
Judging the Quality of Teaching in Lessons: some thoughts prompted by Ofsted's subsidiary guidance on teaching style

COLIN RICHARDS

ABSTRACT Lesson observations involving judgements of teaching quality are a regular feature of classroom life. Such observations and judgements are made by senior and middle managers in schools and also, very significantly, by Ofsted inspectors as a major component of their judgement on the quality of teaching in a school. Using the example of Ofsted inspection, but with arguments that can apply also to routine classroom observation by school managers, this article teases out what can reasonably be said about teaching quality based on observation. It reveals the importance, but also the limitations, of classroom observation. It stresses the tentative, context-specific nature of judgements and the need for observers to have relevant experience and insight into the complexities and imponderables of classroom observation that focuses on teaching quality. It does not deal with the different but allied issue of whether lessons themselves should be graded.

Welcome but Negative Guidance

In publishing paragraphs 64-67 of its subsidiary guidance issued in January 2014 (Ofsted, 2014b), the Office for Standards in Education, Children's Services and Skills (Ofsted) has made a very welcome contribution to the difficult and contested issue of how to judge the quality of teaching. That guidance makes it quite clear that qualitative judgements of teaching cannot, and should not, be characterised unequivocally by the presence or absence of particular features in lessons. Ofsted is emphatic that it does not favour a particular teaching style. According to the guidance, 'good teaching' does not *necessarily* involve independent learning or pupil activity or differentiation. 'Poor' teaching doesn't *necessarily* involve pupil passivity, overlong teacher talk

or the absence of different activities. Yet, though important, it is curious guidance nonetheless. In the space of twenty lines the word 'not' appears eleven times. Eight out of its eleven sentences are negatives. 'Teaching quality' is undoubtedly an elusive concept but it needs more than negative instances to capture something of how it can be recognised and characterised. This article is an attempt to provide more positive guidance in recognising and judging teaching quality in the context of lesson observations. It does not deal with the issues of whether lessons themselves should be given a grade and that grade communicated to the person observed.

Making Judgements of Teaching Quality

As a preliminary to making any judgements, the purpose of any episode of teaching needs to be clarified. The Ofsted handbook (Ofsted, 2014a) is right to stress that '[t]he most important role of teaching is to promote learning and to raise pupils' achievement'. That seems straightforward enough - but is it? Every example of classroom teaching involves at least the intention of contributing to learning in some way, but not necessarily to *new* learning. A lesson might be designed to practise skills already learned. It might be designed to remind children of what they have learned. Its purpose might be to celebrate what has been learned and achieved. It might be designed to motivate, to excite interest or puzzlement about what is to be learned. It might be designed to entertain in order to lay the ground for future learning or to remind pupils of past learning. It does not necessarily have to focus on, or result in, *new* learning. Classroom teaching can serve a variety of purposes; a single observed episode may involve several. The Ofsted inspection handbook pays scant recognition to this multiplicity of purpose, yet this is an undeniable characteristic of classroom teaching. Any judgement of the quality of teaching in a lesson must necessarily take cognizance of its purpose or purposes.

Another preliminary judgement is required. That purpose needs to be suitable or fitting. It needs to be an appropriate response to the learning that has been undertaken before and to that which is anticipated to follow. It is counter-productive, for example, to get pupils to practise a skill which hasn't been adequately taught or learned previously. Spending time recapitulating minor details rather than key facts, ideas or generalisations from previous sessions would not usually meet the appropriateness criterion, nor would a lesson, however motivating or entertaining, which was not to be followed up reasonably soon thereafter by work that built on it. There must be a firm, educationally justifiable connection with the preceding and following lessons. Hopefully, in the context of a particular lesson, that connection is clear and can be justified by the teacher, but is it always obvious to, or sought by, those evaluating the teaching? Within the time constraints of an inspection or an in-school observation it is very difficult in every case to find the time to establish those connections or lack of them through scrutiny of planning or discussion with the teacher or pupils involved. Without that understanding, any judgement

of the quality of the teaching seen is bound to be partial – at least to some degree. Yet the inspection handbook makes no specific reference to this issue except for pointing out that judging the quality of teaching in its most general sense must take account of pupils' learning and progress over time.

Teaching, however skilled, cannot be considered 'good' unless it is meaningfully related to purpose. It cannot be so related if the content is false or only partially accurate - a feature that can be objectively determined by a suitably knowledgeable observer. Distortions, bias or lack of clarity can also be detected, but this involves rather more in the way of subjective judgement. In addition, teaching cannot be considered as 'good' if the material being taught is trivial or worthless; that requires a subjective judgement but one informed by knowledge and understanding of the subject matter and the requirements of the curriculum the school has agreed, or been mandated, to follow. That judgement as to how worthwhile the material being transacted is is central to any appraisal of the quality of teaching, yet it does not feature explicitly in inspection documentation.

According to Ofsted, lesson observations from which the quality of teaching is judged make a major contribution to the overall judgements made about the quality of teaching in the school, though not by aggregating lesson grades awarded following lesson observations. According to the January 2014 inspection framework (Ofsted, 2014c), when judging the latter (and presumably when judging the former, though this is not made explicit), inspectors are required to consider the extent to which 'the teaching in all key stages and subjects promotes pupils' *learning* and *progress* across the curriculum' (my italics). The handbook adds: 'When judging and observing teaching, inspectors must be guided by the response and engagement of pupils and evidence of how well they are *learning*' (my italics). Both requirements raise the issue of how far it is possible to gauge the kind of learning taking place, and its extent, assuming that the purpose of the lesson is to foster *new* learning.

There are complexities and difficulties involving in evaluating *learning* which are not considered in the inspection documentation. Initially two senses need to be distinguished. The first is the *task* sense in which, for example, an inspector observes that pupils in a class are engaged in learning x or y but judges that they haven't yet learnt (i.e. understood) these. If, however, they are judged to have understood x or y then it can be said that they have learnt them in the *achievement* sense of learning.

There are two aspects to learning in the *task* sense, one not directly amenable to observation-based judgement and the other amenable to some degree. In the first, learning involves whatever processes occur in the learners as a result of being taught. The mental strategies, associations or whatever that are involved in learning are unobservable. It may be possible to pick up some clues of this internal processing through, for example, talking to pupils as they work, observing their overt behaviour, or examining their 'working out' of problems on paper or online. But such evidence is indirect, very partial and difficult to

collect. Observers should be very cautious in claiming to be able to 'get at' whatever internal processes are involved in learning.

In contrast, learning in the *task* sense can be judged by examining the behaviour of pupils (i.e. their observable responses to the act of teaching). Neither the handbook nor the inspection framework spells out what some of these might be, except to point out that 'not all aspects of learning, for example pupils' engagement, interest, concentration, determination, resilience and independence, may be seen in a single observation'. While making a useful point, it is a curious list nonetheless; for example, 'resilience' or 'determination' involve qualities needing to be displayed over a time scale greater than a single lesson observation; equally, 'engagement' or lack of it should surely be evident in every lesson, and commented upon if not.

The negative advice beginning 'not all aspects of learning...' needs to be complemented by more positive guidance. Arguably, in judging pupils' observable responses, inspectors have to consider, and answer, such questions as: 'Are pupils engaged in the work?' 'Are they enthused?' 'How are they responding to the teacher's questions?' 'How are they contributing to discussion?' 'Are they responding appropriately to the teacher's monitoring of their learning?' 'Do they appear to be challenged by their work?' However, none of these – enthusiasm, engagement, concentration, active contribution, responding to appropriate monitoring of their learning - is *necessarily* indicative of new learning taking place. For example, a class or a group of children might exhibit all those qualities but simply be involved in revising or applying what they already know. Nevertheless, it seems likely that when all or most of these qualities are in evidence, *task* learning of some kind or other is going on. Second, there is bound to be an element of uncertainty when children's observable responses are being interpreted. For example, they may be feigning these qualities in order to impress the teacher or the inspector. They may even be feigning lack of those qualities so as to appear 'cool' to their peers! Observable responses can have different causes, which the experienced observer needs to try to tease out. Third, inspectors are likely to vary somewhat in their interpretation of what constitutes appropriate 'engagement', 'response', 'contribution', 'challenge', etc. But with appropriate training, experience and discussion among inspectors or other observers it should be possible for their judgements on these aspects to be broadly harmonised and a reasonable degree of consistency of judgement achieved.

The 2014 inspection framework and handbook are on reasonably secure ground; generally valid judgements *can* be, and should be, made about pupils' observable responses to the teaching they receive. It is here that some of the features included in paragraphs 64-67 of the subsidiary guidance come in. It could be that in a particular lesson part of the explanation for pupils' response might rest on the undue extent of teacher talk, the absence of a variety of learning activities or the confusing lesson structure or, in more positive examples, on the use of relevant examples, the skilful use of humour or the evident enthusiasm of the teacher. The features could and should be cited in any

report or feedback to indicate in this *particular instance* how some of them were judged to be influencing the responses, positive or negative, of the pupils. But the citing of the features needs to come *after* the judgement that learning in the task sense is taking place or has taken place; the features help account for, but are not co-terminous with, the learning. And importantly, in a different context those same features may not have the same explanatory value. Hence the importance of Ofsted's subsidiary guidance to the effect that qualitative judgements of teaching cannot, and should not, be characterised unequivocally by the presence or absence of particular features in lessons.

But what about evaluating the learning in its *achievement* sense? The 2014 inspection handbook is quite clear: 'The key objective of lesson observations is to evaluate the quality of teaching and its contribution to learning', and inspectors should ensure that they 'gather evidence about how well individual pupils and particular groups of pupils are making progress ... and assess the extent to which pupils have gained in knowledge' (Ofsted, 2014a, pp. 10-11). Though the guidance that follows is not explicit in detail, presumably this involves inspectors and other observers in answering questions such as 'Does the teaching lead to children understanding what is being presented or required of them?' 'Have pupils gained in knowledge?' 'Are pupils learning as a result of the teacher's checking and interventions?' 'Do they understand how to improve their learning?' 'Has their learning been improved by frequent, detailed and accurate feedback?' But how feasible is it to expect inspectors to be able to make such judgements? Here, inspectors are on less secure ground. Judging the quality of teaching in a lesson by its contribution to learning in the *achievement* sense is at best not straightforward, and in many cases not possible with any satisfactory degree of certainty.

Evaluating how well pupils have learned in a lesson or period of observation involves: (a) knowledge of the understanding and skills which pupils 'bring' to the lesson; (b) knowledge of understanding and skills they 'take away' at the conclusion of the lesson or observation period; and (c) judging whether the extent of the difference between the two is sufficient and appropriate given the purposes of the lesson. Except in lessons involving, for example, the learning of a song, or the learning of specific physical competences, as in PE, dance or drama, which children cannot perform at the beginning of a session but can demonstrably do so at the end, or those involving the learning of factual information which pupils do not know when tested at the start of the lesson but can demonstrate at its end, inspectors do not usually have *sufficient* knowledge of either (a) or (b) and therefore cannot gauge (c) the degree of change in pupils' understanding or knowledge as a result of the lesson. Of course, it is likely that most pupils do learn something new in most lessons, though as stated previously, some sessions might quite appropriately be concerned with practising or applying previous learning. It is also the case that there may be a time lag between a particular episode of teaching and learners' reaction to it. Changes in understanding, in particular, are not detectable to any significant degree through observation or brief discussion

with pupils by inspectors or other observers. It might be possible to evaluate conceptual understanding to a certain extent if inspectors have sufficient time to question closely an appropriate sample of pupils before and after a lesson and to take account of their oral, written and other responses to the teaching received. The Ofsted inspection framework and evaluation schedule do not refer to such close, in-depth questioning. How could they when most lesson observations are for no more than fifty minutes and often for less? In those cases where progress in learning, or lack of it, can be detected by an observer it is important to identify what features of a lesson are likely to have contributed to that outcome, but the caveats discussed above in relation to *task* learning also apply here. But, as the Ofsted subsidiary guidance emphasises, there is no hard and fast connection between a particular feature or features of a lesson and pupils' learning.

The Nature of Judgements of Teaching Quality

If the points made above are accepted it follows that in the majority of cases judgements about the quality of teaching in lessons have to be tentative, and consequently have to be offered as such in any feedback to those who have been observed. There is inevitably a considerable degree of inference involved in the judgements, especially those relating to the extent to which learning has taken place; there is inevitably too an element of professional judgement as to which features of the lesson have contributed to, or inhibited, whatever learning is inferred as having taken place. 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 person being observed. Those being given feedback need to be reassured that because of the tentative, partial, lesson-specific nature of those judgements, no firm generalisations can be made about the overall quality of their teaching based on any one lesson observed.

The tentative, context-specific nature of these judgements makes it problematic, and arguably unnecessary, to restrict summary judgements of teaching quality in lessons to one of four grades. Qualitative judgements do, of course, need to be made, but they need to be tailored to the particular context and not restricted to one of four descriptors when others such as 'very good in respect of x or y' or 'excellent in respect of x and weak in respect of y' might be more appropriate. Such lesson-specific judgements, expressed in a variety of lesson-sensitive ways, need to be taken into account when generalisations about the quality of teaching in a school are being made by the inspection team as a whole but do not require a set of standardised categories or grades.

Acquiring the Expertise to Make Judgements of Teaching Quality

This article has made much of the need to exercise professional judgement when making inferences about learning and about the factors possibly contributing to such learning. But how is expertise in making such judgements acquired and developed?

In considering how an 'expert connoisseur' makes aesthetic judgements, the philosopher Wittgenstein (1980) gets close to characterising the process. He comments:

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 laws.

Applying these insights to inspection implies that professional expertise cannot be acquired from 'a course' or, at least, not just from a course or series of courses. It involves learning from a wide range of teaching and inspection experience in a variety of relevant contexts, preferably not confined to a single geographical area. It involves looking at and comparing a large number of lessons by 'various masters again and again'. It is not like 'learning to calculate' or its equivalent – learning from an inspection rule book or tick list. It involves learning from others more experienced in making judgements of teaching quality who can 'hint' at what is required and who can discuss the complexities and intangibles of classroom observation - hopefully as a result in part of joint observations. Like connoisseurs, inspectors should be able to 'list reasons' for their judgements, but these can never be absolutely 'convincing' given the difficulties involved in interpreting learning. The value of the judgements and the evidence they use to back them up depends on the experience and knowledge of the person making them. Quoting Wittgenstein (1980), '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' enshrined in any inspection handbook or subsidiary guidance.

Only those with the relevant experience and insight into the complexities and imponderables of the inspection/observation process can appreciate, live with and defend the tentative, partial, but necessary judgements involved in reporting on teaching quality in lessons.

Whether lessons themselves should be graded is an allied but different issue which Ofsted has valuably opened up for professional discussion.

References

- Ofsted (2014a) *School Inspection Handbook*. London: Ofsted.
Ofsted (2014b) *Subsidiary Guidance*. London: Ofsted.
Ofsted (2014c) *The Framework for School Inspection*. London: Ofsted.
Wittgenstein, L. (1980) *Reflections on the Philosophy of Psychology*, vol. 1. Oxford: Blackwell.

COLIN RICHARDS is currently Emeritus Professor of Education at the University of Cumbria and has been an honorary professor at the University of Warwick and a visiting professor at the universities of Leicester and Newcastle. A former primary school teacher and deputy head, he was an HMI from 1983 to 1996 specialising in primary education and in teacher education and holding the post of Staff Inspector (Curriculum 5-16) followed by that of Ofsted's Specialist Adviser for Primary Education. Since leaving Ofsted he has maintained a keen but critical interest in the issues of standards, primary and secondary education, the school curriculum, governance and inspection. He is chair of governors of a Cumbrian secondary school which has voted twice not to seek academy status. He has recently been designated as a National Leader of Governance. As a critic of much (though not all) of past and present government education policy, he is a frequent contributor to the national press (in particular to the *Guardian*, the *Observer*, the *Independent*, the *Times Educational Supplement* and *Education Journal*). He treasures the epithet 'an old-fashioned HMI' bestowed on him by a former chief inspector of schools who meant it as a damning criticism!
Correspondence: colin@sparkbridge.freeserve.co.uk