbook's requirements which at first sight seems more possible to implement than the interpretation above. For each of the key judgements – on leadership and management; quality of teaching, learning and assessment; personal development and welfare; and outcomes for pupils – the handbook provides a set of four 'grade descriptors' ('outstanding', 'good', 'requires improvement' and 'inadequate'), each with grade criteria, which our two inspectors can use to create a 'best fit' to the school being inspected. They can then use the grade criteria and descriptors to arrive at their four key judgements – which in turn they can use to arrive at their judgement on the overall effectiveness of a school.

To explore the possibility of using grade descriptors, let us consider what is involved in our inspectors coming to a judgement as to whether the school fits the grade descriptor for 'good' leadership and management.

The grade descriptor for a school with 'good' leadership and management reads:

Good (2)

- Leaders set high expectations of pupils and staff. They lead by example to create a culture of respect and tolerance. The positive relationships between leaders, staff and pupils support the progress of all pupils at the school.
- Leaders and governors are ambitious for all pupils and promote improvement effectively. The school's actions secure improvement in disadvantaged pupils' progress, which is rising, including in English and mathematics.
- Leaders and governors have an accurate and comprehensive understanding of the quality of education at the school. This helps them plan, monitor and refine actions to improve all key aspects of the school's work.

- Leaders and governors use performance management effectively to improve teaching. They use accurate monitoring to identify and spread good practice across the school.
- Teachers value the continuing professional development provided by the school. It is having a positive impact on their teaching and pupils' learning. Teaching is consistently strong across the school or, where it is not, it is improving rapidly.
- Governors hold senior leaders stringently to account for all aspects of the school's performance, including the use of pupil premium, the primary PE and sport premium and SEN funding, ensuring that the skilful deployment of staff and resources delivers good or improving outcomes for pupils.
- The broad and balanced curriculum provides a wide range of opportunities for pupils to learn. The range of subjects and courses helps pupils acquire knowledge, understanding and skills in all aspects of their education, including the humanities and linguistic, mathematical, scientific, technical, social, physical and artistic learning. This supports pupils' good progress. The curriculum also contributes well to pupils' behaviour and welfare, including their physical, mental and personal well-being, safety and spiritual, moral, social and cultural development.
- Leaders consistently promote fundamental British values and pupils' spiritual, moral, social and cultural development.
- Leaders promote equality of opportunity and diversity, resulting in a positive school culture. Staff and pupils work together to prevent any form of direct or indirect discriminatory behaviour. Leaders, staff and pupils do not tolerate prejudiced behaviour.
- Safeguarding is effective. Leaders and staff take appropriate action to identify pupils who may be at risk of neglect, abuse or sexual exploitation, reporting concerns and supporting the needs of those pupils.
- Leaders protect pupils from radicalisation and extremism. Staff are trained and are increasingly vigilant, confident and competent to encourage open discussion with pupils.

Minimally our two inspectors have to compare the school with each of the 11 sub-descriptors (bullet-points) above and come to a judgement about the 'degree of fit'. But note. As with evaluation criteria, many of these sub-descriptors are composite.

Take the first bullet point, for example: 'Leaders set high expectations of pupils and staff. They lead by example to create a culture of respect and tolerance. The positive relationships between leaders, staff and pupils support the progress of all pupils at the school'. This requires inspectors to consider (a) leaders' expectations of pupils; (b) leaders' expectations of staff; (c) whether leaders lead by example; (d) whether the school has a culture of respect and

tolerance; (d) whether *all* (!) pupils make good progress; and (e) how far the positive relationships support that progress. That makes six points to consider and collect evidence for.

Take the last bullet point: 'Leaders protect pupils from radicalisation and extremism. Staff are trained and are increasingly vigilant, confident and competent to encourage open discussion with pupils'. This requires inspectors to consider (a) the steps leaders are taking, and have taken, to protect pupils from radicalisation and extremism; (b) the nature and effectiveness of staff training; (c) how vigilant staff are; (d) how confident they are in this highly sensitive area; (e) the kind of discussion engaged in with pupils; (f) staff competence in engaging in that discussion. That also adds up to six points to consider and collect evidence for.

And yet there are another 11 bullet points related to leadership and management which our inspectors have to consider, each requiring a number of judgements to be made about their various components on the basis of credible evidence.

This example focuses on 'good' leadership and management but each of the four key judgements the two inspectors need to make to judge the overall effectiveness of a school requires consideration of grade descriptors, each with its sub-descriptors (bullet points), most of them composite ones. For our one-form-entry primary school to be judged 'good' for teaching, learning and assessment the inspectors need to consider 11 composite bullet points; for it to be judged 'good' for personal development, behaviour and welfare the inspectors need to consider 14; and for its outcomes to be 'good' they need to consider 7. And remember each bullet point requires a number of judgements, along with evidence, to be made of its components – all in the space of two days.

This analysis strongly implies that using grade descriptors would involve our two inspectors in making an impossibly large number of judgements and amassing far more evidence than is possible under the constraints of a two-day inspection.

Ofsted's Wriggle Room?

In facing the kind of criticism outlined in this article Ofsted has left itself with a little 'wriggle room' with its two statements in the handbook's introduction: (1) 'This handbook should not be regarded as inflexible rules' and (2) 'Inspectors will exercise their professional judgement when using it'. Presumably it means that some at least of the 74 pages of guidance do not have to be followed literally, word for word, but applied flexibly. Does Ofsted expect that flexibility to apply not just to detailed inspection procedures and processes but also to *all*, *some* or *a few* of the evaluation criteria or *all*, *some* or *a few* of the components of the grade descriptors? Are there no limits to the discretion accorded to inspectors as they apply the handbook to the circumstances in the school being inspected? Can inspectors discard whatever evaluation criteria or components of

the grade descriptors they like? Isn't that a recipe for arbitrary judgement and inconsistency across inspections? There *is* a strong argument for an element of professional discretion in relation to how to conduct an inspection in the particular circumstances of a school, but should that discretion extend to deciding which criteria/descriptors to apply and which to ignore?

Believing the Impossible

To illustrate my argument I have used the example of a one-form-entry primary school but exactly the same issues obtain whatever the size or type of school. With its expectations that have grown Topsy-like over the years Ofsted is no longer living in the real world and is demanding the impossible of its inspectors. As in Alice in Wonderland, inspectors have to believe and do so many impossible things – not before breakfast, but before the end of an inspection. I challenge anyone from Ofsted to demonstrate the feasibility of what is expected of ordinary mortals, however well trained, in the inspection handbook. A few months ago Ofsted's National Director tweeted to his colleagues: 'Don't stop thinking like a good teacher/leader. Remember what's practically possible'. Yet on either of the two interpretations discussed in this article, meeting the requirements of the inspection handbook *is* 'practically impossible'.

A Possible Way Forward

As a believer in the importance of inspection, properly conducted and with realistic expectations and demands, I hope that this critique will strengthen, not weaken, school inspection by stressing the unrealistic nature of Ofsted's current demands on both schools and inspectors and by arguing for their reformulation to make them more acceptable and more feasible to implement.

In what way should Ofsted reformulate its expectations? Paradoxically, it should begin by increasing the number of overall judgements inspectors need to make from five to eight. It makes no sense to try to make a single overall judgement in relation to triads such as 'quality of teaching, learning and assessment' or 'personal development, behaviour and welfare'. Separate overall judgements need to be made about each of the triads' very important components. However, there is no need for an 'overall effectiveness' judgement in addition to the eight, especially if a change is made to the grading scale used.

The four-point grading scale from 'outstanding' to 'inadequate' has long since lost its usefulness and has been the cause of endless angst experienced by both teachers and inspectors. Each of the eight major areas of a school's activities need simply to be judged as 'good enough' or 'not good enough', along, of course, with a summary of evidence to back up that binary judgement. So there is no need for an overall effectiveness judgement; the profile of the eight overall judgements will indicate the scope of a school's effectiveness.

The real economies of effort – the rendering of the ‘possible’ from the ‘impossible’ – can only be made by drastically reducing the number of judgements inspectors need to make. Through discussion within Ofsted and genuine consultation with schools and other stakeholders it should be possible to achieve a consensus (unanimity will be impossible!) as to the much smaller number of judgements inspectors need to make to arrive at the overall judgements.

To take a hypothetical example, the 29 sub-judgements (many of them multifaceted) related to ‘leadership and management’ might be reduced to four or five absolutely central ones. Or to take another example – in this case related to grade descriptors – the 11 bullet points (sub-descriptors) for ‘good’ leadership and management might be reduced to a similar number. It is true that many aspects of leadership and management currently considered important would have to be omitted from consideration but that is the case currently too – not every facet of leadership and management is captured in Ofsted’s handbook of inspection. However, the pruning would enable inspectors to focus their efforts on the collection, collation and consideration of evidence to make both their more limited number of judgements and their overall judgement on leadership and management far more credible and well founded.

If reductions (some of them more substantial than those suggested for leadership and management)) were made in relation to the other seven overall judgements the total number of judgements needing to be made could be reduced to manageable proportions.

I suggest these changes as an interim measure to make the ‘impossible’ more possible in the short to medium term. After a quarter of a century in existence, Ofsted requires a fundamental review building on, but going well beyond, the changes introduced since September 2015. The limited proposals set out here are but preliminary forerunners of the reforms necessary to create an inspection system ‘fit for purpose’ and, in the words of Ofsted’s National Director, ‘practically possible’.

COLIN RICHARDS is currently Emeritus Professor of Education at the University of Cumbria and has been a visiting professor at the Universities of Leicester, Warwick and Newcastle. A former primary school teacher and deputy head, he was an HMI from 1983 to 1996 specialising in primary education and in teacher education and holding the post of Staff Inspector (Curriculum 5-16) followed by that of Ofsted’s Specialist Adviser for Primary Education. Since leaving Ofsted he has maintained a keen but critical interest in the issues of standards, primary and secondary education, the school curriculum, governance and inspection. He is chair of governors of a Cumbrian secondary school which has voted twice not to seek academy status. He has served on a number of primary school governing bodies and has recently stood down as a National Leader of Governance. As a critic of much (though not all) of past and present

government education policy he is a frequent contributor to the national press (in particular to the *Guardian*, the *Observer*, the *Independent*, the *Times Educational Supplement* and *Schools Week*). He tweets at @colinsparkbridg and can also be contacted at profcrichards@gmail.com.

