

Surveillance, performativity and normalised practice: the use and impact of graded lesson observations in Further Education colleges

Matt O'Leary

School of Education, University of Wolverhampton, Walsall, UK,

University of Wolverhampton, School of Education, Walsall Campus, Gorway Road, Walsall, WS1 3BD, UK. Email: moleary@wlv.ac.uk

Abstract

In little over a decade the observation of teaching and learning (OTL) has become the cornerstone of Further Education (FE) colleges' quality systems for assuring and improving the professional skills and knowledge base of tutors. Yet OTL remains an under-researched area of inquiry with little known about the impact of its use on the professional identity, learning and development of FE tutors. This paper examines the specific practice of graded OTL and in so doing discusses findings from a mixed-methods study conducted in ten colleges situated across the West Midlands region of England. Data from a questionnaire survey and semi-structured interviews were analysed within a theoretical framework that drew largely on aspects of Foucauldian theory as well as the twin phenomena of new managerialism and performativity. This analysis revealed how OTL has become normalised as a performative tool of managerialist systems designed to assure and improve standards, performance and accountability in teaching and learning. It is argued that FE has now outgrown graded OTL and it is time for a moratorium on its use. Colleges and tutors need to be given greater professional autonomy with OTL and be allowed to develop their own systems that place professional learning and development at the forefront rather than the requirements of performance management systems.

Keywords: lesson observation, observation of teaching and learning (OTL), normalisation, surveillance

Introduction

The sheer volume of initiatives and policies produced under the New Labour government was testament to its efforts to reform the Further Education (FE) landscape in England. A key driver of this reform agenda was the focus on strategies to develop and improve the professional skills and knowledge base of the sector's teaching staff. The effectiveness of such strategies has been critiqued in some studies (e.g. Colley *et al.* 2007), as indeed has the thorny and contentious issue of what constitutes professional development in FE and how it takes place (e.g. Bathmaker & Avis 2005; Gleeson & James 2007; James & Biesta 2007; Orr 2008). Observations of teaching and learning (OTL) play an important role in this domain since they are regarded in some circles as the cornerstone of colleges' quality systems for teaching and learning (Hatton 2008; Ofsted 2008a).

This paper reports on findings from a study into the use of OTL and its impact on the professional identity, learning and development of staff in ten FE colleges in the West Midlands in England. Though the research investigated the full range of OTL in use, discussion in this paper is limited to the common practice of *graded OTL* i.e. where lessons are assessed and graded against Ofsted's 4-point scale (Ofsted 2008b).

The first half of the paper provides the policy backdrop to how and why graded OTL emerged as a key element in Quality Assurance (QA) and Quality Improvement (QI) systems of teaching and learning in FE. As part of this discussion, Ofsted, whose role in shaping FE sector policy has markedly increased in recent years, is identified as the principal hegemonic force behind the development of this practice.

The second half of the paper presents an overview of the study's methodology in which the rationale, sample, context and research methods are outlined. The theoretical framework of the study is examined, centring on the work of Foucault (1977, 1980, 2002) and his notion of normalisation in the context of surveillance. This is followed by a discussion of key themes and issues to emerge from an analysis of the study's data across all ten colleges.

Policy background

The emergence of OTL in FE: putting teaching and learning at the forefront of improvement

OTL is a relatively new phenomenon in FE. It is only during the last two decades that colleges in England have witnessed its widespread use, yet in this short space of time it has become the cornerstone of QA and QI systems for teaching and learning

(Armitage *et al.* 2003). Understanding the context and rationale for its emergence requires the sewing together of a patchwork quilt of policies and initiatives produced as part of the reform agenda that has characterised FE over the last two decades.

One of the key drivers of these reforms has been the prioritisation of teaching and learning as the foundation on which to build *continuous improvement* across the sector (e.g. Finlay *et al.* 2007; James & Biesta 2007; Lucas 2004). This policy focus arose out of a wider political commitment on the part of the New Labour government to promote the role of FE in ‘upskilling the nation’ (DfEE 1999, 2), which in turn acknowledged that ‘success depend[ed] on the skills of people working in the sector’ (DfES 2004, 3). This resulted in the development of a package of reforms aimed at raising standards and improving the quality of teaching and learning, such as the introduction of professional standards (FENTO 1999; LLUK 2006) and the enactment of a ‘Statutory Instrument’ (DfEE 2001a) by New Labour that made it a legal requirement for all teaching staff to gain an appropriate qualification. Such initiatives were designed to improve the professional skills and knowledge base of tutors and it was in this context from which OTL emerged as an important multi-purpose vehicle.

These initiatives were themselves linked to a wider neo-liberal reform agenda intent on transforming the working cultures of public sector institutions by introducing new systems of management from the private sector that were designed to improve levels of performance, productivity and accountability. This approach to management has since acquired the label of *new managerialism* and has become associated with the way in which FE colleges have operated since the early 1990s (e.g. Ball 2001; Gleeson & Gunter 2001; Randle & Brady 1997).

New managerialism is an umbrella term used to refer to a set of systems, processes, attitudes and values. According to Randle and Brady (1997, 125), it was a ‘style of management which emerged in the UK in the early 1980s and gradually spread throughout the Public Sector’. It comprised a package of management techniques taken from the private sector that were considered *successful* and could subsequently ‘be applied as a template for public sector institutions’ (*op. cit.*, p. 121) to improve levels of productivity and performance and to make the workforce more accountable.

A central tenet of new managerialism was the view that workers could no longer be trusted to do their jobs efficiently and effectively (Robson 1998). This led to the introduction of audit systems and mechanisms of accountability and ‘performativity’¹ to monitor output and performance (Ball 2003). The measurement of teachers’ performance and productivity was a key part of this new culture and it was in light of this that OTL emerged as an important means of gathering evidence for colleges’ quality systems and preparing for Ofsted inspections.

During the last decade or so, FE has been subjected to a continuous stream of policies and initiatives aimed at raising standards and improving the quality of teaching and learning, as evidenced by the raft of publications to emerge (e.g. DfEE 1999, 2001b; DfES 2002, 2003, 2004, 2005, 2006; Ofsted 2003, 2004a, 2004b, 2006, 2008a). Coffield and Edward (2009, 373) have likened government policy to ‘a ratchet screwdriver with no reverse movement allowed; only constant forward progression is acceptable’. To reinforce their point they quote from *Learning to Succeed*, one of the first publications to present New Labour’s ‘vision of a lifelong learning society’² (DfEE 1999, 3) and one of the many that stress how important it is that ‘all provision

must be consistently good and continuously improving' (43). The terms *continuous improvement* and *quality* have thus become the mantra of FE sector policy, especially in the domain of teaching and learning.

Since the turn of the millennium two policy developments in particular can be linked to the increased use of OTL in FE, the first saw the introduction of a new set of professional standards and mandatory qualifications for FE tutors by the Further Education National Training Organisation (FENTO) in 1999, subsequently revised by Lifelong Learning UK (LLUK 2006). The second development involved the formation of a new inspection framework, with Ofsted and the Adult Learning Inspectorate (ALI) working alongside each other from 2001 before merging into one inspectorate in 2007.

In 2004, a key government report (DfES 2004) recommended more emphasis on the practical aspects of teaching and highlighted the importance of OTL in particular. It called for 'more effective observation of teaching practice' (16). Another outcome of the report was that LLUK was commissioned to develop a new set of professional teaching standards for FE. The catalyst for this was a set of recommendations made by Ofsted (2003), in which eight HEIs and 23 FE colleges were visited by both Ofsted and ALI inspectors 'as part of a national survey to evaluate the quality and standards of ITT in FE' (2003, 1). The overriding judgement of the report was that the FENTO standards '[did] not provide a satisfactory foundation of professional development for FE teachers at the start of their careers' (p. 2). The report revealed a wide variance between ITT providers in applying the standards; a view later echoed by Lucas' (2004, 42) claim that 'there seems no standards for applying the standards'. In many

cases the standards were simply being mapped to existing ITT programmes in such a way that the application process became purely perfunctory.

OTL was identified as a weakness in FENTO programmes. The FENTO guidelines for OTL were criticised for failing to regulate its use by not specifying precisely how often it should take place and the minimum number of assessed hours of OTL required. This led to a lack of uniformity in practice across providers, an area which both Ofsted and LLUK were keen to rectify. The prominence given to OTL as a key criterion in the assessment of teachers' professional practice in the LLUK standards was thus a noticeable modification to the previous FENTO ones, as evidenced by the requirements below for those following the 'full teaching role' route:

There must be a minimum of 8 observations totalling a minimum of 8 hours. Any single observation must be a minimum of half an hour. (This excludes observed practice completed for PTLLS.) Observations can be formative and summative (LLUK 2007, 23).

This was a clear attempt to address what Ofsted had originally identified as a lack of standardised practice in the assessment of trainees' teaching as well as a sign of their increasing influence over FE policy in the classroom. It is to Ofsted's role in FE policy that the next section of this paper now turns its attention.

The emergence of OTL in FE: the rising role of Ofsted

Along with the introduction of the new LLUK standards and qualifications (2006), the decision to hand over responsibility for all FE inspections to Ofsted marked another significant act in recent years. On the one hand, it could be seen as recognition of how standards in teaching and learning were a high priority in the government's reform agenda in FE. On the other hand, it might also have been indicative of an underlying

aim to ‘police’ the sector (Gleeson *et al.* 2005) in order to ensure that performance targets were met and *continuous improvement* remained at the top of the agenda.

Prior to the implementation of the new inspection regime, from 1993 to September 2001, the Further Education Funding Council (FEFC) had been responsible for carrying out college inspections. The nature of these inspections was quite distinct from the Ofsted model insomuch as they were regarded by staff as more benign, supportive and ‘much less charged’ (Grubb 2000, 717). The replacement of the FEFC by the Learning and Skills Council (LSC) in April 2001 along with the introduction of Ofsted heralded an era of change with *quality* and *continuous improvement* as the drivers of the new regime. Whilst it is important to acknowledge other key players in the ‘drive for excellence’ and improvement, such as the LSC and the Learning and Skills Improvement Service (LSIS - formerly known as the QIA), the impact of these agencies on teaching and learning in FE has arguably been less tangible than that of Ofsted. Ofsted’s involvement in and influence over the FE sector grew considerably during New Labour (Whitehead 2005). It was extended beyond its original remit of inspecting standards in colleges to one of defining them. It has taken a leading role in shaping the QA and QI policy agenda, as evidenced by some of the key publications to emerge in recent years that have influenced management in FE colleges (e.g. Ofsted 2003, 2004a, 2004b, 2006, 2008a).

As discussed in the previous section, one of Ofsted’s criticisms in its survey of ITT provision in the sector was that there was insufficient attention given to OTL under the FENTO programmes (Ofsted 2003). The central role of OTL in self-assessment in FE was also underlined in two parallel reports that were later to become seminal

documents for college senior management: *Why colleges succeed* and *Why colleges fail* (Ofsted 2004a & b), and more recently *How Colleges Improve* (Ofsted 2008a).

In the 2004 reports, ‘underperforming colleges’ were criticised for having OTL schemes that were ‘poorly conceived and implemented’ (2004a, 14). In colleges where there was evidence of ‘a reluctance to grade observations which are viewed as developmental rather than quality control activities’, Ofsted concluded that such schemes were ‘insufficiently robust, partly because the observations are not graded’ (*Ibid.*). The diction of both reports (2004a & b) implied that without the use of a grading scale, the OTL data were deemed less valid and reliable. Similarly in the 2008 report (Ofsted 2008a), graded OTL was emphasised as a vital tool in the performance management of standards in teaching and learning. There were references to OTL on almost every page, highlighting the importance attached to it by Ofsted.

All three reports reveal an underlying value system that favours ‘hard’ (i.e. quantitative) over supposed ‘soft’ (i.e. qualitative) data on the basis that the former is measurable and thus considered more objective and credible. There is an assumption that numerical grades have an objective value comparable to that of calibrated measuring devices such as weighing scales. By attaching a grade to the subjective judgement of the observer, people are seduced into believing that such judgements have greater objectivity and authority than they can, in reality, claim to have. But using a numerical scale of 1-4 does not make the exercise of evaluating practice any more reliable or valid than if the observations were ungraded or subject to an alternative form of summative assessment. In their study into the grading of student teachers on teaching practice placements in Scotland, Cope *et al.* (2003, 682) found

that the success of such practice depended on ‘a clearly reliable and valid system of assessment of the practice of teaching. The evidence available suggests that this does not currently exist’.

This is not a phenomenon specific to observation as a method of assessment, but reflects widely held beliefs among researchers in the field of assessment such as Gipps (1994, 167), who argued back in the 1990s that ‘assessment is not an exact science and we must stop presenting it as such.’ The danger is that the inherent limitations of practice such as graded OTL are often overlooked and the resulting judgments are given more weight and authority than they deserve, as was revealed repeatedly by the value attached to graded OTL in the study’s findings below. But before moving on to the findings, the next sections will examine the rationale for the study, its methodology and its theoretical framework.

Rationale for the study

A survey of previous studies of OTL in England revealed that it remains an undeveloped area of educational research, particularly in terms of recognised empirical work carried out amongst staff working in FE. Apart from a solitary case study by Cockburn (2005) and the recent work of researchers based at the University of Huddersfield into its use in the context of ITT (e.g. Burrows 2008; Ollin 2009; Peake 2006), extant research has tended to be restricted to schools in England (e.g. Marriott 2001; Montgomery 2002; Tilstone 1998; Wragg 1999; Wragg *et al.* 1996). To a lesser extent there have been studies carried out in the HE sector, albeit mainly focusing on the use of peer observation as a means of encouraging reflective practice (e.g. Hammersley-Fletcher & Orsmond 2004, 2005; Peel 2005; Shortland 2004).

Some of these studies have contributed to current knowledge and understanding of the topic, particularly in terms of its use as a formative mechanism. However, their focus has tended to centre on the *pedagogy* of OTL.

There is a lack of empirical studies exploring the perceptions and experiences of in-service tutors, especially in FE. In Jephcote and Salisbury's (2009) study into the working lives of FE tutors and the formation of their professional identities, they remarked that 'there is much research into the initial training of teachers, but much less so into the lives of experienced teachers' (967). This study sought to contribute to empirical work on OTL specifically, as well as to the body of research charting the perceptions of experienced tutors in 'a significantly under-researched sector' (James & Biesta 2007, 2). Besides, recent reports into how OTL has 'become an increasingly common flash point in colleges, triggering local negotiations, and in some places industrial disputes' (UCU 2009, 1) raise questions about the extent to which policy aims are being achieved and highlight the timeliness of this study.

Methodology

This was a mixed-methods study in which quantitative and qualitative methods of inquiry were used, incorporating questionnaires, interviews and document analysis. The study was situated in 10 FE colleges across the West Midlands region of England (see Table 1 below for a profile of the colleges) between May 2008 and March 2010 and its focus was concerned with investigating the ways in which the professional identity, learning and development of FE tutors was being shaped through the use of a particular intervention, namely OTL.

The total sample size was 500, consisting of 50 participants from each college. The sample comprised academic staff and included teaching staff, middle and senior managers. Participants were categorised into three stakeholder groups: (i) 'observees' (i.e. tutors), (ii) 'observers' (i.e. heads of school, advanced practitioners, etc.) and (iii) Senior Management Team (SMT) representative(s) (i.e. the person(s) whose remit it is to oversee the implementation and monitoring of OTL schemes within the institution). The three groups were identified on the grounds that they represented the main stakeholders involved in OTL. It was felt that the representation of all three groups would provide an illuminative approach to OTL by viewing it from different stakeholder perspectives. This 'multi-voiced' perspective (Lather 1991) helped to strengthen the validity of the data by presenting a triangulated perspective (Cohen, Manion & Morrison 2007).

The fieldwork for the study was divided into two phases. In the first phase, questionnaires were sent to a stratified random sample of 50 participants in each college. Curriculum areas were chosen as the stratum by which to select participants. This was to ensure that the sample was broadly inclusive of each college's overall provision and not biased towards particular areas. It was felt that it was important to incorporate a cross-curricular coverage in the sampling strategy, as the enquiry was relevant to all areas of the FE curriculum. The overall response rate was 52.4% with 262 completed questionnaires returned.

The second phase of the research involved a series of semi-structured interviews with staff from a sample of three case study colleges selected from the original group of ten. It was decided that by selecting three colleges with differing profiles in terms of

location, size, student ethnicity (i.e. colleges A, D and J in Table 1 below), not only would it provide a varied sample of interviewees, but also help to capture the diversity of the sector (Huddleston & Unwin 2007), which would ultimately help to reinforce the validity and reliability of the interview sample. Three observees and three observers were interviewed from each of the three colleges, as well as three SMT participants. The observees and observers were selected by a combination of random and purposive sampling; the three SMT participants were all purposively selected on account of their position within their respective colleges, as OTL schemes fell under their remit and as such they were responsible for overseeing this area. Thus there were twenty one interviewees in total: nine each from the observers and observees, and three from the SMT group. Participants from each of the three groups were interviewed with the common aim of exploring their attitudes and beliefs about OTL.

Table 1 – Profiles of participating colleges

College	Location	Size	Number of students enrolled in last year	Percentage of ethnic minority students	Outcome of most recent Ofsted inspection (overall grade)
A	Urban	Large	25,000	30%	2
B	Suburban	Large	14,000	25.3%	2
C	Urban	Medium	10,000	30%	3
D	Rural	Small	1,800	0.3%	3
E	Suburban	Large	20,000	15%	2
F	Suburban	Medium	9,000	unknown	2
G	Urban	Large	15,000	36.5%	2
H	Suburban	Large	17,000	7.2%	1
I	Urban	Medium	10,000	28%	2
J	Urban	Medium	8,000	1.3%	2

Theoretical framework of the study

The study drew on and was informed by the work of Foucault (1977, 1980, 2002) as its theoretical backbone, along with concepts previously discussed relating to theories

of new managerialism (e.g. Randle & Brady 1997; Robson 1998) and performativity (Ball 2001, 2003).

Foucault's work provided a suitable framework for analysing the phenomenon of OTL for several reasons. Some of the key concepts he explored i.e. power-knowledge, surveillance, discourse and normalisation, resonated with themes to emerge from the study's empirical data and the related literature. These key concepts provided a useful lens through which to examine relationships of individual agency and structure, as well as a language with which to describe and discuss the phenomenon of OTL.

For Foucault power and knowledge are inextricably linked. This has particular resonance for OTL given that it is regarded as one of the most important means of collecting knowledge about what happens in classrooms and what it can reveal about teaching and learning. Thus, viewing OTL through a Foucauldian lens inevitably connects its use to power, which means that if we are to construct a detailed understanding of OTL then we must also understand the mechanisms of power that underpin it. Foucault believed that knowledge was a social product created by a number of connected mechanisms. These mechanisms of knowledge production act as what Foucault referred to as 'apparatuses of control' to establish certain forms of knowledge as more legitimate than others. Those with the greatest command of such mechanisms are able to create 'regimes of truth' (Foucault 1980, 102).

Regimes of truth are 'the types of discourse which it [society] accepts and makes function as true' (1980, 131). For Foucault *truth* is not to be understood in the conventional sense as an empirical fact proven and accepted to be true by society, but

as a notion that is ‘linked in a circular relation with systems of power that produce and sustain it’ (2002, 132). These systems of power determine the *rules* or the ‘ordered procedures for the production, regulation, distribution, circulation and operation of statements’ (*Ibid.*). Regimes of truth emerge from the connections formed between these dominant discourses. In paraphrasing Foucault, Brookfield (2005, 138) states that ‘dominant discourses inevitably reflect and support existing power structures’. In the case of OTL, the production of such dominant discourses and regimes of truth is exemplified by agencies like Ofsted, who are the custodians of quality and standards.

An interesting avenue of inquiry explored in this study was the extent to which these dominant discourses shaped how those working in FE made sense of and interpreted the function of OTL, as well as defined the parameters within which they were given licence to do so. As discussed below, this area of inquiry also extended to examining the role of dominant discourses in the process of ‘normalisation’ and how this related to notions of *good practice*.

Normalisation can be defined as the adjustment of behaviour to fall into line with prescribed standards (Perryman 2009). In the case of graded OTL, normalisation can be used as a means of conceptualising the process by which tutors operate within the accepted norms of good practice, a concept largely determined by governmental agencies such as Ofsted. Foucault (1977, 184) asserted that ‘the power of normalization imposes homogeneity’. The ‘homogeneity’ that Foucault refers to is imposed by the requirement for all tutors to demonstrate standardised notions of good practice during OTL. Those that are able to manifest such normalised behaviour form a homogenous community; those that fail to do so are identified through ‘gaps’ in

their assessed performance. The means by which such gaps are measured and levels determined is through a procedure that Foucault referred to as the *examination*, which ‘combines the techniques of an observing hierarchy and those of a normalizing judgement’ (*Ibid.*). In this paper the examination can be seen as a metaphor for graded OTL in FE, where a tutor’s performance is categorised and differentiated by the observer according to Ofsted’s 4-point scale.

Findings and Discussion

The hegemony of QA OTL and the surveillance of standards

Where does OTL fit into college systems and what seems to be its main purpose(s)?

In order to get an overview across all ten colleges of the contexts and purposes for which OTL was used, a specific question was included in the questionnaire that required respondents to indicate the OTL context that best described their most recent experience in the last year. **Table 2** below presents a statistical summary of the responses.

Table 2 – OTL contexts

	Frequency	Percent	Valid Percent	Cumulative Percent
Inspection	26	9.9	9.9	9.9
External Consultation	12	4.6	4.6	14.5
Internal Quality Review	172	65.6	65.6	80.2
Professional Development	32	12.2	12.2	92.4
Other	20	7.6	7.6	100.0
Total	262	100.0	100.0	

The most common response selected by two thirds was the internal quality review option, or the ‘QA model’, which typically mirrors the approach adopted by Ofsted during inspections. That is to say, the lesson is evaluated using the same criteria and

4-point grading scale (Ofsted 2008b), although unlike Ofsted inspectors, there is an expectation that observers will provide formative feedback. The ‘external consultation’ also follows the Ofsted model and tends to be used by colleges as mock inspection, where external consultants are employed to carry out OTL. Thus the first three contexts listed in **Table 2** all adopt a similar QA approach to OTL, which when combined yields over three quarters of responses.

The statistical data from the questionnaires supported claims made in other studies (e.g. Armitage *et al.* 2003) that OTL was mainly associated with QA and performance management systems in FE, although previous work has not been underpinned by the same type of statistical data presented here. These findings were reinforced qualitatively in college OTL policy documents and the comments of interviewees across all three groups. **Table 3** below reveals a sample of interview responses to the question: ‘What do you see as the main function(s) of OTL in your college?’

Participant	Comment
Graham, SMT	To improve standards, to help improve learner experience, success rates, etc.
Paula, SMT	Improving the quality of learning for learners, definitely.
Polly, observer	I think their primary purpose is to maintain standards.
Cyril, observer	I think observations first and foremost are for ensuring that quality is maintained and to ensure our learners are getting the best possible deal.
Debbie, observee	I think the idea of it is to make sure everyone is teaching to the same standards but I also think the college [management] genuinely wants to improve the quality of teaching and learning.
Ryan, observee	One of the reasons is because we’ve got to have them. It’s part of all these quality claims, you know, ensuring standards and improving quality.

Table 3 – Sample of interview responses to main function(s) of OTL

There was a commonality expressed in the diction of the responses that permeated all levels of college hierarchy. The words ‘quality’ and ‘standards’ were conspicuous as either one or both appeared in each comment, indicative of responses to the question given by the vast majority of interviewees. Such language is commensurate with the ‘dominant discourses’ (Foucault 1980) of QA divisions in colleges and relevant policy documentation, as typified by the opening statement of one college’s policy document on OTL, which simply states that ‘the purpose of lesson observation is to improve the quality of learning and teaching within the College’. The notion of *improvement* of the ‘learner experience’ or of the overall ‘quality of teaching and learning’ was also prominent and this reflected a consistent pattern across all colleges, though the extent to which there was agreement among the three groups of participants regarding how successful QA OTL was in achieving the aims of QI was a contested matter discussed in the next section.

The fact that there was such uniformity to the language used by participants to describe the main function(s) of OTL in **Table 3** above was neither a coincidence nor an idiosyncrasy of OTL, but a reflection of a wider discourse associated with the FE reform agenda. Given that college policy was largely determined by government policy and that of key external agencies such as Ofsted, LSC and LSIS, all of whom were linked by a shared agenda of *continuous improvement* and a *framework for excellence* for the sector (e.g. Ofsted 2008a; QIA 2007), it follows that the dominant discourses produced by them were likely to be assimilated and applied by QA divisions in the creation of internal policies across colleges. The repeated reference to quality and standards above in **Table 3** is a case in point, as well as the emphasis not simply on maintaining but on continuously improving them.

The emphasis on continuous improvement, surveillance of standards and the measurement of performance was evident at all levels of college hierarchy and was experienced by tutors and SMT alike. Cyril, a middle manager, described how this impacted on his role as an observer and how he was ‘measured by improving learning performance’, which was mainly judged by ‘improving the grade profile of the teaching staff’ through OTL. Similarly, Graham, a college QA director, pointed out that one of his responsibilities involved providing his SMT colleagues with monthly statistical updates on its internal QA OTL scheme:

On a monthly basis I report to the executive team on the current formal graded observation profile so the stats that they are looking for is the percentage of lessons observed that were good or better and then the percentage of lessons that were observed as inadequate and what we’re doing about it. So they don’t routinely hear about the other observation process in the College, which is the peer observation process.

Graham’s comments epitomised how QA OTL was prioritised over other models such as peer OTL. His comments also brought to light the importance attached to gathering quantitative data on graded OTL, or what Graham later referred to as the ‘much maligned lesson observation profile’ (i.e. statistical data of how many lessons were graded as a 1, 2, 3 or 4). In following on from his previous comment, he was openly critical of such practice:

At the end of the year in our self-assessment report, we will report on the number of ones, twos, threes and fours and I think it’s basically worthless but it’s something that all colleges do at the moment because it’s what Ofsted expects.

The reliance on quantitative performance indicators had, as Graham went on to say, become ‘custom and practice’ and thus an established feature of college performance management systems (QIA 2007). These quantitative data were used to measure and compare year on year OTL grades for self-assessment and inspection purposes and as such were seen as an important tool in tracking progress and measuring performance as part of the continuous improvement agenda.

The decision of colleges to adopt the standard practice of graded OTL for QA purposes is indicative of how Ofsted casts its 'normalising gaze' (Foucault 1977, 184) over the sector and thus exercises a form of panoptic control. It seems that Ofsted has hegemonised the FE workforce to view the main function of OTL as a performance indicator for categorising tutors and their professional practice according to its 4-point scale.

Phil, an observer working in the field of business and professional studies, neatly summarised Ofsted's panoptic presence when he said that 'the whole quality system in terms of teaching and learning is based around collecting evidence for Ofsted.' Perryman (2006, 2007, 2009) has argued in her work in the schools' sector that the gathering of evidence is a constant process that is not confined to the period of an inspection itself or even just the build-up to it. Ofsted expects such data to be gathered by colleges on an on-going basis as part of their self-assessment. The culture of collecting numerical/written evidence of continuous improvement is, as the term suggests, continuous. In the words of Foucault, 'surveillance is permanent in its effects, even if it is discontinuous in its action' (1977, 201).

Some tutors displayed feelings of resentment that they were 'being checked up on' (Gavin), reinforcing the notion of their work being 'policed' (Gleeson *et al.* 2005), while others accepted it as an inevitable part of the job such as Debbie. She summed up the feeling of many of her colleagues:

The truth is though that most of us actually just find it a pain because it's extra work. We're busy anyway and there is that nugget of resentment there that I would really rather put my thought and effort into planning this lesson and creating resources instead of having to spend all this time putting this paperwork together for this observation ... I suppose you could say it's sort of a necessary evil though.

Debbie's reference to OTL as a 'necessary evil' was symptomatic of the paradoxes associated with its use but it also offered a balanced and realistic reflection of the thoughts of many tutors. That is to say, they were committed to providing the best learner experience they could and thus acknowledged the role that OTL potentially had to play in this. Yet they resented the additional bureaucracy involved in having to compile the necessary 'evidence' associated with QA OTL.

Combining QA & QI purposes: a union of incompatibles?

A trend to emerge across data sets was how QA requirements took precedence over QI or the 'development' needs of tutors. OTL's potential as a tool for professional development thus tended to get lost. Cathie, a professional development manager, provided a concise insight into some of the difficulties faced by observers in attempting to dovetail two seemingly conflicting purposes into one OTL scheme:

I find it difficult to manage two very different roles for observations under the one umbrella, you know the quality assurance, Ofsted-type approach versus developmental and supportive feedback for the teacher. If I'm honest, I'm not sure I succeed in achieving both and I feel guilty about that because I'm a teacher trainer at heart and so instinctively I feel the need to help teachers develop.

Cathie's candid comments highlighted how observers often found themselves having to juggle the QA demands of colleges' audit systems, yet also build in time for supportive feedback and dialogue with observees that would facilitate their CPD and ultimately lead to improvements in the quality of teaching and learning. While many observers expressed a commitment to maintaining the importance of the latter, it was clear that their ability to uphold such a commitment was compromised by the prioritisation of the QA agenda in colleges coupled with practical time constraints. For one observer, the contrast between QA and QI driven OTL was epitomised in her differing roles as an ITT mentor and an observer for the college's internal QA OTL:

The [college] observation process for me is very different in its focus to the observations I do in my role as a teacher training mentor. The former is a quality control process aimed at providing quantifiable data whereas the latter is genuinely developmental and truly aimed at quality improvement. The latter is therefore eminently more valuable.

What arguably defines the ‘value’ associated with QI-type OTL is the nature of the relationship between observer and observee and the opportunity that the latter has to engage in substantive dialogue. In the QA context the data revealed a clear delineation of power between observer and observee based on hierarchical seniority, where the observer often took on the role of ‘judgement maker’. Such an approach can represent a significant threat to the developmental potential of OTL, as has been commented in previous work (O’Leary 2006). Or as one observee put it, ‘if observation is something that is *done to us*, then it’s unlikely to help us improve as teachers’.

The prioritisation of QA over QI did not go unnoticed by observees. A recurring theme to emerge from the qualitative data (both questionnaires and interviews) was the perceived lack of benefit of QA OTL to their CPD. Many tutors commented that the only beneficiary was the college SMT in terms of providing them with the necessary data to compare levels of performance to national benchmarks. Some tutors referred to college-wide OTL as a ‘tick-box’ exercise that seemed more concerned with satisfying the requirements of Ofsted than their development needs.

OTL as ‘part of the furniture’: standardising systems and normalising practice

Despite its short history, evidence across data sets suggested that OTL had quickly become embedded into colleges and was a well-established feature of QA systems. There seemed to be a consensus and clarity about its purpose(s) and the criteria used to assess tutors’ practice. As Jackie, a tutor, aptly remarked, ‘observation’s become part of the furniture in colleges now’. When asked to comment on the clarity of the

purpose of their most recent OTL, there was an overwhelming level of agreement in questionnaire responses, which manifested a joint appreciation on the part of observers and observees as to why OTL was being used, as shown in **Table 4** below.

Table 4 – The purpose of the observation was clear to me

	Frequency	Percent	Valid Percent	Cumulative Percent
Strongly Agree	148	56.5	56.5	56.5
Agree	104	39.7	39.7	96.2
Disagree	8	3.1	3.1	99.2
Strongly Disagree	2	.8	.8	100.0
Total	262	100.0	100.0	

A related pattern to emerge across data sets was how the established use of OTL had resulted in the standardisation of ‘systems’ associated with it i.e. the procedures involved and the type of documentation expected and produced by college staff. Phil, an observer, acknowledged that the model of OTL used in his college was one that sought to replicate the Ofsted approach during inspection visits. This was typical of how many colleges had incorporated it into their self-assessment systems as part of a ‘mock Ofsted’ in order to prepare staff for the ‘real thing’. It reinforced Phil’s earlier comment in the previous section that QA systems for teaching and learning tended to revolve around ‘collecting evidence for Ofsted’. As Avis (2003: p. 322) argued:

Although providers have a degree of autonomy over self-assessment, the Common Inspection Framework forms the basis of that process setting the terrain upon which self-assessment documentation is constructed.

As many interviewees confirmed, this typically involved making a ‘course file’ available to the observer, which was expected to include a current scheme of work, present and past lesson plans, pen portraits of the learners, samples of learners’ assessed work, etc. Besides the standardisation of documentation, the data revealed

recurring patterns of *normalisation* of classroom practice i.e. tutors consciously shaping what they did during OTL to fit in with prescribed notions of ‘good’ or ‘outstanding’ practice as outlined by agencies such as Ofsted and LLUK (e.g. Ofsted 2008b; LLUK 2006). There were many instances in the data of tutors being encouraged to adopt templates of ‘good practice’ so as to achieve a high grade. Such templates were often cascaded down from SMT as Graham, a QA director, revealed:

I sent everyone a support pack and information about the common inspection framework and key questions and top tips to try and give them a focus, probably too much! The one that was the most liked was the top tips because it was almost like a checklist for good lessons. I had a number of grade ones saying ‘all I did was go down that checklist and tick everything off’ and I thought wow, this works.

Terry, an engineering tutor with over twenty five years’ experience, provided a candid and detailed account of how he followed a similar ‘checklist’ to the one referred to by Graham above in the planning and delivery of a recent OTL to achieve a grade one:

So you know your lesson plan inside out. You make sure there’s a plenary, a couple of plenaries in there at the start and the end of the lesson. Put a load of crap in with regards to getting the students to do some sort of learning activity at the beginning to show that they have learnt from the previous week’s work, put your core tasks in and don’t forget that old chestnut about “differentiating” in the tasks you include! Give them a little quiz, move on to the next one and then make sure you do a good summary and do a nice little feedback session with them. Fiddle your scheme of work so you’re doing the lesson that you want to do, make sure that all the handouts have got the college logo on them and they’re all nice and neat with no smudges, do a lot of questioning, do a lot of walking around, then bring some work in with you so you can show that you’re giving them adequate feedback.

Terry was openly cynical of what was required to secure a high grade. His knowledge of ‘which boxes to tick’ was indicative of many astute tutors’ pragmatic response to the use of graded OTL and the need to ‘play the game’ (Jeffrey & Woods 1998). In other words, they were able to assimilate those features of pedagogy that had been identified as part of a *toolkit* for ‘good’ or ‘outstanding’ lessons and subsequently apply them to their own teaching during OTL. This resulted in such practice becoming normalised and adopted as the *default* model for all those tutors striving to achieve a high grade, which itself raises questions concerning the validity and

reliability of graded OTL as a means of assessing classroom performance, though discussion of this topic is beyond the scope of this paper.

Finally, questionnaire and interview data across all three participant groups repeatedly referred to the use of graded OTL as a form of ‘quality control’ of tutors’ teaching, with evidence of both punitive and rewarding consequences. This typically involved those tutors who were consistently awarded a grade one or two being granted an ‘OTL holiday’, which exempted them from being observed for the following year. In contrast, those who received a grade three or four were assigned a mentor whose responsibility it was to ‘up their performance’ in a short space of time.

There was also compelling evidence across data sets to suggest a process of implicit and explicit labelling of tutors’ teaching performance despite the hackneyed stance that it was the ‘learning’ that was being judged and not the tutor. The depth of feeling expressed in the comments of one questionnaire respondent below neatly encapsulated some of the tangible effects of grading tutors’ performance:

The grading of observations is divisive – we are given tables of how many people got which grade – it has almost become unhealthy competition – it’s unnatural too. Personally I hate the process though I get good grades. I live in fear of failing next time.

In short, the use of graded OTL in some colleges appeared to have had a restrictive and often negative impact on tutors’ professional identities and their notions of self.

Conclusion

The relentless drive for continuous improvement in teaching and learning over the last two decades in FE has resulted in increased levels of evaluative surveillance of tutors.

In keeping with a managerialist philosophy underpinned by the notions of

performativity and accountability, graded OTL has emerged as one of the key tools to measure and control what tutors do in the classroom. Such strategies have reduced the assessment of teaching and learning to a simplified rating scale and in so doing required tutors to ‘organise themselves as a response to targets’ (Ball, 2003: p. 215). Through the dominant discourses of government policy makers and agencies such as Ofsted, FE colleges have been hegemonised into viewing the main function of OTL as a performative tool for categorising tutors and their professional practice.

During this period of unrelenting reform, professionalism and professional identity have been routinely prescribed and codified by a managerialist agenda cascaded from central government, to aligned agencies and subsequently to colleges. The end result is that what it means to be a professional in FE has been continually subjected to and shaped by shifting political priorities. Through the lens of graded OTL, those who exhibit the characteristics of ‘preferred models’ of professionalism are rewarded, though this demands a sense of compliance and allegiance to the prescribed systems that form part of such models. Alongside these policy developments there has been a simultaneous reduction in the autonomy of colleges and tutors, with limited opportunities for them to shape and influence their professional development and identity.

The findings of this study have reinforced the idea that OTL has a significant role to play in improving the quality of teaching and learning as well as contributing to a greater understanding of these processes. However, in order to realise its potential to do so, there needs to be a fundamental re-conceptualisation of its current use in the FE sector. Such a course of action should aim to release OTL from the shackles of

managerialist control by which it is currently bound and to put the professional needs of tutors at the forefront rather than the requirements of performance management systems. In order for this to happen and to enable colleges to effect such change, policy makers need to devolve a greater degree of autonomy and trust to FE teaching staff to determine how mechanisms like OTL might best be used to meet their professional needs. One positive step towards this would be to implement an immediate moratorium on the use of graded OTL in colleges.

References

- Armitage, A. Byrant, R. Dunnill, R. Hammersley, M. Hayes, D. Hudson, A. and Lawless, S. 2003. *Teaching and training in post-compulsory education*. 2nd ed., Buckingham: OUP.
- Avis, J. 2003. Re-thinking trust in a performative culture: the case of education. *Journal of Education Policy*, 18, no. 3: 315-332.
- Ball, S. 2001. Performativities and fabrications in the education economy: Towards the performative society. In *The performing school: Managing teaching and learning in a performance culture*, ed. D. Gleeson and C. Husbands, 210–26. London: Routledge-Falmer.
- Ball, S. 2003. The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18, no. 2: 215-228.
- Bathmaker, A., and Avis, J. 2005. Becoming a lecturer in further education in England: the construction of professional identity and the role of communities of practice. *Journal of Education for Teaching*. 31, no. 1: 47-62.
- Brookfield, S. D. 2005. *The Power of Critical Theory for Adult Learning and Teaching*. Maidenhead: Open University Press.

- Burrows, J. 2008. *Trainee Perceptions of Observation*. Huddersfield: Huddersfield Consortium.
- Cockburn, J. 2005. Perspectives and politics of classroom observation. *Research in Post-Compulsory Education*, 10, no. 3: 373-388.
- Coffield, F., and Edward, S. 2009. 'Rolling out 'good', 'best' and 'excellent' practice. What next? Perfect practice?' *British Educational Research Journal*, 35, 3: 371-390.
- Cohen, L., Manion, L., and Morrison, K. 2007. *Research Methods in Education – 6th Edition*. London: Routledge.
- Colley, H., James, D., and Diment, K. 2007. Unbecoming teachers: towards a more dynamic notion of professional participation. *Journal of Education Policy*, 22, no. 2: 173-193.
- Cope, P., Bruce, A., McNally, J., and Wilson, G. 2003. Grading the practice of teaching: an unholy union of incompatibles. *Assessment & Evaluation in Higher Education*, 28, no. 6: 673 -684.
- Department for Education and Employment (DfEE). 1999. *The Learning and Skills Council Prospectus: Learning to Succeed*. London: HMSO.
- Department for Education and Employment (DfEE). 2001a. *The Further Education Teachers' Qualifications (England) Regulations 2001 - Statutory instrument 1209*. London: HMSO.
- Department for Education and Employment (DfEE). 2001b. *Raising Standards in Post-16 Learning: Self-Assessment and Development Plans*. London: HMSO.
- Department for Education and Skills (DfES). 2002. *Success for all: reforming further education and training*. London: DfES.
- Department for Education and Skills (DfES). 2003. *The future of initial teacher education for the learning and skills sector: an agenda for reform*. London: DfES.

- Department for Education and Skills (DfES). 2004. *Equipping our Teachers for the Future: Reforming Initial Teacher Training for the Learning and Skills Sector*. London: DfES.
- Department for Education and Skills (DfES). 2005. *Excellent teachers* Sherwood Park: DfES Publications.
- Department for Education and Skills (DfES). 2006. *Further education: raising skills, improving life chances*. CM 6768. London: The Stationery Office.
- Finlay, I., Spours, K., Steer, R., Coffield, F., Gregson, M., and Hodgson, A. 2007. 'The heart of what we do': policies on teaching, learning and assessment in the learning and skills sector', *Journal of Vocational Education & Training*, 59, 2: 137-153.
- Foucault, M. 1977. *Discipline and Punish – The Birth of the Prison*. Harmondsworth: Penguin.
- Foucault, M. 1980. *Power/Knowledge – Selected Interviews and Other Writings 1972-1977*. Brighton: The Harvester Press.
- Foucault, M. 2002. *Michel Foucault – Power – Essential Works of Foucault 1954-1984: Vol. 3.*, ed. James. D. Faubion. London: Penguin Books.
- Further Education National Training Organisation (FENTO). 1999. *National Standards for Teaching and Supporting Learning in Further Education in England and Wales*. London: FENTO.
- Gipps, C. 1994. *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer Press.
- Gleeson, D., and Gunter, H. 2001. The Performing School and the Modernisation of Teachers (pp. 139-158), in Gleeson, D. & Husbands, C. (eds), *The Performing School*. London: Routledge.

- Gleeson, D., Davies, J. and Wheeler, E. 2005. On the making and taking of professionalism in the further education workplace. *British Journal of Sociology of Education*, 26, no. 4: 445-460.
- Gleeson, D. and James, D. 2007. 'The paradox of professionalism in English Further Education: a TLC project perspective'. *Educational Review*, 59, no. 4: 451-467.
- Grubb, W. 2000. Opening Classrooms and Improving Teaching: Lessons from School Inspections in England. *Teachers College Record*, 102, 4: 696-723.
- Hammersley-Fletcher, L., and Orsmond, P. 2004. Evaluating our peers: is peer observation a meaningful process? *Studies in Higher Education*, 29, no. 4, 489-503.
- Hammersley-Fletcher, L., and Orsmond, P. 2005. Reflecting on reflective practices within peer observation. *Studies in Higher Education*, 30, no. 2: 213-224.
- Hatton, P. 2008. 'Observing teaching is central to improvement.' *The Talisman*, Issue 70. London: Ofsted Publications Centre.
- Huddleston, P., and Unwin, L. 2007. *Teaching and Learning in Further Education: Diversity and Change, 3rd Edition*. London: Routledge-Falmer.
- James, D., and Biesta, G. (eds) 2007. *Improving learning cultures in Further Education*. London: Routledge.
- Jeffrey, B., and Woods, P. 1998. *Testing Teachers: the effect of school inspections on primary teachers*. London: Falmer Press.
- Lather, P. 1991. *Getting Smart*. London: Routledge.
- Jephcote, M., and Salisbury, J. 2009. Further education teachers' accounts of their professional identities, *Teaching and Teacher Education*, 25, no.7: 966-972.
- Lifelong Learning UK (LLUK) 2006. *New overarching professional standards for teachers, tutors and trainers in the lifelong learning sector*. London: LLUK.

- Lifelong Learning UK (LLUK) 2007. *Guidance for awarding institutions on teacher roles and initial teaching qualifications*. London: LLUK.
- Lucas, N. 2004. The 'FENTO Fandango': national standards, compulsory teaching qualifications and the growing regulation of FE college teachers. *Journal of Further and Higher Education*, 28, no. 1: 35-51.
- Lyotard, J-F. 1984. *The Postmodern Condition: a report on knowledge*, vol. 10. Manchester: Manchester University Press.
- Marriott, G. 2001. *Observing Teachers at Work*. Oxford: Heinemann.
- Montgomery, D. 2002. *Helping Teachers Develop through Classroom Observation*, 2nd Edition. London: David Fulton Publishers.
- Ofsted. 2003. *The initial training of further education teachers: a survey*. London: Ofsted Publications Centre.
- Ofsted. 2004a. *Why Colleges Fail*. London: Ofsted Publications Centre.
- Ofsted. 2004b. *Why Colleges Succeed*. London: Ofsted Publications Centre.
- Ofsted. 2006. *The initial training of further education teachers: findings from 2004/05 inspections of courses leading to national award body qualifications*. London: Ofsted Publications Centre.
- Ofsted. 2008a. *How Colleges Improve*. London: Ofsted Publications Centre.
- Ofsted. 2008b. *A Handbook for Inspecting Colleges*. London: Ofsted Publications Centre.
- O'Leary, M. 2006. Can inspectors really improve the quality of teaching in the PCE sector? Classroom observations under the microscope. *Research in post-compulsory education*, 11, no. 2: 191-198.
- Ollin, R. 2009. *The grading of teaching observations: implications for teacher educators in Higher Education partnerships*. Huddersfield. Huddersfield Consortium.

- Orr, K. 2008. 'Room for improvement? The impact of compulsory professional development for teachers in England's further education sector', *Professional Development in Education*, 34: 1: 97-108.
- Peake, G. 2006. *Observation of the Practice of Teaching*. Huddersfield. Huddersfield Consortium.
- Peel, D. 2005. Peer Observation as a transformatory tool? *Teaching in Higher Education*, 10, no. 4: 489-504.
- Perryman, J. 2006. Panoptic performativity and inspection regimes: disciplinary mechanisms and life under special measures. *Journal of Education Policy*. 21, no. 2: 147-161.
- Perryman, J. 2007. Inspection and emotion. *Cambridge Journal of Education*. 37, no. 2: 173-190.
- Perryman, J. 2009. 'Inspection and the fabrication of professional and performative processes'. *Journal of Education Policy*, 24, no. 5: 611-631.
- Quality Improvement Agency (QIA) 2007. *Pursuing excellence: the National Improvement Strategy for the further education system*. Coventry: QIA.
- Randle, K., and Brady, M. 1997. Managerialism and professionalism in the 'cinderella service'. *Journal of Vocational Education and Training*. 49, no.1: 121-139.
- Robson, J. 1998. A Profession in Crisis: status, culture and identity in the further education college. *Journal of Vocational Education and Training*. 50, no. 4: 585-607.
- Shortland, S. 2004. Peer Observation: a tool for staff development or compliance? *Journal of Further and Higher Education*, 28, no. 2: 219-228.
- University and College Union (UCU) 2009. *Lesson Observation UCU Guidelines*. [Accessed 22/02/2010] Available online at: <http://www.ucu.org.uk/index.cfm?articleid=2969>.

Whitehead, S. 2005. 'Performativity Culture and the FE Professional'

Management in Education. 19, no. 3: 16-18.

Wragg, E.C. 1999. *An Introduction to Classroom Observation*, 2nd Edition. London: Routledge.

Wragg, E. C., Wikeley, F. J., Wragg, C. M., and Haynes, G. S. 1996. *Teacher Appraisal Observed*. London: Routledge.

¹ Ball acknowledges the use of the term 'performativity' as originating in the work of Lyotard (1984), who used it to describe the obsession that postmodern society has with efficiency and effectiveness and how these concepts are measured through an industrial model of input equals output. Ball (2003, 215) describes performativity as a culture that 'requires individual practitioners to organise themselves as a response to targets, indicators and evaluations, to set aside personal beliefs and commitments and live an existence of calculation'.

² It is generally accepted that the Green Paper *The Learning Age* (1998) was the first publication to encapsulate New Labour's vision of lifelong learning and the White Paper *Learning to Succeed* (1999) acknowledged this in its introductory statement.