

Ofsted's revised school inspection framework: experiences and implications

Steven J. Courtney, *School of Education, University of Manchester, Manchester, UK.*

Corresponding author email: stevenjohncourtney@gmail.com

Abstract

Ofsted's introduction from January 2012 of a revised framework for inspecting schools was intended to sharpen the focus on teaching and learning. In this paper I focus on headteachers' experiences of inspection under this framework, their views of its principles and implications.

The paper draws on findings from a mixed-methods study of headteachers, using survey and interview data to show that inspections have become more focused on pupils' attainment and progress. Headteachers intend to prioritise these and other judged areas over those no longer explicitly judged. Whilst broadly agreeing with the framework's principles, half of the surveyed headteachers say inspection was less positive; the interviewees attribute this to variation in inspector quality and rigidity in the (interpretation of the) framework. Schools serving socio-economically disadvantaged areas might find it harder to obtain a good rating, with implications for headteacher recruitment and retention.

The paper argues that the framework is more successful as a tool for controlling behaviour than for improving schools. It suggests that improved inspector training, a focus on a broad, values-driven agenda by headteachers and increased recognition and engendering of contextual diversity by and through future frameworks should realise improvements.

Keywords: Ofsted; inspection; headteachers; framework; experiences; control; improvement; panoptic performativity.

Introduction

Ofsted (Office for Standards in Education), the non-ministerial department which inspects schools in England, introduced a revised inspection framework for use from January 2012 in response to the education White Paper published by the coalition government (DfE 2010, 69). The White Paper declared that Ofsted should 'focus on just four things-pupil achievement, the quality of teaching, leadership and management, and the behaviour and safety of pupils. Ofsted piloted this framework in 145 schools in summer 2011.

This paper takes as its starting-point the 'well-known phenomenon that organizations will concentrate their efforts on those things they are judged on' (Muijs and Chapman 2009, 41). In other words, as previous versions have done (Ouston, Fidler and Earley 1997), this framework is likely to change what leaders prioritise and hence the education system more widely. A number of scholars go further, claiming that Ofsted privileges its corporate

conceptualisation of educational processes and enforces compliance with it through a culture of performativity within a managerialist discourse which it structures through its inspection regime (Ball 2008; Gilroy and Wilcox 1997; Hoyle and Wallace 2007; Perryman 2006). Its inspection framework operationalises this compliance; schools which do not achieve the standards it prescribes can be closed (Perryman 2010) or its leader dismissed (Roberts 2005). Its influence extends beyond the relatively brief periods of inspection; many leaders internalise its definition of success (Hoyle and Wallace 2007) and subject themselves and school staff to intense surveillance in order to ensure that daily practice corresponds as closely as possible to the Ofsted-sanctioned ideal (Ball 2008) whilst retaining the illusion of freedom (Ball 2003). This phenomenon is recognised in the literature as ‘control of control’ (Power 1997) or panoptic performativity (Perryman 2006).

Inspection as discursive mechanism is constituted of individuals’ experiences. Headteachers are crucially important in an English education system which exemplifies ‘high stakes accountability’ (Muijs and Chapman 2009); their experiences influence profoundly how the school responds both during and after inspection (Ouston, Fidler and Earley 1997). Consequently, their views have been researched on a range of aspects related to inspection experiences, including post-inspection actions and outcomes (Ouston, Fidler and Earley 1997), and the impact of years’ experience (Ferguson, Earley and Ouston 1999) and school context (Chapman 2002) on headteachers’ attitudes to inspection. These scholars were exploring how inspection might lead to school improvement; this paper concerns itself principally with how the framework, as both the product of a managerialist discourse and a mechanism for its reproduction, operationalises the ‘invisible web of power’ (Shore and Wright 1999, 561) underpinning the present audit culture (Power 1997). As Shore and Wright predicted, the invisible has become visible in this unstable period of the framework’s introduction; this paper profits from this brief visibility to report findings from an exploratory study into headteachers’ experiences and views of the framework, supported by the Economic and Social Research Council [grant number ES/J500094/1], to assess both the impact of the framework and the extent to which it functions as a device for controlling not just heads’ practices, but their conceptualisations of what counts as valuable and valid purposes of education.

Methods

Research design

This study used mixed methods, based on what Creswell, Clark, Gutmann and Hanson (2003) call a concurrent triangulation design. Semi-structured interviews and an online survey collecting Likert-scale data were used to strengthen findings and mitigate methodological shortcomings, with the former providing richer layers of meaning to complement and illuminate the generalisations supplied by the latter (Greene, Caracelli and Graham 1989). Triangulation was both methodological (Cohen, Manion, and Morrison 2007) and multilevel (between the national and the individual, see Youngs and Piggot-Irvine 2012). The data were collected sequentially, with some survey respondents volunteering for interview, but analysed concurrently to answer the following research questions:

1. What are headteachers’ experiences of inspection under the January 2012 framework?
2. What are headteachers’ views of the framework’s provisions and objectives?
3. What do headteachers believe will be the implications of the framework?

The data were integrated at the interpretation stage through comparing the results of the two strands. Where incompatibilities arise, these are described and possible explanations given.

Epistemologically, this study is premised on ‘weak constructivism’ (Schwandt 2003); whilst foregrounding context and subjectivity in the social construction of participants’ realities, it both avoids the more extreme interpretation of constructivism holding that social acts are context-specific to the point of mutual meaninglessness, and allows the poststructuralist possibility of discursively-produced commonalities which may be explored and expressed quantitatively.

Data analysis

The survey results were analysed using SPSS. The interviews were transcribed and coded using NVivo software to highlight patterns, make connections and collapse or expand these as necessary into themes. Data coding here was influenced by the research questions, the literature, panoptic performativity theory and emergent survey findings. Inevitably, the rich interview data produced more themes than exemplify or illuminate the quantitative findings; these themes are noted in Table 1, but time and resource constraints prevent them from being fully explored here.

Survey sample

All English secondary schools inspected under the 2012 framework and whose report was published on Ofsted’s website between 1 February and 31 March 2012 were located using that website’s search function. Middle schools deemed secondary were excluded, producing 175 schools whose headteacher was eligible to participate. These were emailed a link to an online survey, piloted with a headteacher involved in the framework’s piloting. The email was sent to the school administration where the headteacher’s email address was unavailable or inoperative. One reminder was issued after two weeks. Out of the 36 (21%) respondents, three (8%) led schools graded Outstanding, fifteen (42%) were graded Good, fourteen (39%) were Satisfactory and four (11%) Inadequate. This compared with 5%, 36%, 42% and 17% respectively in each category amongst the population of 175 inspected schools eligible to respond. A Chi-Square Test showed that this difference between the proportion of ratings in the responding and eligible sample was not statistically significant ($\chi^2=2.3$, DF=3, $p>.5$). This, along with a non-significant gender difference between the responding and eligible sample, ($\chi^2=1.3$, DF=1, $p>.25$) strengthens claims regarding the sample’s representativeness; statistically, leaders of schools rated poorly, or highly, or male or female leaders, did *not* respond to the survey in significantly larger numbers, skewing the results. A large majority was leading a school for the first time (78%) and a similar number had over 24 years’ teaching experience (72%). The percentage of respondents whose school’s Ofsted rating had improved was 25%; those retaining their rating constituted 39% and those whose school’s rating had worsened was 36%. These ratings were compared to those received by secondary schools in inspections from 1 September 2010 to 31 August 2011 (Ofsted 2011). Despite a lower number of Outstanding ratings under the new framework, (8% against 14%), and a higher number receiving Inadequate, (11% against 8%), a weighted Chi-Square goodness-of-fit test revealed that these differences are not statistically significant ($\chi^2=1.5$, DF=3, $p>.68$). The decrease in the number of Outstanding ratings can be explained by the non-inspection of schools already rated as such. The increase in the number of Inadequate ratings *cannot* be explained by the prioritisation for inspection of schools previously rated Satisfactory under the 2007 framework whose outcomes had declined, since a 2 x 3 Chi-Square test showed no statistically significant association between the framework a school was last inspected under (2007 or 2009, indicating time since last inspection) and rating movement up or down (or none) ($\chi^2=2.4$, DF=2, $p>.29$). This increase is therefore more likely to be attributable to the new framework itself (and may in fact be higher, since an (untested) impression formed

during sampling was that Inadequate schools took longer for their report to be published, and so may have missed the sampling window).

Interview group

Surveyed headteachers could volunteer for a semi-structured interview. Twelve did so, of whom six were selected who covered a range of geographical locations and inspection outcomes; one whose school was rated Outstanding, one Good, two Satisfactory and two Inadequate; one school had improved its Ofsted rating, one had retained its rating and four had declined. Whilst thereby avoiding obvious skewing, the group has no statistical claim to representativeness. One interviewee “came out” as an Ofsted lead inspector during interview. This person’s data often constitutes a dissonant case, deepening my understanding of inspection experiences. Further group characteristics are withheld because of the risk of identification. One headteacher was interviewed in person for 45 minutes, the rest by telephone for around 35 minutes. For clarity, in the following section, surveyed headteachers are referred to as respondents, and interviewed headteachers collectively as interviewees and individually as HT 1-6.

Findings

Interview themes

Three overarching themes emerged: framework-derived experiences, inspection team-derived experiences and implications of the framework (see Table 1). Variability in the quality of inspectors is a significant factor underpinning all other sub-themes within inspection team-derived experiences. This variability may explain the fact that many of the sub-themes represented a continuum of experiences or views. For example, under *collaboration*, some interviewees found they were involved closely in the discussions leading to judgments, others found themselves excluded and one experienced both situations throughout the inspection. Equally significant are the absent themes: behaviour and safety, newly privileged in the framework, featured only as a context for other themes. Its appearance in the survey data is probably an artefact of the specificity of that instrument.

<i>Framework-derived experiences</i>	<i>Inspection team-derived experiences</i>	<i>Implications of framework</i>
Emotional response.	<div style="display: flex; align-items: center; justify-content: center;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg);"> Variability in inspector quality. </div> </div>	Changes to school practices.
Increased focus on teaching, learning and achievement.		Penalises schools serving disadvantaged communities.
Framework flexibility/rigidity.		Implications for careers.
Self evaluation.		Implications for school system.
Data-driven inspections.		
Limiting judgments.		
Not capturing the whole school.		

Table 1. Themes from the interview data.

Headteachers' experiences of inspection

The first section of the questionnaire asked respondents about their inspection experiences under the new framework, focusing on the extent to which the framework's objectives were realised (see Table 2). Results show that largely, inspections *were* more focused on the four judgment areas. For a majority of respondents, the framework does not, however, adequately capture schools' activities, contexts and achievements. Nor has the framework produced inspections which are experienced more positively.

Table 2. Headteachers' experiences of inspection under the 2012 framework.

Key to headings in tables 2-3.

1=Strongly agree

2=Agree

3=Neither agree nor disagree

4=Disagree

5=Strongly disagree

6=Don't know

Figures (discounting headings) represent percentages and may not sum owing to rounding.

	1	2	3	4	5	6
There was more focus on pupils' progress	56	42	0	0	3	0
Pupils' behaviour mattered more	39	42	8	11	0	0
Attendance and punctuality data were given more attention	8	42	36	8	6	0
How we deal with bullying was more of a focus	14	36	31	17	3	0
Nothing has disappeared: inspected areas have just moved	0	42	3	42	14	0
Inspectors spoke to more children	31	39	6	17	0	0
Teachers were observed more	56	33	6	6	0	0
There are limiting judgments	44	42	8	0	0	6
Ofsted wanted to know what my school is normally like	17	33	14	28	8	0
This inspection directly affected my teaching staff more	28	33	11	28	0	0
I was more involved in the inspection than previously	11	42	19	22	3	3
This inspection captured what my school is all about	17	30	6	25	22	0
I agree with the inspectors' judgments about my school	19	36	6	25	6	0
This inspection was a more positive experience	3	22	25	22	28	0

All respondents agreed or strongly agreed that there was more focus on pupils' progress, a finding corroborated by five of the six interviewees. HT 5, for example, said:

My view is that if we had low attainment, even if it's below floor, but actually our progress measures for Maths were very strong, they would have come to a different view about standards.

HT 6 agreed, remarking that a key difference with this framework was 'the very, very sharp focus on achievement and progress'. 81% thought that pupils' behaviour mattered more, and only 20% disagreed that how the school deals with bullying was emphasised more. Attendance and punctuality data were given more attention according to 50%, with another 36% noticing no difference. Respondents were not in agreement over whether the framework streamlines priorities; 56% believed some areas are no longer inspected, whilst a large

minority, 42%, felt that nothing has disappeared from the framework. HT 4 was one of five interviewees noticing the absences:

I think that...all of the other stuff, if you like, which is important to schools and would have got recognition under the previous framework around community cohesion and all of that soft stuff that's not teaching and learning in the classroom as it were, that's all gone... (HT4)

For HT 3, however, community cohesion still featured in the inspection:

...it was lovely to see that even though there wasn't a community cohesion judgment, that actually went into the pot and was given recognition in the report.

Twenty percent more respondents believed Ofsted wanted to know what their school is normally like (53%) than did not (33%). HT 6 supported this objective; 'Ofsted's key purpose is... [to ask], what do children typically get in a typical day, or a typical week in this school, in terms of their experience of lessons?' HT 3 was similarly supportive, disapproving of 'people [who] performed on the day'. One way in which this typicality was assessed was through inspectors talking to more children; 70% of respondents agreed or agreed strongly that this happened, against only 17% who disagreed. The desired increased focus on teaching and learning has resulted in more lesson observations, as 89% of respondents report, which may explain the large majority saying their teaching staff was more directly affected (61%). This was corroborated by the qualitative findings, in which *increased focus on teaching and learning* emerged as a major theme affecting inspections; most of the interviewees felt this had *always* been a strong focus of theirs:

On one level, it [the change in framework] didn't change what we do, because obviously my job as a leader is to get outstanding teaching and learning and outstanding achievement. (HT 2)

HT 5 also supported the survey finding, noticing 'a stronger focus on classroom observation', and HT 4 went further, reporting that the inspection 'was completely focused on what they saw in the classroom', but felt that the lead inspector was unfair in interpreting that data. Although the substantive issue here is inspector quality, the subjective way in which different forms of data are selected and interpreted by inspectors to match the emergent narrative created by attainment and value-added (VA) progress measures¹ in (a) core subject(s) is a key finding. For instance, HT 6 said the inspectors 'weren't absolutely convinced ... they were seeing evidence in the classroom to back up our [higher] judgments'. HT 6's capacity to argue the case was reduced owing to the school's poorer value-added data in a key subject; these provided an impression of the school which remained unaltered during better lessons the inspectors subsequently observed. During HT 2's inspection, a similar dissonance occurred between teaching standards observed during inspection and those recorded by the school (though higher up the judgment scale); here, the interviewee was able to persuade the inspectors:

...because our ... value-added data was so good, and our database on teacher assessments was so thorough, it was possible to show that the dissonance was simply stuff happening on the day, and teachers over-talking a bit, rather than a real problem. (HT 2)

The use of data per se did not feature in the questionnaire because there was little indication in the framework that it would change significantly. The interviewees disagreed, for them its

increase and changed use under the new framework was a major theme. Two interviewees felt that the inspectors scrutinised English and Maths more than under the former framework and applied any weaknesses found there to their interpretation of the school's performance more generally:

And whereas we're outstandingly good for everything else, we're poor for maths. And because of that, under the old framework ... they would have looked at the big picture and made a judgment about the school as a whole. And we very much had the impression that with this, they weren't interested in the big picture. (HT 5)

HT 6, too, found that data on progress in Maths influenced the inspection, though with justification:

I think, [x]% of my youngsters made the expected progress in Maths last year, and that's just not acceptable, and we weren't inclined to argue that, and there's no hiding place from that. And we could not demur from their view that therefore the standards were inadequate.

These interview findings support the 86% of respondents who felt that limiting judgments apply in this framework, with none disagreeing. HT 1 was particularly angry about this:

Ofsted ... just said no, it [the overall effectiveness judgment] will have to go down as Satisfactory because Achievement must be at least Good. They took one line out of the Achievement criteria; Achievement must be at least Good, and abandoned the rest of the Achievement criteria.

Whilst it is most often the judgment for overall effectiveness which is limited by that for pupils' achievement, HT 5 felt that this latter may limit others also, '[e]ffectively, there was an issue within maths, therefore, leadership cannot be good'. The percentage of respondents believing that they were more involved in the inspection than previously was 53%, with 25% disagreeing. In the interviews, this developed into a significant theme also expressing a continuum of experiences; *level of collaboration with inspectors*. HT 2 and HT 3 (with dissimilar ratings) reported a high level of collaboration:

...there were two or three things where I said, 'You know where it says that, I would have read it as that', and it was 'Did you? I read it more as this, mm, yeah, but I can see where you're coming from with that.'... I didn't feel there was any imposed outcome on anything. (HT 3)

HT 1 found the inspection team 'too rigid', but attributed this largely to its inexperience with the framework rather than any characteristic of the inspectors, 'they were still very much working to strict guidelines on what they could give and not give'. HT 4 agreed that at this early stage, inspectors were 'very much following the wording':

And their issue was that whilst our attainment was outstanding, ... and our progress was high... it wasn't rapid, because it's been at that level for the last two or three years. And my argument was it's actually quite difficult to be rapid when you're already at a high point, but they weren't having any of that! (HT 4)

HT 3, speaking as an Ofsted inspector, describes how inspectors were trained for the framework:

...in the guidance for us all, ... you look for the general sense of a level, you don't go down and say, I see you haven't, der der, and therefore I can't give you a good on that, for example. So it's a lot less prescriptive.

This contradicts most others' experiences; HT 6 is typical, 'they kept going to the Inadequate ... if any of these bullet points is placed yes, therefore the judgment has to be a four'.

Next, the survey sought respondents' views of the validity and reliability of the inspection process and judgments. They were evenly divided regarding the former; 47% agreed that the inspection captured their school's essence, against 47% disagreeing. The interviewees reflected this split, with a range of inspection outcomes represented in each position, the following exemplifies the former:

[The inspectors said] '...this is what's going to be in the report, can I tell you some of the things that I noticed and which you may wish to think about'. How fabulous is that? ... But they'd noticed it going round, wasn't on the framework, they knew because they're educationalists that it was important. (HT 3)

HT 5, however, felt strongly that the inspection did not attempt to incorporate the school's achievements:

...the school has got significant strengths. And we felt that those were ignored. And so, for example, we've had no fixed-term exclusions for two years. We see that as a pretty significant achievement. As far as they were concerned, they looked, are exclusions an issue? No, they're not. They moved on. So they were looking for problems...

Again, the underlying theme uniting these experiences is variability in the quality and judgment of inspectors, raised spontaneously by all six interviewees, including the inspector, who said, 'inspectors are human, they do all have their own hobby-horses, they do have a mixed background' (HT 3). This theme supports the survey finding that 56% of respondents did and 39% did not agree with the inspection judgment.

Finally, 50% of respondents indicated that this inspection was a less positive experience, with 25% finding it more positive. Since the data are ordinal, a non-parametric Kruskal-Wallis test was performed to establish if the likelihood of experiencing the inspection positively was significantly different for respondents with differing levels of teaching experience; results gave a χ^2 of 7.31, but with a non-significant probability value of 0.063. A Mann-Whitney test confirmed that schools whose Ofsted rating worsened had a significantly less positive experience than those whose rating improved, ($U=18$, $Z=-2.8$, $p<.01$). Amongst the interviewees, two of the six found their inspection a positive experience. One of these two reported an improved rating; these proportions align with the quantitative model. The second was the Ofsted inspector, whose views might be expected to reflect a "bi-cultural" understanding. This latter enjoyed 'an honest and open relationship' with the inspectors, 'the dialogue was very good, and it wasn't chitty-chatty, it was good, solid, professional stuff' (HT 3). One interviewee, HT 6, found the inspectors and the process 'rigorous', but didn't think 'that Ofsted were unfair, ... the judgment was fair and accurate'. To different degrees, the others all found the inspection to be a negative one, for HT 1 it was 'frustrating', for HT 4 it was 'deeply traumatic ... [and] unpleasant' for HT 5 'it felt like a completely negative experience'. The reasons for these experiences are varied; however, they can be placed along a continuum with *framework-related* and *inspector-related* at either end. The framework was criticised for its rigidity and its assumptions; HT 1, for instance, thought

they ‘had a fair team restricted by, really, the framework’, and HT 5 believed it to be ‘deficit-led’, saying:

If you’re marking a child’s work, you don’t start with what they’ve done wrong, you start with what they’ve done right. If you want to move them on. If Ofsted wanted to move schools on, they would do the same.

Mid-way along the continuum are examples of *inspectors’ rigid interpretation of the framework*, including HT 4’s argument about *rapid and sustained* progress. Finally, inspection experiences can be negative wholly because of the attitude and quality of the inspectors. HT 4, for example, found some of them ‘unfriendly ... confrontational and downright rude’, producing an ‘unnecessarily adversarial’ inspection.

Headteachers’ views of the provisions and objectives behind the January 2012 framework

The second part of the questionnaire sought to establish the extent to which respondents agreed with the provisions and objectives of the January 2012 framework, especially where these differed significantly from the last framework (see Table 3). Not all such were listed so as not to over-burden respondents.

Table 3. Headteachers’ views of the principles and objectives behind the 2012 framework.

Figures (discounting headings) represent percentages and may not sum owing to rounding.

	1	2	3	4	5	6
The last framework was too centred on checking compliance with government policies	8	42	17	22	3	8
VA is better at judging pupils’ progress than CVA	3	28	25	25	19	0
The new framework is right to try to be more focused	6	50	17	19	3	6
Behaviour deserves to receive more focus in inspections	8	31	39	22	0	0
Schools rated Outstanding should be inspected less frequently	6	31	17	36	11	0
The increased focus on progress is welcome	8	72	6	8	6	0
The levels of progress the inspectors want are achievable	0	47	11	31	11	0
SEN students are well catered-for within this framework	11	50	14	22	3	0
Inspections shouldn’t assess whether pupils eat healthily	8	33	42	17	0	0

Table 3 shows that respondents broadly support the principles underpinning the January 2012 framework, with around twice as many agreeing as disagreeing that the former framework centred overly on checking policy compliance, and welcoming this framework’s tighter focus. The interview data confirmed this; ‘I do think that the narrow emphasis, the focus on the four areas, is easier to cope with than the dozens of areas that were there before, and I think that’s helpful’ (HT 5). Furthermore, only 17% of respondents thought that pupils’ diets, exemplifying a former judgment area, should still be judged. Two areas receive more attention in this framework; behaviour and safety, and pupils’ progress. Over 80% of respondents agree that progress should feature more strongly in inspections, however, many are concerned about the rates expected, with as many disagreeing (and more fervently) as agreeing that these are possible. The increased focus on behaviour receives a weaker endorsement, with under 40% supporting it and a similar number expressing no view either way, a statistic corroborated by behaviour’s absence from the interviewees’ data. Any potential conflict between the focus on progress and the needs of pupils with Special Educational Needs is largely refuted here, with more than twice as many respondents

agreeing as disagreeing that these students are well catered-for in the framework. Two features of the 2012 framework receive less support. Eleven percent more respondents disagreed (47%) than agreed (36%) that schools rated Outstanding should be inspected less frequently than others, and finally, the abolition of CVA is the least popular of the new framework's features, with 45% of respondents disagreeing that its replacement, VA, is a better way of judging pupils' progress, against 31% in agreement, although those respondents expressing a strong view were seven times more likely to disagree than agree. Both views were expressed by the interviewees in broadly similar proportions to the survey findings, though were not linked necessarily to whether the school had benefited from CVA. HT 6, whose school had, said:

But when you ... started to look at the levels of progress data for Maths and English, you started to realise, actually, this is, the CVA judgments are disguising some really quite significant issues with youngsters' progress in the core subjects.

HT 2, whose excellent value-added outcomes contributed to the school's rating of Outstanding, nevertheless pointed out the system-wide implications of the abolition of CVA, high value-added outcomes being, for HT 2, easier to obtain for schools with high attainment, which are often schools in socio-economically more advantaged areas:

I do totally get why they're saying, you know, standards for children shouldn't be related to their context, but the judgment they make on the school *should*, because it just is harder ... I've got no axe to grind because we got Outstanding ... but it is made easier by having enough children that come from quite motivated homes ...

Headteachers' views of the implications of the January 2012 framework

In the final section of the survey, respondents were asked if they will change the focus they place on a number of areas which are either newly privileged in, which have survived in some form in or which have disappeared from the framework. Broadly, respondents intend to focus more on those areas appearing in the framework, and less on those which no longer do so (see Table 4).

Of the four areas constituting the graded judgments, two; the quality of teaching and leadership and management, survive from the former framework. Results show that in future, 78% of respondents intend to focus even more than presently on teaching, and 50% will focus more on leadership and management. A third judgment area, behaviour and safety, receives more emphasis in this framework than in the former, yet in a finding consistent with that showing that respondents are unconvinced of the value of its "promotion", most indicated here that they would not change the importance they attach to either constituent element. Pupils' progress is a newly-privileged component of the fourth inspected judgment, achievement. Here, two thirds of respondents reported that they would focus more on this element in future. However, amongst the interviewees, only HT 1 intended to focus more on progress in future. This is probably a product of the instrument; the questionnaire asked specifically about progress, whereas the interviewees may have prioritised other areas in response to a more open question. Whilst not separately judged, pupils' literacy features strongly in the new framework: over 70% of respondents state they intend to focus more on it. This was supported by the qualitative data; HT 2, HT 5 and HT 6 volunteered this as an area for future emphasis. Despite the abolition of the Self Evaluation Form (SEF), the principle of self evaluation remains within inspection (Ofsted 2012a). Respondents agree, with 61% indicating they would retain the level of emphasis they place on it. For many of the interviewees, however, it is a delicate and consequence-laden process in which the school attempts to second-guess the inspectors rather than simply assess as inspectors do 'without

fear or favour’ (Ofsted 2012a, 4) the school’s performance. ‘If you write it too negatively, they’ll hold you to it, and if you write it too positively, they’ll say you don’t understand what’s going on’ (HT 5). Three items exemplified areas removed from the framework; of these, two; the school’s community engagement and pupils’ health, show an intention by a majority of respondents to de-prioritise them in future, and a third intend to focus less on the third; the care, guidance and support offered to pupils. HT 3 might be unsurprised by this, warning that ‘there are other heads who actually needed that poke on the community thing to make them take notice of it’, and believes that the government has let community cohesion slip off the agenda ‘at their peril’.

The qualitative data provided two further themes regarding the implications of the framework beyond its impact on schools’ practices and priorities. The first was *implications for headteachers’ careers*. HT 6 is no longer the school’s headteacher following the poor inspection result, saying that, ‘convincing those higher in education that a headteacher who’s led a school from being Good to being a Special Measure school in two years, you know, that’s a bit of a tough sell, and I understand that’. A further two plan to retire before their next inspection, one intends to ‘just carry on’ despite becoming ‘disillusioned’ (HT 4). On the other hand, HT 2 ‘just want[s] to enjoy running an Outstanding school and tweaking the bits that still need tweaking and having a role in outreach’, but, like HT 6, makes the point that outstanding candidates might be dissuaded from becoming head of a school serving a disadvantaged community because of lower opportunities for Outstanding ratings, or the increased risk of job loss; ‘it’s just not fair ... if you lose really, really good heads from schools in challenging circumstances because the best you can hope for is Good’. HT 3 has profited from the experience, despite a poor rating:

We in fact got a report that has helped me re-launch the school because we got some very good comments in it that I’ve been able to use, and ... do some phoenix-out-of-the-ashes stuff in a very positive way. It helped me personally because it did say ‘the headteacher has correctly diagnosed the areas for improvement and has already ...’. So that has helped me establish myself.

The second theme was *implications for the wider school system*. HT 2, along with HT 1 and 6, believes that ‘it’s an impossible framework to get Outstanding on if you’re in challenging circumstances. And that seems really profoundly unfair’. HT 5 felt that the framework was part of ‘quite a narrow, anti-school agenda’, which seeks to create ‘a climate of fear’ rather than improve schools.

Table 4. Headteachers’ views of the implications of the 2012 framework for their practice. All figures represent percentages.

	More focus	No change	Less focus
Self-evaluation will receive	28	61	11
Care, guidance and support will receive	11	58	31
The quality of teaching will receive	78	22	0
Pupils’ literacy skills will receive	72	28	0
My school’s engagement with the community will receive	0	36	64
Behaviour will receive	31	61	8
Pupils’ safety will receive	8	86	6
Leadership and management across the school will receive	50	44	6
Pupils’ progress will receive	67	33	0
Pupils’ health will receive	0	44	56
Developing evidence bases for use in inspections will receive	25	50	25

Discussion

Inspections *are* largely more focused under this framework; this focus is predominantly on one of the four inspected areas- pupils' achievement, defined as their attainment and progress (without contextual factors), especially in the core subjects of Maths and English. Such is the dominance of the core subjects in determining achievement that low(er) attainment or progress there is sufficient to influence negatively other judgment areas, as HT 5 and 6 attest. This emphasis on attainment and progress effectively means that this framework produces particularly data-driven inspection outcomes, and to a lesser extent, experiences. This dissonance is attributable to the privileging of conclusions drawn from achievement data over those from the increased number of lesson observations taking place, which are a key feature of this framework. This may not persist after September 2012, when changes to the framework mean that it will be impossible to obtain an overall rating of Outstanding without achieving that rating for Quality of Teaching.

This study confirms that this framework "raises the bar". Not only are more schools receiving a rating of Inadequate, but the qualitative data suggests how this might be happening. One way is through the exigencies of the framework itself. Another is inspectors interpreting the framework rigidly, despite HT 3's report that this is not how they have been trained, and confirmed in the evaluation schedule accompanying the framework (Ofsted 2012a). Some interviewees attributed this rigidity to the framework's novelty, though the planned revision described above might imply that flux and instability, and concomitant rigidity, are *purposive*, just as in Higher Education (Shore and Wright 1999). The third way in which the bar has been raised is through the application of limiting judgments. These are logically rather than structurally justified, and so rarely appear as such in the framework or its accompanying guidance. For example, HT 5 reported the inspectors believed that leadership could not be good if attainment in Maths was low; evidence again of (core) attainment's primacy. Leadership's function is conceptualised primarily as procuring higher scores in standardised tests, despite evidence that this is incongruent with a world-class education system (see, e.g., Hargreaves and Harris, 2011).

Raising the bar in the way described above disproportionately penalises schools serving more disadvantaged communities. This is compounded through the abolition of CVA, whose measures took into account the various contextual factors affecting achievement and so benefitted these schools particularly. This decontextualisation fits a larger pattern of standardisation. Mansell (2011) describes several devices, including the inspection framework, which constitute the state's apparatus of standardisation through which it controls and homogenises pupils' "learning" experiences. Others include policy context, national assessment models and teacher/leader training programmes. These things being equal, schools should fare similarly in inspection. That they do not is seen as evidence that technical implementation, especially of leadership and teaching, is the 'weak' key variable, explaining why these two feature amongst the four judgment areas. Mansell proposes another variable which it is politically expedient to ignore; pupil motivation, especially insofar as it relates to social deprivation. The moral argument that poor background should not excuse low achievement is persuasive, and has persuaded many heads, but to construct policy as if pupils' social background had no effect on achievement is quite different, and has the effect of blaming low achievement on schools. As HT 2 and 6 point out, this will have consequences for headteacher recruitment and retention in such schools.

Within these broad trends lies enormous variability in the quality of inspectors, affecting how inspection is experienced as well as its outcome, since judgment formation is a subjective process (Gilroy and Wilcox 1997). HT 3 says most inspectors are passionate about school improvement and believe in the inspection model. Nevertheless, not every school currently receives a professional and competent service. Magnifying these differences is the

power-play inherent in inspections, for as Ferguson et al. note, '[t]here can be no genuine partnership between schools and inspectors while the balance of power is so uneven and the penalties for failing the inspection involve such high stakes' (1999).

The framework and this variability in inspector quality have contributed to the large number of headteachers who have found their inspection to be less positive than under the former framework. The correlation between holding this view and leading a school whose rating has declined, along with the probability that "declining" schools are more likely to be serving more socio-economically disadvantaged areas owing to the use of VA measures, means that this framework may well contribute to low morale in exactly those schools where improvements are most needed if standards, however they are defined, are to be systemically raised. The Chief Inspector of Schools, Sir Michael Wilshaw, has said that inducing low morale indicates effective leadership (Abrams 2012); this view contradicts findings by Chapman and Harris (2004, 224) into factors fostering improvement in schools facing challenging circumstances (SFCCs), '[e]ffective leadership in improving SFCCs was characterized by the building of leadership teams that motivated, raised morale and sustained performance over time'. Sir Michael's attitude marks a move away from the 'blend of challenge and support' (Chapman and Harris 2004, 219) which had characterised policies under the former Labour government and towards one redolent of the Conservative period following the 1988 Education Reform Act, focusing mostly on challenge; *both* being required to secure improvement (Barber 1994).

That the framework is structured and implemented in ways which contradict research into school improvement supports its interpretation as being as much a device for controlling what happens in schools as improving them. This study supports this view, finding that many headteachers have embraced the narrow focus on pupils' achievement and are structuring their schools in response. For example, fully a third of the respondents intends to devote less time to the care, support and guidance offered to pupils. The 64% of leaders who intend to spend less time developing links with their community, seemingly because Ofsted no longer privileges it, will do so in contradiction of Chapman and Harris' findings that these are vital for improving SFCCs. Rather than improving the validity of inspections by broadening their terms; finding ways to judge "softer" outcomes, or deontological processes, validity may in future be improved instead by narrowing schools' missions to suit the inspection model. As in Mansell's (2011) question about whether pupils' gains in understanding match rises in attainment in this climate of 'hyper-accountability', it may be asked whether, through *focusing*, some schools will simply become better at inspection at the expense of providing a wide and rich learning experience. In any case, valid questions are raised concerning the purposes of education, and what *vision* leaders are permitted to have in such a climate (see Wright 2001).

Panopticism, where the object of inspection regulates itself, is structurally facilitated in this framework through privileging *typicality*. Here, Ofsted assesses whether "the Ofsted way" is being implemented even in its absence. In this reading, releasing Outstanding schools from routine inspection may be interpreted not as their liberation, but as the success of a policy of panopticism. These schools have internalised the criteria and will henceforth inspect themselves. Panopticism is also encouraged through self-evaluation. The abolition of the Self-Evaluation Form (SEF) sent a clear message that the coalition government viewed it as expendable bureaucracy. However, much thinking, including by Ofsted itself (Ofsted 2004), had seen an *increasing* role for self-evaluation, characterised by a more integrative approach (Plowright 2007). This has not been realised in this framework, which, whilst claiming inspections will 'focus on the needs of schools by... taking account of schools' self-evaluation' (Ofsted 2012, 11) in fact does this by structuring it as a mechanism for behavioural control rather than as an independent strand of evidence. For example, in its

guidance, one key difference between the judgments for “notice to improve” and “special measures” (the latter being worse) is that in a school requiring “special measures”, ‘self-evaluation lacks rigour and is wide of the mark in its conclusions so that leadership and management are inadequate’ (Ofsted 2012b, 19). Here (since in this statement, ‘the mark’ is defined by Ofsted), failing to evaluate the school in the same way as the inspectorate is a sign of weak leadership, confirming Gilroy and Wilcox’s (1997) view that Ofsted privileges and imposes its “reality” despite evidence that Ofsted’s judgments are value-laden (see also Woods and Jeffrey 1998) This interpretation has been internalised by the interviewees, who uniformly wished for their (school’s) judgments to mirror the inspectors’ to validate their leadership. Self-evaluating *fearlessly* would require a significant redistribution of power between schools and Ofsted.

Recommendations for practice and policy

Ten years ago, Chapman (2002) published three principles which should underpin any future framework aiming to improve schools: *context specificity* is explicitly rejected in this latest formulation; *change at all levels* is arguably there, but with a reduced purpose; and *post-inspection relationships* exist predominantly for those schools receiving Inadequate ratings; these relationships being about monitoring rather than support. With the benefit of hindsight, I am less convinced than Chapman was of the truth of the premise on which his principles are based, i.e., that the inspection regime is designed primarily to bring about school improvement. If that were so, I doubt it would look the way it currently does. Additionally, whilst most inspectors *are* probably motivated by the desire to improve schools, the variation in their quality is too great to guarantee it. I recommend therefore better recruitment and training procedures for inspectors to facilitate an improved inspection experience. As HT 3 noted, many of the opportunities for improvement happen in conversations between headteacher and inspectors rather than deriving directly from the framework or the report. Competence in building professional relationships characterised by insight, curiosity and empathy should be key to improving these opportunities.

This framework may penalise schools serving socio-economically disadvantaged communities. I recommend that more research be conducted to confirm or refute this hypothesis, given its potential consequences for headteacher recruitment and teacher morale in such schools. I recommend also that headteachers engage in conversations at all levels in their schools concerning conceptualisations of the broad purposes of education, and work to achieve those purposes. Where resources and the curriculum have narrowed, I recommend that they assess the impact of that on pupils’ attainment of those broader purposes and widen them.

Finally, I recommend that any future framework recognise the importance of the contextual in evaluating schools’ work, not as an excuse for poor performance, but as a starting-point for conversations about, for example, local priorities and styles of leadership. Indeed, it should actively foster contextual solutions through structural flexibility. A standardising model has not delivered notable systemic improvements, and it is unlikely to do so. Developing a framework which encourages the diversity necessary to engender change in a range of contexts is likely to be a more profitable avenue to explore, and may reinvigorate purposeful, values-driven leadership, especially in England’s least-advantaged schools.

Notes

¹ Value-added progress measures were used from 2002 in published performance data for English schools until eclipsed by contextual value-added measures (CVA) from 2006, and are re-privileged in this framework. VA describes the difference a school has made to individual

children's outcomes, comparing that child's expected outcomes, based on prior attainment plus the national median for that level, with actual outcomes. CVA included in its statistical modelling other factors, such as ethnicity, gender and entitlement to free school meals. In abolishing it, the Department for Education described it as 'morally wrong to have an attainment measure which entrenches low aspirations for children because of their background' (DfE 2010: 68).

Acknowledgements

I would like to thank the headteachers who participated in this project, particularly those six whom I interviewed; Chris Chapman for his guidance throughout this project; Michael Wigelsworth for invaluable methodological advice and Jan Webber, inspections specialist at the Association of School and College Leaders, for sharing her expertise with me.

References

- Abrams, F. 2012. Is the New Chief Inspector of Schools Just an Instrument of Government? *The Guardian*, January 23. <http://www.guardian.co.uk/education/2012/jan/23/chief-inspector-schools-michael-wilshaw>. (accessed 13 August 2012).
- Ball, S.J. 2008. The Legacy of ERA, Privatization and the Policy Ratchet. *Educational Management Administration & Leadership* 36, no. 2: 185–199.
- Ball, S.J. 2003. The Teacher's Soul and the Terrors of Performativity. *Journal of Education Policy* 18, no. 2: 215–228.
- Barber, M. 1994. Power and Control in Education 1944–2004. *British Journal of Educational Studies* 42, no. 4: 348–362.
- Chapman, C. 2002. Ofsted and School Improvement: Teachers' Perceptions of the Inspection Process in Schools Facing Challenging Circumstances. *School Leadership & Management* 22, no. 3: 257–272.
- Chapman, C., and A. Harris. 2004. Improving Schools in Difficult and Challenging Contexts: Strategies for Improvement. *Educational Research* 46, no. 3: 219–228.
- Cohen, L., L. Manion, and K. Morrison. 2007. *Research Methods in Education*. Sixth. Abingdon: Routledge.
- Creswell, J.W., V.L. Plano Clark, M. Gutmann, and W.E. Hanson. 2003. Advanced Mixed Methods Research Designs. In *Handbook of Mixed Methods in Social and Behavioral Research*, ed. A. Tashakkori and C. Teddlie, 209–240. Thousand Oaks, CA: Sage.
- Department for Education (DfE). 2010. *The Importance of Teaching: The Schools White Paper 2010*.

- Ferguson, N., P. Earley, and J. Ouston. 1999. New Heads , OFSTED Inspections and the Prospects for School Improvement. *Educational Research* 41, no. 3: 241–249.
- Gilroy, P., and B. Wilcox. 1997. Ofsted, Criteria and the Nature of Social Understanding: A Wittgensteinian Critique of the Practice of Educational Judgement. *British Journal of Educational Studies* 45, no. 1: 22–38.
- Greene, J.C., V.J. Caracelli, and W.F. Graham. 1989. Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis* 11, no. 3: 255–274.
- Hargreaves, A., and A. Harris. 2011. *Performance Beyond Expectations*. National College for School Leadership.
- Hoyle, E., and M. Wallace. 2007. Educational Reform: An Ironic Perspective. *Educational Management Administration & Leadership* 35, no. 1: 9–25.
- Mansell, W. 2011. Improving Exam Results , but to What End ? The Limitations of New Labour’ s Control Mechanism for Schools: Assessment-based Accountability. *Journal of Educational Administration and History* 43, no. 4: 291–308.
- Muijs, D., and C. Chapman. 2009. Accountability for Improvement: Rhetoric or Reality? In *Radical Reforms*, ed. C. Chapman and H. Gunter, 28–41. Abingdon: Routledge.
- Ofsted. 2004. *The Future of Inspection: a Consultation Paper*.
- Ofsted. 2011. *Section 5 Inspection Judgement Summary by Phase 2010-11*. www.ofsted.gov.uk/resources/inspection-judgements-201011-for-maintained-schools. (accessed 13 August 2012).
- Ofsted. 2012a. *The Evaluation Schedule for the Inspection of Maintained Schools and Academies*.
- Ofsted. 2012b. *Conducting School Inspections*.
- Ofsted. 2012c. *The Framework for School Inspection from January 2012*.
- Ouston, J., B. Fidler, and P. Earley. 1997. What Do Schools Do After OFSTED School Inspections-or Before? *School Leadership & Management* 17, no. 1: 94–104.
- Perryman, J. 2010. Improvement After Inspection. *Improving Schools* 13, no. 2: 182–196.
- Perryman, J. 2006. Panoptic Performativity and School Inspection Regimes: Disciplinary Mechanisms and Life Under Special Measures. *Journal of Education Policy* 21, no. 2: 147–161.
- Plowright, D. 2007. Self-evaluation and Ofsted Inspection: Developing an Integrative Model of School Improvement. *Educational Management Administration & Leadership* 35, no. 3: 373–393.

- Power, M. 1997. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Roberts, A. 2005. Transposing a Culture: Reflections on the Leadership of a Closing school(R). *Improving Schools* 8, no. 3: 237–253.
- Schwandt, T.A. 2003. Three Epistemological Stances for Qualitative Inquiry: Interpretivism, Hermeneutics, and Social Constructionism. In *The Landscape of Qualitative Research: Theories and Issues*, ed. N.K. Denzin and Y.S. Lincoln, 292–331. Thousand Oaks, CA: Sage.
- Shore, C., and S. Wright. 1999. Audit Culture and Anthropology : Neo-Liberalism in British Higher Education. *The Journal of the Royal Anthropological Institute* 5, no. 4: 557–575.
- Woods, P., and B. Jeffrey. 1998. Choosing Positions : Living the Contradictions of OFSTED. *British Journal of Sociology of Education* 19, no. 4: 547–570.
- Wright, N. 2001. Leadership, Bastard Leadership and Managerialism: Confronting Twin Paradoxes in the Blair Education Project. *Educational Management Administration & Leadership* 29, no.3: 275–290.
- Youngs, H., and E. Piggot-Irvine. 2012. The Application of a Multiphase Triangulation Approach to Mixed Methods: The Research of an Aspiring School Principal Development Program. *Journal of Mixed Methods Research* 6, no. 3: 184–198.