

Discussion piece:

taken from:



<http://cem.org/blog/414/>

Classroom observation: it's harder than you think

JAN 09 2014 Professor Robert Coe

We've all done it: observed another teacher's lesson and made a judgement about how effective the teaching was. Instinctively it feels valid. I am a good teacher; I'll know a good lesson when I see one. We've all experienced it from the other side – being observed – but this time the feeling may be more mixed. Sometimes you get real insight from someone who sees what you don't, questions what you take for granted and makes you think differently. Sometimes they just tell you what they would have done, or focus on some trivial irrelevance.

In September, I gave a talk at the [ResearchED2013 conference](#). Earlier in the day, Tom Bennett had talked about fashions that had been adopted widely, but uncritically, by teachers, such as Brain Gym, VAK, Left/Right-Brainedness. I sensed a hint of smugness among the delegates: they had never embraced such flakiness, or if they had it was short-lived and emphatically renounced. I couldn't resist trying to challenge any feelings of being 'holier-than-thou' by pointing out something they were still doing, but for which evidential justification was no better than the flaky stuff: classroom observation.

The evidence shows that when untrained observers are asked to judge the quality of a lesson, there is likely to be only modest agreement among them. Worse still, even if they do agree that what they see is good practice, it often actually isn't. I will briefly outline some of this evidence, and then try to explain how something that feels so right can actually be so wrong.

Research Evidence: Can observers spot good teaching?

There are two key issues here. The first concerns the extent to which the judgements made independently by two observers who see the same lesson would agree: in other words, **reliability**.

Fortunately, a number of research studies have looked at the reliability of classroom observation ratings. For example, the recent [Measures of Effective Teaching Project](#), funded by the Gates Foundation in the US, used [five different observation protocols](#), all supported by a scientific development and validation process, with substantial training for observers, and a test they are required to pass. These are probably the gold standard in observation (see [here](#) and [here](#) for more details). The [reported reliabilities](#) of observation instruments used in the MET study range from 0.24 to 0.68.

One way to understand these values is to estimate the percentage of judgements that would agree if two raters watch the same lesson. Using Ofsted's categories, if a lesson is judged 'Outstanding' by one observer, the probability that a second observer would give a different judgement is between 51% and 78%.

For observations conducted by Ofsted inspectors or professional colleagues, 'training' in observation is generally not of the quality and scale used in these studies, and no evidence of reliability is available. Hence, we are probably justified in assuming that the true value will be close to the worst case. In other words, if your lesson is judged 'Outstanding', do whatever you can to avoid getting a second opinion: three times out of four you would be downgraded. If your lesson is judged 'Inadequate' there is a 90% chance that a second observer would give a different rating.

The second key issue is **validity**: if you get a high rating, does it really mean you are an effective teacher? Unfortunately, the evidence here is even more worrying.

Strong et al. (2011) used value-added scores to identify 'effective' and 'ineffective' teachers, showed videos of them teaching to observers and asked them to say which teachers were in which group. In both the experiments where the observers were not trained in observation, the proportion correctly identified by experienced teachers and head teachers was below the 50% that would be expected by pure chance. At this level of accuracy, fewer than 1% of those judged to be 'Inadequate' are genuinely inadequate; of those rated 'Outstanding', only 4% actually produce outstanding learning gains; overall, 63% of judgements will be wrong.

How can the research evidence be so out of line with our intuition?

The belief that we know good teaching when we see it is so strong that it is a real challenge to be told that research does not support it. Faced with such a conflict, our instinct may be to reject the research. But here are five reasons why our belief may be wrong:

1. Observation produces a strong emotional response

When we sit in a classroom we absolutely know what we like. It is hard not to project our own preferences for particular styles or behaviours onto the situation, and compare what we see with what we think we would have done. Unfortunately, what we like may not correspond well with what helps others to learn.

2. Learning is invisible

Many people claim that you can tell whether children are learning by observing what happens in classrooms. We may think that if we focus on what the students are doing, rather than the teacher's performance, we will see learning happening. But learning is invisible and if we don't judge it by teacher behaviours, we have to rely on observable proxies for student learning. **I have listed some of these before:**

- Students are busy: lots of work is done (especially written work)
- Students are engaged, interested and motivated
- Students are getting attention: feedback, explanations
- Classroom is ordered, calm and under control
- Curriculum has been 'covered' (i.e. presented to students in some form)
- (At least some) students have supplied correct answers (whether or not they really understood them, could reproduce them independently or knew them already).

Of course, they are all related to learning but it is quite possible for any or all of them to be observed without any actual learning taking place.

3. Accepted 'good practice' may be more fashionable than effective

If we do look at the teacher, we will try to match what we see with what we know to be good pedagogy. Unfortunately, we are limited here by two things: our ability to define and operationalise specific pedagogic practices, and our knowledge about whether they really are effective. Moreover, if a group of observers are shown the same lesson they might well disagree about whether each of these practices has been seen.

4. We assume that if you can do it you can spot it

For experienced teachers, their classroom behaviour is mostly automated and sub-conscious. If you do not know exactly what it is you do in a way you could describe explicitly, it may be difficult to recognise when you see someone else do the same. Even an effective teacher may not understand fully which bits of their practice really make a difference. And if the observing teacher is experienced but not particularly effective themselves, it may be even less likely they will be able to identify effective practices.

5. We don't believe observation can miss so much

A number of quite surprising studies in psychology show that when people try to focus on observing particular things, they can miss an extraordinary amount, a phenomenon known as inattention blindness. For example, in [one study](#) around half the observers failed to spot a gorilla walking through the middle of a basketball game when their attention was focused on specific events.

What should we do?

A number of people have written blogs recently about classroom observation, including [@joe_kirby](#), [@learningspy](#), [@oldandrewuk](#), [@headguruteacher](#), [@tombennett71](#), [@Cazzypot](#), [@HeyMissSmith](#), [@StuartLock](#), [@Katiesarahlou](#), [@samfr](#). Various suggestions include that we should disband or reform Ofsted, that lesson observation should be used only formatively, or that we should stop doing it altogether.

My position is this: I am not against Ofsted; I am certainly not against inspection; nor am I against classroom observation. In fact there are good reasons for wanting to observe teaching. It is hard to imagine a credible evaluation system that doesn't include some observation; when done properly it does contribute to a valid judgement of teaching quality; and the process of having to think harder about how to do it should help us to understand better what effective pedagogy means. However, if we are going to do it we should do it in the most defensible way possible. Specifically, I think we should:

- Stop assuming that untrained observers can either make valid judgements or provide feedback that improves anything
- Apply a critical research standard and the best existing knowledge to the process of developing, implementing and validating observation protocols
- Ensure that good evidence supports any uses or interpretations we make for observations. It follows that appropriate caveats around the limits of such uses should be clearly stated and the use should not go beyond what is justified
- Undertake robustly evaluated research to investigate how feedback from lesson observation might be used to improve teaching quality (EEF already has [one such study](#) underway).

This certainly does have implications for Ofsted, and I have [written](#) and [spoken](#) previously about other aspects of the inspection process that 'require improvement'. However, it also has implications for every school and college leader who wants to use lesson observation as part of an evaluation or quality improvement process.

We all need to raise our game on this one.

Slides from an event organised by Teacher Development Trust and TeachFirst on 13 Jan 2014, "The role of lesson observation in England's schools", can be found [here](#).

Video of the event can be found [here](#).

Posted by [CEM Administrator](#)