

Responsive not coercive inspection

A restoration and reinterpretation of the values, principles and procedures of HM Inspectorate (abolished in 1992)

Colin Richards

Abstract

After widespread concern following the tragic death of a headteacher, the appointment of a new chief inspector for Ofsted and the prospect of a Labour government offer an opportunity for a radical appraisal of the current inspection regime. This article discusses some basic concepts and principles which might contribute to a system that is responsive, not coercive, and one that is infused by ‘a duty of care’.

Keywords: Ofsted; HMI; judgment; duty of care; quality of education

In 1992, Ofsted was set up to assess the quality of education in each school (and through that to assess the state of the English school system as a whole). It was set up partly to inform parents and other stakeholders. It was not established primarily or directly to improve schools or the system generally. It aimed to provide impartial, independent, professionally subjective judgments to assist leaders, governors, teachers and local authority officers in evaluating their policies and provision so as to help determine priorities for development, including sustaining good elements and working towards improvement in others. It was for schools to consider the inspection findings and recommendations, not for schools to slavishly accept them all.

Thirty years on, that original limited but valuable purpose has been forgotten, even perhaps corrupted and overtaken by more grandiose claims. Ofsted is viewed by its leaders and supporters as providing authoritative, ‘objective’, incontestable findings and recommendations which have to be acted upon. Its detractors contest these claims as unwarranted and overblown. Any reform of the current system would do well to recognise the value of the more limited purposes set out in the original legislation.

What follows is a personal re-interpretation and updating of an approach to school inspection undertaken by HM Inspectorate before its abolition in 1992.¹ It is offered as set of principles, concepts, insights and procedures which could be used to underpin a redesigned inspection system post-Ofsted which, faithful to ‘a duty of care’, is responsive not coercive, and which takes more cognisance of the subtleties and uncertainties involved in trying to assess the quality of education in a timely and sensitive manner. It offers no neat, detailed, prescriptive blueprint or framework to be readily adopted by any successor body to Ofsted, but proposes a set of principles underlying an inspection

system defensible in terms of the possibilities and limitations involved in the exercise of professional judgment in educational contexts.

The nature of professional judgment

More so than in successive Ofsted formulations, any renewed approach to school inspection needs to recognise and make central the exercise of professional judgment. That is easy to say but very difficult to characterise. Ofsted made no attempt to characterise it, despite making sporadic and ritualistic references to its importance. Never quoted by Ofsted, Geoffrey Vickers got close to it in his book *The Art of Judgment*. He valuably distinguished two aspects of that kind of overall appreciative judgment which school inspectors have to make – reality judgments involving facts about the state of any system or organisation such as a school being inspected, and value judgments ‘making judgments about the significance of these facts’.² Note: Vickers characterised appreciative judgment as an ‘art’, not as a science, nor as an art informed by science – as some Ofsted publications vainly tried to do.

In an inspection, reality judgments are derived from observations and discussions with leaders, teachers, pupils and governors. Such judgments can seem to be firmly rooted in objective reality, but this not the strictly the case; crucially, these judgments are mediated through inspectors’ past experience and values. Different inspectors may legitimately detect and report facts differently. Reality judgments cannot be characterised as totally objective.

Similarly, the judgments that inspectors inevitably have to make about the value of what they observe or deduce ‘cannot be *proved* correct or incorrect; they can only be *approved* as right or *condemned* as wrong by the exercise of another value judgment’³ – this makes any quality judgment inevitably contestable. In any reconceptualisation of inspection post-Ofsted, the notion of objectivity needs to be replaced by that of ‘value-informed judgment’ or ‘value-informed appreciation’. All this implies that inspection cannot, and should not, claim to be any more than the professional subjective judgment of a group of experienced, expert observers. As such, the findings of any inspection are open to interpretation and never definitive. Post-Covid, post-Spielman inspections should respect, not conceal, the uncertainties of the judgmental process.

To minimise but not remove those inevitable uncertainties, inspections need to rely on the collective, moderated judgment and experience of the inspectors. As Vickers stressed, appreciative judgment and subsequent decision-making are part of a social, collective process. They are taken within, and depend on, a net of professional communication, which is meaningful only through a vast, partly organised accumulation of largely shared professional assumptions and expectations constantly being developed and changed by the activity of inspection itself. Collective judgment-

making needs to be based on participants' wide experience of a variety of institutions in different educational contexts nationwide. It is 'forged' or metaphorically 'hammered out' through lengthy discussion and deliberation with other similarly experienced colleagues during the period of the inspection. The result is a collective, unique and internally moderated set of judgments – an appreciation, not a set of off-the shelf criteria imported from elsewhere. The notion of collective, 'hammered out' appreciative judgment is crucial. No published report, especially one determining or affecting a school's future, should be the work of one individual alone. Even in the smallest school or setting, an individual's judgements need to be moderated with the perspectives of at least one other inspector to arrive at a defensible, moderated judgment of the whole school or setting. Any replacement system for Ofsted needs to meet that requirement in the interests of the school and of the inspectors themselves.

Practising professional judgment

In his book *The Educational Imagination*,⁴ Elliot Eisner argued that evaluating educational provision is more akin to making aesthetic judgments than to conducting a scientific enquiry. Drawing from his experience of English education in the 1970s, he memorably characterised inspection by HM Inspectorate as a form of educational connoisseurship, not bound by clear-cut, straightforward, incontestable criteria but very valuable nonetheless. 'Connoisseurship' is a flattering term, not claimed as such by those inspectors observed by Eisner, but it gets closer to the reality of the inspection process than viewing it as a straightforward process of directly comparing practice with a set of statements listed in an inspection handbook or on a crib-sheet.

In a different context, the philosopher Ludwig Wittgenstein described how a connoisseur of art makes aesthetic judgments, and in doing so got close to helping understand the nature of inspection judgements, how they are acquired and how they are justified. He commented:

We learn certain things only through long experience and not from a course in school. How, for instance, does one develop the eye of a connoisseur? Someone says, for example, 'This picture was not painted by such-and-such a master'. He may not be able to give any good reasons for his verdict. How did he learn it? Could someone have taught him? Yes – not in the same way as one learns to calculate. A great deal of experience was necessary. That is, the learner probably had to look at and compare a large number of pictures by various masters again and again. In doing this he could have been given hints. Well, that was the process of learning. But then he looked at a picture and made a judgment about it. In most cases he was able to list his reasons for his judgment, but generally it wasn't they that were convincing ... The value of the evidence varies with the experience and the knowledge of the person providing

it, and this is more or less the *only* way of weighing such evidence since it cannot be evaluated by appeal to any system of general principles or universal law.⁵

Applying these insights to inspection, the required expertise cannot be acquired simply from ‘a course’, from a period of training, from an aide-memoire or from limited experience in a small number of schools. It involves learning from a wide range of teaching and inspection experience in a variety of relevant contexts, national as well as regional or local. It involves looking at, and comparing, a large number of educational activities by ‘various masters again and again’. It is not like learning from an inspection rule book, tick list or aide-memoire. It involves learning from others more experienced in making judgments of teaching quality who can ‘hint’ at what is required and who can discuss the complexities and intangibles of classroom observation. Like Wittgenstein’s connoisseurs, inspectors should be able to ‘list reasons’ for their judgments but, as he commented, these can never be absolutely ‘convincing’ given the difficulties and complexities involved. The value of the judgments and the evidence they use to back them up depends essentially on the experience and knowledge of the persons making them. To repeat Wittgenstein’s comment, ‘this is more or less the *only* way of weighing such evidence since it cannot be evaluated by appeal to any system of general principles or universal laws’ – whether these are enshrined in an inspection handbook, in subsidiary guidance or an aide-memoire.

Such a characterisation of inspection as a kind of aesthetic appreciation is a far cry from the formulaic, rule-bound, supposedly objective scrutiny by an organisation such as Ofsted.

The reliability and validity of inspection judgments

One of the major and most common criticisms levelled at Ofsted judgments has been their lack of reliability and validity. Teachers’ professional organisations, professional interest group and academics such as Coffield have offered this critique.⁶ Ofsted responded by claiming to use social-science research methods to improve the validity and reliability of its judgments. But can notions of validity and reliability be straightforwardly applied to the making and justification of such educational judgments?

To answer this question, let us consider another area of activity – theatre criticism – which is in many ways analogous to school inspection. Theatre critics observe and appraise a performance or run of performances, as school inspectors do. They judge the quality of the acting, production and direction; likewise, inspectors judge the quality of teaching, curriculum and school leadership. Critics judge how far the performance reflects the content and intentions of the text; similarly, under the current Ofsted framework inspectors comment on the rationale and implementation of the ‘text’ of the curriculum. Critics assess the reactions of the audience; likewise,

inspectors assess students' responses.

Critics judge the quality of what they see. So do inspectors. Critics do not, and cannot, measure what they see on any numerical scale, nor can inspectors. Critics make their judgments based largely on their experience of similar, though never identical, productions; likewise, their inspectorial counterparts relate their judgements to analogous experiences in other schools. The criteria theatre critics use are largely intuitive, impressionistic and cannot be reduced to a checklist of clear, unambiguous components.

It does not make sense to ask of theatre criticism that it be reliable and valid in the ways such terms are usually used in educational parlance. The same applies to school inspection. Both are value-laden enterprises with the concept of 'quality' at their heart and thus subject to a different kind of logic than educational measurement. Academic and professional critics need to recognise this. So does any successor to Ofsted.

Some limitations of inspection judgments

Because of the complex mix of reality and value judgments involved in the act of educational connoisseurship, an inspection team can never claim that their interpretation of a school is the definitive one. Nor can they claim that their unique set of judgments can be directly or robustly compared with the equally unique set of judgments of a school in a different context or even with the judgments of the same school (which never remains 'the same school') inspected at a different time. Each set of inspection judgments is in a sense *sui generis*. Direct comparison of inspection judgements over time or from 'similar' school to school is at best highly problematic and at worst totally invalid. Yet Ofsted persisted in this practice without questioning it. Its successor needs to revisit this issue.

Inspectors can only report, interpret and evaluate activities seen at a particular point in time. They cannot comment with any plausibility on what has happened in the past or predict what will happen in the future. In particular, they cannot comment with any authority or conviction about progress over time, whether by groups of students or by the school as a whole, since they were not present to observe or assess it. Admittedly they may have past documentation or a past inspection report to refer to; but they do not have access to their predecessors' assumptions, expectations or deliberations for comparison, nor can they know with any certainty what has transpired in the interval between inspections. Performance data from the past may be available but such data are fallible, contestable, variously interpretable and only very partial as indicators of progress, performance or quality. Such data can only be interpreted in the light of close knowledge of the context in which they were generated yet this is denied the inspectors using them at a future date. As a consequence, the judgments inspectors make can only

be as 'they seemed to them at the time'. That does not mean that the judgements are not valuable in the short-to-medium term as a basis for professional reflection and development and for communication with governors, parents and other stakeholders. Such reports can be, and in many cases have proved to be, very valuable as an 'outside' appreciation of a school's work, including those aspects requiring further work, as well as those considered praiseworthy by the inspection team. The time-specific, 'instant' nature of inspection judgments and inspectors' inability to comment meaningfully on progress, whether by the school or by its students, need to be more fully recognised in any re-valuation of inspection policy and practice post-Ofsted.

Inspecting the quality of teaching

At the heart of inspection is a professional judgment about the quality of teaching as reflected in pupils' learning. This was recognised in statute when Ofsted was established in 1992. More recently, it has become a contentious issue with Ofsted inspectors accused of favouring specific teaching methods. Under the most recent inspection framework, its centrality was downgraded and teaching came to be seen essentially as a vehicle for 'delivering' (rather than transacting) the content of the curriculum. Arguably, downgrading the making of judgments about the quality of teaching in a school endangers the professional credibility of inspection and threatens to leave performance data as the main source of evidence used in reporting on the quality of education. Any successor to Ofsted must re-examine how to deal with the inspection of teaching and learning.

In recent years, teachers' professional associations have expressed anxieties over inspectors' preferred teaching methods influencing their judgments. However, evaluating the quality of teaching need not, should not, involve looking for particular teaching methods and then gauging their effectiveness in terms of promoting learning. Rather the reverse. Inspectors should look for evidence of pupils' learning in terms of their observable responses to teaching and then should work back to highlight those factors that may have promoted, or hindered, their learning. This involves close observation and discussion with both staff and students. 'The unanticipated success of the wrong method' as judged by students' responses to the teaching they receive, needs to be recognised and celebrated. Similarly, 'the unanticipated failure of the right method' needs acknowledgment in future inspection reports.

Judgments about the quality of teaching in lessons and in the school as a whole are properly tentative and consequently should be offered as such in any feedback to those whose work has been observed. There is inevitably a considerable degree of inference involved in feedback, especially about the extent to which learning has taken place. There is inevitably too an element of professional judgment as to which features of a

lesson have contributed to, or inhibited, learning. That tentativeness is crucial to the context in which any feedback is being given. It offers the opportunity in dialogue for other tentative, evidence-based, interpretations to be offered by the teacher who has been observed.

(The same degree of tentativeness needs to be applied also to other aspects of provision, including the quality of leadership and management of heads and other senior and middle leaders.)

Inspecting the curriculum

Though featuring strongly in early Ofsted inspection frameworks, the curriculum had not been a major focus of attention in school inspections between 2010 and 2019. From 2019 it was seen as the core of what Ofsted termed ‘the quality of education’.⁷ Its Education Inspection Framework focused on three aspects: intent, implementation and impact. These were to be assessed as a result of discussion with school and subject leaders, scrutiny of documentation and observation of work in class. These are certainly important aspects of curriculum planning and management and they can be inspected with varying degrees of plausibility and certainty. *Intent* is probably the easiest to characterise in general terms, though deciding on whether the curriculum is ‘ambitious’ enough or whether it is ‘coherently planned and sequenced towards cumulatively sufficient knowledge and skills for future learning’⁸ is far from straightforward and inevitably shot through with value judgments which are far from uncontentious. *Implementation* involves appropriate sampling of lessons by inspectors with activities observed being matched with written or oral expressions of intent – a tricky but not impossible undertaking provided inspectors have the necessary time and subject expertise. *Impact* is the most uncertain to judge given the restricted timescale of a typical inspection. Talking with students about their progress could in theory be a major source of evidence, provided lengthy, in-depth discussions with a representative range of different groups and ages are built into inspection schedules, but these are all but impossible within the time constraints of whole-school inspections. Likewise, in-depth examination of students’ work, cross-referenced to evidence of *intent* and *implementation*, may be possible in theory, but very problematic and inevitably impressionistic in practice.

In its most recent iteration, the inspection framework controversially stresses the centrality of memory in learning and requires inspectors to assess the impact of a school’s curriculum by questioning pupils as to how much they remember from previous lessons. But despite Ofsted’s stress on memory and reliance on other aspects of so-called ‘cognitive science’, no one, whether primary-aged child or school inspector, can possibly be expected to have total, or near-total recall of what they have experienced

a week, a month or a year ago. The most that can be expected is partial recall, and even that requires sensitive prompting by the questioner who needs to know the context of the experience being discussed and, in particular, the language originally used to mediate that original experience. But how can an inspector possibly have this?

Then there are the conditions necessary for that sensitive questioning of children to take place. Exploring children's past learning in any depth requires time and a quiet atmosphere to delve beyond immediate answers to ascertain understanding rather than superficial recall. Can they always, or even often, be found within a crowded inspection setting?

All of this throws doubt on the value Ofsted places on interviewing children as a way of exploring the impact of the curriculum and on the emphasis it places on memory as a proxy for learning and understanding. And yet in report after report, inspectors have reported 'gaps' in children's knowledge as evidence that a school's curriculum did not have its intended impact.

Perhaps most significant of all – and never acknowledged by Ofsted – there was a major and fundamental lacuna in its approach to the school curriculum. It focused entirely on how well the curriculum was planned and implemented and on its apparent impact. It did not allow evaluation of the *worthwhileness* of what has been designed and implemented. It assumed that the current legally mandated national curriculum framework was both good and incontestable. It did not permit inspectors to comment on any inherent deficiencies in the content of the officially approved curriculum – only deficiencies in its planning and management by schools. This is deeply problematic. The rationale, aims, concepts and content all need to be considered *worthwhile* for a school's curriculum to be judged 'good', and inspection criteria should embody that value dimension. Ofsted's claim that it was able to report on the quality of the curriculum is thus a very partial and flawed one.

Inspection grading

The evaluation of teaching, the curriculum and other aspects of the school is inevitably qualitative: nothing speaks for itself, everything needs interpreting and that interpretation inevitably involves value judgments and the use of broad qualitative descriptors such as 'good', 'very good', 'excellent', 'satisfactory', 'reasonable', 'fair', 'poor' etc. In advance of an inspection there can be no stipulation as to which qualitative terms are to be used; the terms must 'fit' the perceptions of the activity or activities being evaluated. They cannot be reduced to just four numerical grades, as under the current Ofsted framework; reality is much more complex than a four-fold categorisation. The over-simplification fails to take into account the many-varied facets of educational reality which can only be captured (and then only in part) in well-crafted prose.

Inspection teams need the freedom to dispense with artificial, misleading constructs such as overall inspection gradings, and to present schools in their idiosyncratic variety with idiosyncratic descriptors to match. Each inspection report has to be bespoke – not a formulaic account with minimal variation from school to school. That qualitative richness needs to be built into a re-interpreted inspection system.

Inspection recommendations

No school, however notionally ‘outstanding’, is perfect. There is always more to learn from the experience of other schools, and inspectors can help bring that experience to bear in discussion with school leaders. Inspections should result in recommendations, not in diktats about ‘what the school needs to do to improve’. Inspectors should raise issues a school needs to consider, not necessarily to act on; that is a crucial distinction. However, there needs to be a professional obligation on the part of schools to respond publicly on how they have considered and responded to those recommendations, even if it is to reject them in part. This would reinforce a view of inspection as providing a set of provisional, tentative, time-specific judgments which inform, rather than necessarily override, the similarly provisional, tentative and time-specific judgments of staff, governors and parents. Providing recommendations for schools to consider, rather than to comply with, would serve to respect rather than undermine the professional judgment of staff.

Inspecting with a duty of care

Inspection is a very fraught and stressful process for school leaders and teachers, and also for inspectors. It needs to be conducted with ‘a duty of care’ by all parties.⁹ As Peter Hennessy points out (perhaps optimistically?). ‘it’s a phrase and concept we all understand, yet it is a fairly recent one in UK history’.¹⁰

He attributes the phrase to Lord Atkin, who characterised it in the following terms:

You must take reasonable care to avoid acts or omissions which you can reasonably foresee would be likely to injure your neighbour. Who then in law is your neighbour? The answer seems to be persons who are so *closely and directly* by my act that I ought reasonably to have them in contemplation as being so affected when I am directing my mind to the acts or omissions which are called in question.¹¹

This is particularly pertinent to inspection when comments made by inspectors can either reinforce or challenge teachers’ deep-felt identity of self with teaching – the result of their deeply emotional bond with their work (often as they would say, their ‘life’s work’). Inspectorial criticism of their professional practice, however minor or however justified, is so very often taken as a criticism of them as people.¹² Not only that,

but such negative comments are likely to be remembered and worried over for the rest of a teacher's professional life. Inspectors, including His Majesty's Chief Inspector, too often failed to recognise or forget that fact. They needed to tread warily and carefully, but too often did not.

It would be very easy to provide sensational examples (such as that of Ruth Perry) to illustrate the negative effects of Ofsted inspections, but the widespread extent of everyday distress before, during and after the event of an inspection is even more telling testimony, provided it is recognised for what it is – the result of an Ofsted system, not always sufficiently responsive to personal sensitivities. It is fair to say that such shortcomings have been mitigated to some extent by many, though not all, inspection teams, but a sense of alienation and of 'being done *to* rather than done *with*' have characterised so many teachers' reactions, even when their schools have received favourable inspection reports, let alone soul-destroying ones.

In a letter published in the *British Medical Journal*, Waters and McKee argue that Ofsted: should publicly accept that it has a duty of care to teachers (and to its inspectors, some of whom are also traumatised). 'We believe that it has such a duty, with failure to uphold it amounting to negligence. While Ofsted inspections place great emphasis on safeguarding by school staff, we have struggled to find evidence that Ofsted has reflected in detail on its own safeguarding responsibilities. It should also concede that it has lost the confidence of teachers and work with them to rebuild it'.¹³

To sum up, the teaching profession needs to be held accountable (as does Ofsted itself), but in a way which embodies a duty of care and which preserves both accountability and humanity. Any successor to Ofsted needs to consider what that duty of care involves and to monitor how far it characterises all inspection activities and pronouncements.

Moving forward

Following the appointment of a new chief inspector, and with the prospect of a Labour government, there will be strong arguments for the replacement of what many see as an adversarial inspection system with a more consensual one focused on commonly agreed principles and based on greater awareness of the nature of inspection as an appreciative process – tentative and provisional, though still very valuable when conducted in a context-sensitive fashion.

This article is intended to contribute to that much-needed debate about the future of school inspection post-Covid and post-Spielman. Teachers' unions, subject-associations, parents' groups, the Chartered College of Teaching and personnel from universities, local authorities and multi-academy trusts should be involved, along with Ofsted itself and the Department for Education. But until the chief inspector's replacement is in post,

Ofsted is unlikely to be willing, or to be able, to orchestrate a fundamental reappraisal that would retain the confidence of the various interests involved. Hopefully, her successor will be willing to engage. If not, then perhaps a Labour government will find a replacement more in tune with its educational mission.

The current negative, punitive mindset of Ofsted – and its corollary, the fearful, negative mindset of schools – need resetting as a matter of urgency. A period of three years or so free from the threat of routine inspection would create a period of calm reflection and renewal for the much-needed reconstitution of Ofsted to take place. His Majesty's Inspectors of Schools could continue to conduct national surveys on key issues and be on hand to inspect the very small number of schools where very serious problems are detected as a result of the surveys. That period of suspension would help to mitigate, even perhaps largely remove, the fear-ridden negative mindset that is inhibiting much-needed reform.

The papers in this issue of *FORUM* should help inform that much-needed review.

As Hennessy stresses in *A Duty of Care*, a post-Covid world necessitates the re-examination of a wide-range of previously held assumptions, policies and practices, not only in education but in other policy areas, such as health and social care. School inspection should not be an exception to that fundamental review

Colin Richards joined HM Inspectorate in 1983. He later became Staff Inspector (Curriculum), Staff Inspector (Primary Education) and, finally, Ofsted's Specialist Adviser for Primary Education. He treasures the description 'an old-fashioned HMI' bestowed on him by an HMCI who did not deem it a compliment.

profrichards@gmail.com

Notes

1. Colin Richards, 'Thoughts on whole-school inspection post-Covid', *London Review of Education*, 8(3), 2020, pp511-18.
2. Geoffrey Vickers, *The Art of Judgment: a study of policy-making*, London, Longman, 1983, p40.
3. Vickers, *op. cit.*, p71.
4. Elliot Eisner, *The Educational Imagination*, London, Collier-Macmillan, 1979.
5. Ray Monk, *Wittgenstein*, London, Granta Books, 2005, pp103-4.
6. Frank Coffield, *Will the leopard change its spots? A new mode of inspection for Ofsted*, London, UCL IoE Press, 2017.
7. Ofsted, *Education Inspection Framework*, London, Ofsted, 2019.
8. Ofsted, *op. cit.*, p90.

9. Sarah Waters and Martin McKee, 'Ofsted :a case of official negligence?', *British Medical Journal*, 381, 1147, 2023.
10. Peter Hennessy, *A Duty of Care*, London, Allen Lane, 2023, pxvi.
11. Hennessy, *op. cit.*, pxviii, my emphasis.
12. Colin Richards, 'The high price of inspection', *The Guardian*, 19 June 1997.
13. Waters and McKee, *op. cit.*, p1147.