‘How educational research gets done: looking through the lens of a study into lesson observation’ Keynote talk delivered by Matt O’Leary at the HELS Faculty Research Conference 13th July 2016

Introduction

On first reading the opening title of this keynote address, you would be completely justified in thinking that it sounds rather prescriptive or even obnoxious. It makes it sound as though there is A WAY of carrying out educational research and the talk intends to explain what that way is and how you should go about doing it. Well, rest assured, nothing could be further from the truth. I realise that for some researchers that may indeed be the case when it comes to aligning themselves to a particular research paradigm, but not for me as a mixed methods researcher. For mixed methods researchers, ‘the world is not exclusively quantitative or qualitative; it is not an either/or world but a mixed world’ (Cohen, Manion & Morrison 2011, 22). I am a firm believer that both paradigms and their associated methodological approaches have their strengths and weaknesses, as others have argued and that ‘even greater strength can come from their appropriate combination’ (Gorard & Taylor 2004, 1).

One of the key aims of my talk was to focus on the ‘doing’ of educational research through the lens of a national project that I was commissioned to undertake on behalf of the University and College Union (UCU) in 2012-13, which ultimately resulted in the publication of a final project report along with several other outputs. During my talk I used this particular project as a window through which to explore some of the practicalities of real world research and the various stages involved in carrying out a piece of research from start to finish, whilst also including a glimpse of some of the findings from the project and its implications for future models of teacher evaluation, along with my recommendations on how the education sector might best make use of observation as a method.

The talk discussed the policy backdrop to the research topic, the origins of the project, the research design, along with some of the practical considerations involved in analysing the research data, disseminating the findings and the implications for future practice. During the
course of this short paper I attempt to touch on each of these aspects briefly in my discussion and in so doing cover the following sections:

1. Policy context and the origins of the project
2. Research design and making sense of the data
3. A peek at some of the findings
4. Implications for future practice

1. The policy context and the origins of the project

Global interest in improving education systems has risen sharply in recent years. Fuelled by the ever-growing importance of comparative performance data from international assessment systems such as the OECD’s (Organisation for Economic Co-operation and Development) flagship Programme for International Student Assessment (PISA), the drive for continuous improvement in educational standards has undoubtedly become a high priority for many governments worldwide (Meyer & Benavot 2013). With increasing links being made between a country’s economic competitiveness and its levels of educational achievement, politicians and policy makers are keen to identify the formula for success. The importance of teachers and the quality of teaching in student achievement have figured prominently in recent studies, with particular interest in research exploring teacher effectiveness in the hope of pinpointing the skills and qualities displayed by the ‘effective teacher’ (e.g. Darling-Hammond 2005; Stronge et al 2011). As Sellar and Lingard (2014) have highlighted in recent work, the ‘Holy Grail’ for the OECD would be to align the performance data from PISA with other performance metrics used to measure teacher impact and effectiveness such as the Teaching and Learning International Survey (TALIS). This interest has filtered down into national education policies and priorities and has resulted in an increasing focus on measuring the impact of teachers’ work. One of the key tools used to do this has been through the process of observing classroom practice.

Lesson observation is a polemical issue that has attracted much discussion amongst the education community in recent years. In the Further Education (FE) sector, where this research was carried out, such discussion is not new, but has been simmering away on the sector stove for up to a decade now. In the lead up to this project, the use of observation
had already been identified as an increasingly common flash point in colleges, triggering local negotiations, and in some places industrial disputes. The level of discontent associated with its use had even resulted in the boycotting of lesson observations altogether by union members in some colleges. Where industrial disputes had occurred, this had often been in response to what practitioners perceived as draconian and punitive observation policies imposed on them by senior management. Such policies largely sought to link the outcomes of formal, graded lesson observations to disciplinary or capability procedures. Thus the policy backdrop to the project was one in which observation had become a highly charged and contested area of practice for the FE workforce.

2. Research design and making sense of the data

The project adopted a mixed-methods approach involving quantitative and qualitative methods of inquiry. Online questionnaires, semi-structured interviews and focus groups were the main research tools used as part of a triangulated framework to address the project’s research questions. The sample for the first phase of the data collection (online questionnaire) comprised UCU members working in the FE sector and ranged from part-time tutors to senior managers. The second phase involved staff from several colleges across England, including UCU members and non-members. A purposive sampling strategy was used to select the colleges. It was purposive in the sense that I wanted to ensure a geographical spread, thus colleges were selected from the north, the midlands and the south of England.

The amount of data generated during the course of the project was overwhelming to the extent that I felt in danger of drowning in data even before I had started to analyse it. What became clear early on in the data collection stage was that the use of lesson observation and its impact on the professional lives of the FE workforce was something that all the participants involved in this study felt very strongly about. To say that there was no shortage of opinions in all of the data collected for the project does not do justice to the magnitude of the responses. The qualitative responses in the online questionnaire are a good example of this. At the end of the questionnaire respondents were given the option of writing additional qualitative comments about the topic. Just under half of those respondents who completed the survey \( n = 1619 \) wrote comments in this section, which in itself is testament
to the fact that lesson observation was a topic of significant interest to them. To put this into perspective, these comments amounted to over 100,000 words of text or the equivalent of over 350 pages in Microsoft Word. In short, whether it was written or verbal comments, lesson observation was undoubtedly a topic that generated a lot of discussion among the study’s participants and it was clearly something about which they had a lot to say and wanted to make sure their voices were heard.

3. A peek at some of the findings
To get an overview of the contexts and purposes for which lesson observation was used in FE, a specific question was included in the online survey that required respondents to indicate the context that best described their most recent experience of observation. Figure 1 below presents a summary of the responses to that question.

![Table](image)

**Figure 1 – Contexts of lesson observation**

As Figure 1 illustrates, only just over a tenth of respondents (11%) chose peer review/development, suggesting this particular use of observation was relatively marginal in FE, with much of it taking place in the context of Initial Teacher Education (ITE) courses. The most common response selected by over two thirds (68.6%) was the Internal Quality
Assurance (QA) scheme, which typically mimics the approach adopted by Ofsted when carrying out lesson observations during inspections i.e. where the lesson is evaluated and graded against its 4-point scale. Similarly, the context of ‘external consultation’ follows the Ofsted model and is used as a ‘Mocksted’, where external consultants are employed to carry out observations across the institution. So when combined, the first three contexts listed in Figure 1, all of which adopted a similar performance management approach, amounted to over four fifths (84%) of responses.

Although there was evidence of ‘alternative models’ in practice, they were relatively marginalised and tended to operate on the peripheries of most formal systems of accountability. For example, ungraded models of observation were in use in some institutions, though only accounted for a tenth of current practice. Similarly, peer observation, whilst not uncommon, occurred mainly as an informal, unaccredited activity that staff undertook on a voluntary basis. These alternative models were rarely viewed by senior management teams (SMTs) with the same level of importance as their performative counterparts and tended to be valued more highly by practitioners. Furthermore, there was evidence of apprehension among some providers in implementing alternative and/or ungraded models of observation on a formal basis for fear of going against normalised practice and leaving themselves open to increased scrutiny from Ofsted.

The fact that lesson observation has become associated with performance management in education over the last two decades is no accident but a consequence of the wider political and economic ‘neo-liberal reform agenda’ that has sought to ‘transform the working cultures of public sector institutions’ (O’Leary 2014: 11). During this period, the medium of classroom observation has evolved into a catch-all crucible in which performance data and evidence of improvements in teaching and learning simultaneously bubble away. Driving this reform agenda are three interwoven policy ‘technologies’ that have shaped teachers’ work, notably ‘markets’, ‘managerialism’ and ‘performativity’ (Ball 2003). The way teachers experience these ‘technologies’ in their working lives is concretised through systems of quality assurance, target setting, continuous improvement etc, or what Green (2011) neatly refers to as the language of ‘managerialese’.
In discussing the impact of high-stakes testing on teachers’ work, Stevenson and Wood (2013) argue that ‘it is the pincer movement of markets and managerialism that have combined to effectively and radically re-shape teachers’ experience of work’ (p. 43), underpinned by the aim of making the complex processes of teaching and learning measurable. Smith and O’Leary (2013) refer to this as ‘managerialist positivism’, where the complexity of the teaching and learning process is reduced to the presentation of quantitative performance data, which is valued more highly than the qualitative kind on the premise that its outward-facing measurability imbues it with an increased authenticity and credibility:

Managerialist positivism functions to provide and impose quantitative wholeness on the unfinished totality of the present for the (political) purposes of policy intervention and the allocation of financial resource. It can be seen as the ideological veil that normalises the representation of complex sociological and qualitative phenomena in reductive and numerical forms (Smith and O’Leary 2013: 246).

4. Implications for future practice
The project’s key findings revealed an appetite for change among the majority of participants as to how observation is currently used in FE, as well as the ways in which teacher performance is assessed and managed. The first recommendation to emerge was the need for observation to be exploited more as a formative, supportive intervention with the emphasis on helping teachers to improve their practice rather than the current deficit model that focused mainly on attempting to measure and judge it in an isolated, episodic manner. Thus the professional needs of practitioners should be placed at the forefront of any observation scheme rather than the production of performance management data. Providing both observees and observers with opportunities to engage in substantive professional dialogue as part of the observation process is vital, though if this is going to work in practice then sufficient time needs to be allocated and embedded into teachers’ workloads at the start of each academic year. This time allocation needs to allow for: 1) a pre-observation meeting; 2) feedback and 3) feed-forward meetings as well as the observation of the lesson itself as a minimum commitment.
Following on from this, the second recommendation identified the need for a shift towards more peer-based models of observation. While only a small minority of participants expressed a desire to see an end to the use of lesson observation per se as a form of teacher appraisal, the majority acknowledged that it had an important role to play in teacher assessment and development. They did so, however, on the proviso that certain models were deemed to be more beneficial than others, particularly peer-based models with a focus on enhancing professional learning. These have the potential to redress some of the power imbalances associated with top-down, deficit models and encourage a greater sharing of practice and dialogue that could be mutually beneficial for observer and observee. This is perhaps best encapsulated in the comments of one of the project’s participants:

The best model has to be one based on a genuine spirit of enquiry and research. To explore what’s happening in that messy business of learning, and to be a starting point for professional discussion and debate. This means that both people involved in the discussion - teacher and observer need to be equal partners in the process, both working to improve things for teachers and learners. This means sharing a common purpose - why are we doing this and what do we both need to get out of it? (Vera, observer)

Thirdly, there was an overwhelming desire to move away from a system that relies heavily on a narrow evidence base to a more fully-inclusive, multi-dimensional model of assessing teaching competence and performance. The current reliance on annual graded observations as a means of measuring a practitioner’s professional competence was considered an inequitable and reductive practice. Given the misgivings surrounding the validity and reliability of observation as a method of assessment discussed at length in the project’s report, there can be no justification in linking the outcome of formal observations directly to an institution’s capability or disciplinary policy. The project’s findings highlighted an increasing demand to make the process of teacher evaluation more inclusive by extending it beyond the lens of lesson observation and drawing on other sources of evidence (e.g. student feedback, student achievement data, peer review, self-assessment, external verification) so as to ensure a more triangulated and reliable evidence base for assessment.
How such different data sources might be harnessed into a coherent framework of assessment undoubtedly represents a challenge, but it is a challenge worth tackling for all those committed to the ongoing improvement of teaching and learning.

Fourthly, there is a need to explore alternative approaches to the current, dominant model of graded lesson observation. These alternatives should seek to draw on elements of existing practices but also make greater use of recent advances in the research of observation and cognate areas that lend themselves to exploiting the potential of observation as a medium for professional enquiry and learning. Such alternative approaches might therefore wish to draw on models like ‘differentiated observation’ or ‘lesson study’ as a starting point (see, for example, O’Leary 2014, Chapter 8, for an account of alternative models of observation).

Concluding comments
For any of the recommendations discussed above to lead to meaningful and sustained change, this requires a root and branch reform of the way in which observation is conceptualised and engaged with as a form of educational intervention. As a starting point, educational leaders, inspectors, policy makers and practitioners alike need to break free from the assessment straitjacket that currently constrains the perception and implementation of observation. In my view, this is arguably the biggest obstacle that prevents the English education system from harnessing the potential of observation as a tool for meaningful and sustained professional learning and development.

However, in order for this to happen, there needs to be a ‘thinking outside the box’ when it comes to how observation may be used as a source of evidence in the educational arena. Tinkering with prevailing normalised models of observation is, at best, only likely to have minimal impact and offer short-term solutions to longstanding issues. Although removing the graded element would certainly represent a step in the right direction, for example, it cannot be considered a panacea in itself. In a similar vein, recent calls for the abolition of lesson observation from the inspection process are a classic example of ‘throwing the baby out with the bath water’ and as such represent a knee-jerk reaction to a much more complex problem than the one they claim to solve. Ultimately, what both of these strategies
fail to address are the deep-rooted political and epistemological issues surrounding the use of observation as a method of assessment. At the heart of any such discussion is the acceptance that the use of observation is not purely an act of pedagogy but one that is underpinned by issues of hierarchical power and professional trust. Until these issues are acknowledged and discussed by education professionals in an open forum, then any attempts at reforming the way in which the sector makes use of observation are unlikely to progress.

References