



Research and analysis

Research commentary: assessing the quality of education

Published 26 June 2019

Contents

[Introduction](#)

[Inspecting the quality of education](#)

[International seminar on lesson observations](#)

[Validity](#)

[Reliability](#)

[Next steps](#)

Introduction

Daniel Muijs, Ofsted's Deputy Director, Research and Evaluation, summarises our lesson observation and workbook scrutiny research.

Lesson visits and work scrutiny have long been part of Ofsted's inspection toolbox. Doubtless many of you will remember graded lessons. This is a practice we have obviously abandoned in light of both unintended consequences (we certainly do not want to see the return of 'Ofsted lessons') and issues around reliability and validity.

In the [common inspection framework](#), inspectors use both lesson observation and work scrutiny mainly in a qualitative way. This allows inspectors to collect rich data and use their expert knowledge in making judgements. It also avoids the unintended consequences that can occur when using checklists in high-stakes inspections.

We did not, however, have much evidence on the validity or reliability of each specific inspection method we were using. This is not necessarily a problem. The evidence collected through these methods was only ever intended to be an indicator of processes at whole-school level.

However, as we developed the [new education inspection framework \(EIF\)](#), we felt it was time to revisit methodologies of inspection as well as framework content. We therefore set out a programme of work to look at our inspection methodology.

Inspecting the quality of education

The main aspect of this was looking in detail at how we could inspect quality of education within a framework that puts curriculum at the heart of inspection. The approach we have developed uses 'deep dives' into a selection of subjects as its core [methodology](#).

This approach has 3 main elements:

- a top-level view of quality of education developed during conversations between the lead inspector and the principal
- a deep dive into a selection of subjects during which the inspectors will gather evidence on quality from a range of sources
- a meeting during which the inspectors will bring all the evidence together

As part of these deep dives, inspectors will carry out lesson visits and scrutinise a selection of pupils' work. Alongside the deep dives looking at quality of education, inspectors will also be collecting evidence on leadership and management, personal development, and behaviour and attitudes. This last judgement can also benefit from evidence provided through lesson visits.

Because they are an important part of the deep dive approach, we wanted to make sure that the evidence we collect through lesson visits and workbook scrutiny allows us to reach a judgement that is valid and reliable. We felt this was important because the deep dive approach puts more scrutiny on subjects or equivalent as the unit of analysis. The evidence from this is then collated to come to the judgement of quality of education at whole-school level.

International seminar on lesson observations

To help us do this, we set up an [international seminar on lesson observation](#). This seminar suggested that we were right to take the focus off the classroom level. It also confirmed all our concerns about grading lessons and using rating scales. The international experts did, however, feel that we needed greater clarity on what inspectors were looking for in classrooms. We therefore ran a [large-scale research project, building on the international evidence base on lesson observation](#). Our purpose was different from that of most work in the field, however. Lesson observation research typically looks at individual lessons for

factors relating to teacher or teaching effectiveness. But we want to use lessons as one indicator of quality of education at the subject level. That is why we are calling our approach 'lesson visits' rather than 'lesson observation'.

While lesson observation has this research tradition behind it, this is not the case for work scrutiny. Nevertheless, we know from inspection how important it is to look at pupils' work, so here too we felt it would be useful to look at the validity and reliability of what we are doing. We therefore carried out a [study on validity and reliability of work scrutiny](#) as well. This was conceived as a smaller scale and more exploratory piece of work due to the very limited external evidence base we can draw on. We will do further work on this in the future.

Validity

The main aim of both studies was to look at validity. The most important thing to get right is that we are looking at the right things. In addition, of course, we want to do this reliably. So in both cases we first set out to learn what things lesson visits and work scrutiny are most useful for in the context of an inspection framework focused on quality of education. We then wanted to see whether we could collect the necessary evidence reliably.

In the lesson visits study, we first set out to determine, based on the literature on lesson observation research, the draft EIF and discussion with inspector colleagues, what areas would be worth exploring.

We ended up looking at 3 main areas:

- teaching quality
- classroom and behaviour management
- curriculum implementation

We developed 18 quality indicators across these 3 areas and added a 5-point scoring guide. This is similar to the approach we used in [phase 3 of our curriculum study](#). As with the curriculum study, the indicators were meant purely for research purposes. Quantitative indicators allow us to do quantitative analysis, which is useful when you're looking at questions of validity and reliability. They are, however, problematic on inspection, because they lead to undesirable consequences – as we found when we used graded lesson observation in the past.

We followed a similar approach for the work scrutiny study. We drew on the framework and conversations with inspectors to create indicators for:

- building on previous learning
- depth and breadth of coverage
- pupils' progress

- opportunities to revisit and practise what they know

In the [lesson visits study](#), inspectors visited 22 schools and 15 colleges. So that we could look at reliability, we used paired observations in which inspectors independently rated the lesson on the indicators. In total, inspectors completed 346 paired observations across 74 departments. Lesson visits lasted for between 15 and 30 minutes, which is similar to what happens on inspection currently. As a lesson cannot be understood other than as part of a sequence of lessons on a particular topic, inspectors had a conversation with subject leads at the beginning of the visit to get a sense of the subject journey. They also spoke to the teachers and a small number of pupils they had observed about where this lesson fitted into their curriculum progression. In addition to the visits, we ran a number of focus groups with the inspectors to get their insights on how they felt the observations went, making this essentially a mixed methods study.

The [work scrutiny study](#) also used a mixed methods approach. We looked at over 300 workbooks from Years 3, 4, 5, 8 and 9, covering mathematics, English, history and geography, science and French. Nine HMI independently looked at the work, with 2 HMIs scrutinising each book so that we could calculate reliability estimates. HMI looked at work both within and outside their own subject specialisms. As with the lesson observation study, we ran a focus group to get inspectors' views on the process and on what they thought were the most useful ways in which we could carry out work scrutiny.

A weakness in the design of both studies was that there was limited training for observers and raters. This was a major disadvantage, as previous research suggests that substantial amounts of training are needed to generate reliability. In the run-up to the introduction of EIF, we are delivering what is probably the most extensive programme of inspector training we have ever done at this stage. We are continuing our work on developing subject guidance, working with expert groups for each subject. In the first phases, this had to focus on developing inspectors' understanding of the content and background of EIF. Training on inspection methodology happened after we had completed the research project, once we had developed the deep dive approach that was informed by the findings from these studies.

We wanted to find out what lesson visits and work scrutiny are most useful for on inspection. We also wanted to discover what information they can provide us on quality of education and behaviour and attitudes as part of our arsenal of inspection tools. This is fundamentally a question of validity. Here, the findings are positive, especially in schools. In the lesson observation study, we found that observers clearly distinguished behaviour from teaching and curriculum. This confirms the model we are using in EIF, where we have separate judgements for behaviour and attitudes and for quality of education. This acknowledges that good behaviour management is a pre-condition for learning, but that in itself does not guarantee a good quality of education.

Interestingly, the items measuring teaching quality and those measuring curriculum quality were closely related, so they formed one factor rather than two

different ones. This means that observers do not clearly distinguish the two. This again confirms the model we developed in EIF, where we see both as part of the overarching 'quality of education' concept. It also follows the intent of our model, in that the lesson is seen as part of a sequence with the delivery of a curriculum.

We also found that, on average, observers' ratings of behaviour were slightly higher than the ratings for teaching and curriculum. It was not uncommon to view lessons where pupils' behaviour was exemplary, yet the quality of teaching observed was not strong.

Inspectors themselves told us that they found the structure we provided useful, because it helped them to be clear on what they were and were not looking at.

The value of focusing on a relatively limited number of areas, and of having a clear set of things to look at, was also obvious in the work scrutiny study. This helped inspectors focus on the essential aspects of the quality of education, while minimising the effect of irrelevant factors such as neatness of handwriting.

Inspectors agreed that using the indicators 'allowed them to delve under the surface'. They also agreed that a focus on knowledge sequencing and depth and breadth of content coverage allowed them to engage with curriculum content in a deeper way.

Overall, inspectors were confident using the indicators. However, the limitation of indicator systems and rubrics was apparent in how difficult inspectors found it to distinguish between adjacent points on the 5-point scale.

Reliability

The use of paired lesson visits and paired work scrutiny was so that we could measure reliability. In schools, we found substantial levels of reliability in primary and on behaviour in secondary. We found only good levels of reliability on curriculum and teaching in secondary. That the secondary curriculum and teaching measures were lower was due to inspectors looking at lessons outside of their subject expertise. To counter this, we are developing subject-specific guidance for inspectors in all subjects. This is in collaboration with expert groups. Overall, though, within a high-inference (subjective) model that requires a lot of observer judgement, this is a positive outcome.

The picture for work scrutiny showed good but not substantial levels of reliability overall. Depth and breadth of coverage, building on prior learning, and pupil progress showed better reliability than opportunities for practice. Overall, reliability was quite a bit higher in primary than in secondary.

Our model for lesson visits did not show the same level of reliability in further education and skills (FES) settings as it did in schools. This is because the model we proposed does not fit with all delivery methods and contexts in FES. The

model is essentially classroom-based, which makes it less suitable for FES outside A-level provision. Expectations of behaviour, for example, vary considerably in FES depending on the age and type of learners in a particular setting. In addition, the behaviour indicators we used clearly do not work for adults in technical classes and apprentices in a work setting. We are therefore working on developing a model that is more suited to the FES context. Similarly, work scrutiny was not considered an appropriate activity to carry out in most FES contexts.

Next steps

This in-depth look at 2 important parts of our inspection methodology has been incredibly useful and has informed the development of the deep dive methodology. The study has reinforced both the value of lesson visits and work scrutiny and has pointed to the limitations of these methods. That is why we see both as part of a range of data we are collecting on quality of education or behaviour and attitudes. It is also why conversations are so important in our inspection model. Conversations with leaders, teachers and pupils can provide the vital context we need to understand what we see. These conversations are central to the deep dive model. However, relying purely on what people tell us would be problematic in an inspection context. This is why we need first-hand evidence from pupils' work and lesson visits among other sources, such as attainment in national tests.

What these findings also suggest to us is the value of a clear focus on a more limited number of judgement areas when doing lesson visits and work scrutiny. We have therefore built this into inspector training for EIF. The studies also confirmed the limitations of using indicators and rating scales, such as the difficulties we found in distinguishing adjacent scale points.

As an inspectorate, we know that using indicators and scales leads to a range of unintended consequences, such as standardisation and potential gaming. Indicators and scales do not allow us to fully understand why something is observed in books or lessons. The judgement of inspectors must therefore be central to inspection. What these findings do suggest, however, is that inspector training is important to ensure reliability and validity in collecting first-hand data. That is why we are running the training programme mentioned earlier. The deep dive approach that built on these findings is itself subject to extensive piloting (we have piloted EIF and the new methodology in over 150 schools), so we can make sure we have got it right when we launch the framework in September.

These 2 studies are an illustration of our commitment to making sure what we do is evidence-based, not just in terms of content but of methodologies used. This will not be our final look at inspection methodology. We will be looking at observation methods in FES, and are running a large-scale evaluation alongside the introduction of EIF from September. We are committed to developing the

most evidence-based inspection framework ever, and we will continue to deliver on that promise as we implement EIF.

Is this page useful? [Yes](#) [No](#)

[Is there anything wrong with this page?](#)

Prepare for EU Exit

[Prepare your business or organisation for the UK leaving the EU](#)

[Living in Europe after the UK leaves the EU](#)

[Prepare for EU Exit if you live in the UK](#)

[Continue to live in the UK after it leaves the EU](#)

Services and information

[Benefits](#)

[Education and learning](#)

[Births, deaths, marriages and care](#)

[Employing people](#)

[Business and self-employed](#)

[Environment and countryside](#)

[Childcare and parenting](#)

[Housing and local services](#)

[Citizenship and living in the UK](#)

[Money and tax](#)

[Crime, justice and the law](#)

[Passports, travel and living abroad](#)

[Disabled people](#)

[Visas and immigration](#)

[Driving and transport](#)

[Working, jobs and pensions](#)

Departments and policy

[How government works](#)

[Departments](#)

[Worldwide](#)

[Services](#)

[Guidance and regulation](#)

[News and communications](#)

[Policy papers and consultations](#)

[Transparency and freedom of information releases](#)

[Help](#) [Cookies](#) [Contact](#) [Accessibility](#) [Terms and conditions](#)
[Rhestr o Wasanaethau Cymraeg](#) Built by the [Government Digital Service](#)

OGI All content is available under the [Open Government Licence v3.0](#), except where otherwise stated



© Crown copyright