Navigating uncertainty: a future direction for Ofsted?

Purpose of this discussion paper

Ofsted reform is often a live policy discussion but with the prospect of a new HMCI in January 2024 and a general election in the same year, it is almost inevitable that there will be heightened debates in the next twelve months about the future role and shape of the inspectorate.

This paper is intended to support such discussions by exploring key tensions and trade-offs that exist in inspection policy and practice. We think setting these out is an important first step before moving into the 'nuts and bolts' of what a future framework should look like. Too often calls are made for Ofsted to focus on X or to look beyond Y, but the broad paradigm – and the concerns sometimes expressed about inspection - remain the same because the issues outlined in this paper are not always properly understood. We think that exploring these tensions as a starting point creates the space to think truly differently and constructively about the future of inspection.

Accordingly, this is a discussion paper and not a list of firm proposals. We hope it offers some useful concepts to support thinking about the future of inspection policy and practice. In the coming months we will be engaging more with CST members about these concepts and the potential implications for inspection reform.

Ofsted plays an important role in the current system, primarily:

- Providing assurance to Parliament and insight for regulators
- Providing insight for parents

Given these aims, it is important that inspection is considered to be valid by those who use inspection outcomes and those affected by them.

It is also important that the benefits of the current approach to inspection are weighed against the potential unintended consequences.

Concerns raised about inspection practices should not be summarily dismissed as producer interest. In a mature system, stakeholders are able to observe and take account of competing tensions while reaching a balanced view.

Autonomy and control

A state funded system with autonomous institutions (eg state funded independent schools – academies) requires some form of oversight in order to prevent failure and inform systemic responses to failure, such as rebrokering. This provides a good rationale for the continuation of a form of inspection.

In a high autonomy system, where individual institutions have a good degree of leeway in relation to curriculum, pedagogy, resources etc, it is arguably important that this system of oversight is

independent, rigorous and carries consequences for institutions found not be reaching particular standards. High stakes inspection is a way of exerting a form of control over standards and practices in a highly autonomous system.

However, high stakes inspection brings with it risks of systemic distortion, including:

- Anxiety and stress, particularly among leaders
- Inspection methodologies that unhelpfully or inappropriately influence school practice

In a sense, the 'control' imposed through a high stakes inspection & accountability system provides a counterbalance to the risks that can be associated with high autonomy. The balance, however, needs to be right.

Validity & reliability

Inspection judgements are inferences made by inspectors.

In order for an inference to be valid it must avoid (or limit) construct underrepresentation. For example, an inference about the curriculum should be based on a sufficient scope of things considered to make up the construct of 'the curriculum'. An inference about the quality of education must encompass a sufficient scope of the things deemed to make up the 'quality of education'.

This has implications for individual inspections – for example, where an inspection team fails to gather sufficient evidence to support the inference made.

It also has implications for how we view the broader underlying constructs, such as the 'quality of education', enshrined within the framework and handbook. Does the 'quality of education' sufficiently capture the intended construct? There is a view among some that the current conception of quality of education overplays specific elements of curriculum design.

One solution to this would be to provide counterbalance through a greater focus on other aspects of the construct which are currently underrepresented. There would of course be clear implications for the duration of inspection; inspecting more would require more time. And there are already reports that the current inspection methodology, especially under Section 8, is too rushed. Increasing inspection time would also require more inspectors (for which there is potentially a draw away from teaching and leading schools), as well as a financial cost. Given the current funding pressures it is hard to make the case that additional funding for inspection should be a priority.

Alternatively, the construct itself which is being assessed could be changed such that it more readily matches the available scope of evidence. For example, rather than calling the judgement 'quality of education', should it be called 'curriculum sequencing'?

Is the answer for inspection to move away from 'big' graded judgements of schools? It could be argued that smaller scale, more focused judgements would be harder for parents to navigate, but one might equally take the view that if the construct itself is not considered to be represented with validity in the graded judgement then there is little point in it. In simple terms, is an overall 'good' judgement a sufficiently valid construct as it stands?

Inspection frameworks over a number of years have struggled with the amalgamation of the available evidence into a single overriding – and valid – construct expressed as a graded judgement.

Another issue to consider: how confident can we be that the inference being made is not a reflection of something else in fact – known as construct-irrelevant variance? This occurs when random factors, other than the thing itself, impact the inference being made. So, although a judgement may be termed 'quality of education', might there be factors other than that construct at play (eg socio-economic factors, an individual inspector's preferences, the timing of the inspection etc). Where this is the case, construct-irrelevant variance works to undermine the validity of the inference being made.

This from Dylan Wiliam (on the topic of assessment) is instructive of the tensions at play:

"For a given amount of time, we can cast our net widely and get some not particularly reliable information about a large number of aspects, or we can focus our attention on much more limited aspects and get more reliable information…Sometimes we need a floodlight to get a perspective on a wide area, and sometimes we need a spotlight, getting clear information about a small area. What matters most is whether the trade-off between reliability and other aspects of validity is more or less appropriate for the particular situation."

Given the high stakes nature of Ofsted graded judgements, and the uses to which inspection judgments are put, it is important to get this balance right. Inspection tends to have a wide-ranging focus (the floodlight Wiliam refers to above), which may be appropriate for the purposes of inspecting a whole school, but arguably the judgements and conclusions emerging from inspection conceal the inherent trade-offs that exist when inspecting something as broad and complex as a school.

Inference and consistency - another way of thinking about validity and reliability

In order for a high stakes system to carry the confidence of stakeholders it must be seen as being consistent so that outcomes have a level of predictability and are not seen as arbitrary. These seems particularly important where the stakes are high and outcomes which feel arbitrary are likely to be even more problematic.

Such a requirement lends itself to an inspection system that prioritises consistency. A key means for achieving greater consistency is to reduce the degree of subjective interpretation (inference) that

has to be undertaken by inspectors. This is because it reduces the risk of construct-irrelevant variance (see above).

However, a lower inference methodology can bring a trade-off: it can reduce the scope of the inspection activity (which impacts on construct representation, as explained above).

Given the aims of inspection and its enactment as a public-facing undertaking, it may not be appropriate to adopt a high inference approach which risks issuing invalid judgements.

On the other hand, prioritising consistency by reducing inference may lead to an inspection system which lacks the breadth, depth and nuance necessary to provide useful information to stakeholders.

To some degree this can be seen in the differences between Ofsted's 2015 and 2019 inspection frameworks. Broadly, we might say the 2015 framework prioritised lower inference inspection by focusing more on published school performance data. Arguably, the cost of this approach was that inspection provided limited insight beyond what already existed in league tables.

In contrast, the 2019 Education Inspection Framework was established to reveal information which could not be seen in league tables, most notably curriculum quality, through a higher inference inspection model. But the trade-off in doing so may be that inspection outcomes are harder to predict and stakeholders may feel that outcomes are more arbitrary.

Resolving this tension can be difficult but possible solutions could include:

- Maintaining a higher inference model of inspection but reducing high stakes (because consistency is not sufficient to support high stakes)
- Move to a model of inspection where the nature of the outcome (judgment/grade/comment) is more in keeping with the degree of inference involved. For example, providing more cautious written comments, rather than graded judgements, on aspects where inspection methodology is higher inference.

Conclusion

The purpose of this short paper is to set out some of the key tensions at play within the paradigm of inspection. Collectively, these tensions reach beyond the specifics of particular frameworks and pose a deeper question: ***how certain can inspectors be about a school they inspect?***

Ofsted is in the business of drawing inferences and its graded judgements are the ultimate and enshrined expression of those inferences. The current paradigm of inspection asserts an absoluteness of judgement, one in which complexity and tentativeness are played down in favour of clarity and authority.

However, this approach risks overlooking the reality that our ability to be certain depends on the nature of thing we are drawing inferences about. For example, is it the case that inspectors can be more or less certain when judging behaviour than when judging personal development or the quality of education?

A counter argument sometimes made against establishing an inspection approach more comfortable with uncertainty is that parents want and need Ofsted to judge and communicate with absolute certainty and authority in order to inform them about matters such as school choice. However, recent research from Public First suggests that, particularly for working class 'Red Wall' parents, Ofsted's output is not as significant in informing decisions as policy makers might assume.

The lesson here is not necessarily that Ofsted's judgments and reports do not matter, but perhaps that parents' use of inspection outcomes does not require the level of certainty and starkness that is often assumed; it may be that a more exploratory approach to inspection might be just as useful to parents as one which appears to valorize a sense of certainty and confidence above all else.

The case made in this paper is not that the Ofsted's judgements are 'invalid' per se, but rather it asks whether they are all *sufficiently* valid – and whether inspectors can be sufficiently certain - for the purposes to which judgements are put, and whether the cost in terms of unintended consequence is too great.

It is occasionally suggested that Ofsted should abandon judgements altogether in favour of purely narrative reporting, or that Ofsted should judge some other aspect(s) of schools. This paper does not assert either of these positions. Rather, it makes a separate and more fundamental point: inspection should set off from the position that inferences made about schools, the expression of those inferences (including judgements), and the use of those inferences, should be appropriate to the level of certainty that can be obtained on the matter in question.

This speaks both to the ***culture*** and ***methodology*** of inspection. A more exploratory approach to inspection – one more comfortable navigating uncertainty - could manifest in a range of ways, including:

- How inspectors gather evidence

- How inspectors consider evidence
- How inspection conversations are conducted
- The type of judgements inspection produces
- How these judgements are communicated in conversations and public records such as reports

That there should be four or five graded sub-judgments and that these should aggregate to a single overall grade has become sacrosanct. This paper offers the view, free from prejudice of the outcome, that policy makers may find it fruitful to re-evaluate this assumption.

For discussion

Building on the concepts explored in this paper, here are three fundamental questions that it might be fruitful to consider:

- What can we know about a school that is valuable to know?
- How certain can we be when drawing inferences about those things?
- How can we inspect, judge and report inferences in a way that is appropriate to the level of certainty?

Confederation of School Trusts

The Confederation of School Trusts (CST) is the national organisation and sector body for School Trusts in England advocating for, connecting and supporting executive and governance leaders. We are a membership organisation of organisations. This means that the organisation – the Trust – is the member. Our mission is to build an excellent education system in England – every school part of a strong and sustainable group in which every child is a powerful learner and adults learn and develop together as teachers and leaders.

Tel: 0115 917 0142
Email: admin@cstuk.org.uk
Address: Suite 1, Whiteley Mill Offices, 39 Nottingham Road, Stapleford, NG9 8AD
@CSTvoice
cstuk.org.uk

The Voice of School Trusts

Selflessness | Integrity | Objectivity | Accountability | Openness | Honesty | Leadership