



**Developing a National Framework for the
Effective Use of Lesson Observation in Further
Education**

Project Report for UCU

By Dr Matt O'Leary

November 2013

| Contents | Pages |
|---|--------------|
| <i>Figures and tables</i> | 3-4 |
| <i>Acknowledgements</i> | 5 |
| 1. Executive summary | 6-7 |
| 2. Introduction | 8-10 |
| 3. Literature review of lesson observation: understanding its role in schools, colleges and universities in England | 11-29 |
| 4. Research methodology | 30-36 |
| 5. Findings and discussion | 37-85 |
| 6. Conclusions and recommendations | 86-91 |
| 7. References | 92-97 |
| 8. Appendices | 98-104 |

Figures and Tables

Pages

Figures

| | | |
|----|---|----|
| 1 | Gender | 37 |
| 2 | Employment status | 38 |
| 3 | Employment capacity | 38 |
| 4 | Years of teaching experience | 39 |
| 5 | Observer, observee or both? | 40 |
| 6 | Contexts of lesson observation | 40 |
| 7 | Models of lesson observation | 41 |
| 8 | Use of unannounced lesson observations | 42 |
| 9 | Graded lesson observations | 44 |
| 10 | Ungraded lesson observations | 48 |
| 11 | Lesson observation feedback | 51 |
| 12 | Unannounced lesson observations | 53 |
| 13 | Key thematic categories of the project's findings | 56 |
| 14 | Multi-dimensional model of teacher appraisal | 69 |

Tables

| | | |
|----|--|----|
| 1 | Sample of qualitative comments on graded observations | 47 |
| 2 | Sample of qualitative comments on ungraded observations | 50 |
| 3 | Sample of qualitative comments on lesson observation feedback | 52 |
| 4 | Sample of qualitative comments on unannounced observations | 54 |
| 5 | Summary of key themes and issues | 57 |
| 6 | Sample of qualitative comments on observation as a punitive tool | 58 |
| 7 | Sample of qualitative comments on observation as a 'box-ticking' exercise | 60 |
| 8 | Sample of qualitative comments on stress and anxiety associated with current observation regimes | 62 |
| 9 | Sample of qualitative comments on emphasis on judging and measuring rather than supporting improvements in teaching and learning | 63 |
| 10 | Sample of qualitative comments on preparation time for observations | 65 |
| 11 | Sample of qualitative comments on need for trust in the professionalism of teaching staff | 66 |
| 12 | Sample of qualitative comments on unfair reliance on observations to judge professional competence | 68 |
| 13 | Sample of qualitative comments on concerns about the validity and reliability of judgement | 70 |
| 14 | Sample of qualitative comments on inauthenticity of observations | 72 |

| | | |
|----|---|----|
| 15 | Sample of qualitative comments on importance of subject specialist & fully trained observers | 74 |
| 16 | Sample of qualitative comments on professional credibility of observers | 76 |
| 17 | Sample of qualitative comments on importance of observation as a tool for professional learning | 77 |
| 18 | Sample of qualitative comments on value of peer observation | 79 |

Acknowledgements

I would like to take this opportunity to thank everyone who contributed to this research project. From those UCU members who completed the online questionnaire, to those who participated in the interviews and focus groups. Without the cooperation or generosity of these people in willingly giving up their time to take part, I would not have been able to gather such a rich collection of data, nor indeed compile a report that brings together such a comprehensive insight into the important topic of lesson observation in the FE sector. I would also like to thank Terry Pearson, an independent education consultant, for his contribution as a 'critical friend' in on-going discussions regarding the project, along with his comments on an earlier draft of the report, all of which have proven very valuable.

Dr Matt O'Leary

November 2013

1. Executive summary

This report explores one of the most widely debated and hotly contested initiatives to affect teaching staff in the FE sector in recent times, that of **lesson observation**. The report captures the views of thousands of UCU members working in a wide range of contexts and institutions and as such represents the largest and most extensive account of the topic to date.

Even before beginning to discuss the project's findings, what emerges very clearly from this study is that the use of lesson observation and its impact on the professional lives of the FE workforce is 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. 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.

Naturally the data presented in this report encompasses a breadth of views, as one might expect from the size and diverse representation of the sample (see section 3), though there were numerous aspects of the topic on which there was an overriding consensus among participants. One of the main findings to emerge from this study was the widespread discontent felt amongst UCU members towards the use of lesson observation as a form of teacher assessment. This dissatisfaction was particularly targeted at graded models of observation, which have become the norm in FE over the last two decades (e.g. O'Leary 2013a). These were repeatedly criticised by a significant majority of participants for being little more than a 'box-ticking' exercise and, in some instances, a 'disciplinary stick' with which 'to beat staff'. In relation to this, graded observations were also identified by many respondents as being a major cause of increased levels of stress and anxiety amongst teaching staff.

Another compelling finding to emerge from the study's data was the increasing appetite for change to how observation was used in many institutions across the sector. 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/approaches to observation were deemed to be more beneficial than others, particularly peer-

based models with a focus on enhancing professional learning and development. Furthermore, many participants expressed the need to explore alternative approaches and to move away from current normalised models of graded observations driven by performance management agendas. What these alternative approaches might look like or consist of differed from one institution to another, though there are common features that seemed to apply regardless of contextual variables and these are explored in depth in sections five and six of this report.

There is little doubt that lesson observation has become increasingly associated with the monitoring of standards and teacher accountability in the sector over the last two decades. To say that respondents were fully supportive of the view that poor teaching is not something that should be tolerated and that every effort should be made to eradicate it wherever it occurs might seem to be stating the obvious. How this should be done, however, was a matter of some debate. A recurring theme in the study's data highlighted the shortcomings surrounding the current reliance of many institutions on graded lesson observation as the main – and sometimes sole – means of identifying and eradicating 'poor teaching', though opinions differed according to the employment role of some participants. Whilst teaching staff recorded high levels of disagreement regarding the effectiveness of graded observations, these views were not necessarily shared by senior managers, although it has to be acknowledged that the latter represented a very small percentage (n = 20) of the overall sample.

Many of the shortcomings expressed concerning graded lesson observations centred on the topic of assessment, in particular the key principles of validity and reliability of observation as a method of assessment. In other words, a viewpoint expressed by the majority of the study's participants was that it was neither valid nor reliable to make a conclusive judgement about someone's professional competence based on 'snapshots' or isolated, episodic performances. It was felt that any overall judgement needed to be inclusive of other key performance indicators (KPIs) such as student achievement rates, student evaluations, self-evaluations, peer reviews etc.

In short, this report raises serious questions about the fitness for purpose of prevailing observation assessment systems in FE and the extent to which these systems are able to achieve their purported goals. The overriding message from practitioners was that current, normalised models of graded lesson observations have minimal, if any, positive impact on raising the quality of teaching and learning across the sector. In many instances, they appear to have become a perfunctory mechanism with observers as well as observees questioning their effectiveness as a method of assessment. The views of practitioners working at different levels within the sector all point to one pressing outcome and that is the need for a change to current practices. The findings from this report can thus be seen as a mandate for change, along with providing an accompanying set of concrete recommendations to effect such change.

2. Introduction

In recent years lesson observation has become one of the most important methods of judging the performance of teaching staff in the FE sector both internally for colleges and externally for agencies such as Ofsted. The policy focus of the current coalition government suggests that this reliance on lesson observation as a key tool for assessing the quality of teaching and learning looks set to become even greater as teachers enter a new era of heightened scrutiny and performativity (e.g. DfE 2010; Ofsted 2012). Given its importance, it is somewhat surprising then that very little is known about the effects of observation on individual FE tutors or indeed its role in improving the quality of teaching and learning in general. This project seeks to make significant inroads in addressing that gap in knowledge.

Recent research carried out in FE has argued that the use of lesson observation has been predominantly shaped by a performance management agenda (e.g. Boocock 2013; O'Leary 2012a, 2012b, 2013a, 2013b). This has been reflected in the contexts and purposes for which observation has been used for assessing the performance of experienced, in-service teachers but also in the training and assessment of novice, pre-service teachers. This performance driven focus has culminated in a prescribed and codified model of what it means to be an effective teaching professional in some circles, with limited opportunities for the use of observation to stimulate collaborative discussion about the process of teaching and learning. One of the principal aims of this research project was to construct an informed understanding of how observation operates across the sector, what the benefits and shortcomings are of current approaches and how observation might be harnessed most effectively to create collaborative networks between teachers whose collective goal is to enrich and support the continuous improvement of teaching and learning in classrooms.

Furthermore, recent reports into how lesson observation has 'become an increasingly common flash point in colleges, triggering local negotiations, and in some places industrial disputes' (UCU 2009: 1) have raised questions about the extent to which policy aims are being achieved and highlight the timeliness of this study. Over the last few years, UCU members have been involved in boycotting lesson observations in numerous colleges (e.g. UCU FE News, May 2012). There has also been an increase in some colleges linking lesson observation outcomes with formal disciplinary procedures, heavily influenced by Ofsted's recent policy shift to re-classify grade 3 from 'satisfactory' to 'requires improvement' (Ofsted 2012: 6). Yet, at the same time, there is also evidence of a move away from graded lesson observations in some colleges.

In December 2012, I was invited to UCU's head office to meet and talk with the London FE regional committee and the national FE Executive committee about my previous research into lesson observation, along with the specific focus of this project. On both occasions, I was struck by the intensity of feelings and concerns

expressed by all those present at the meetings with regard to members' experiences of observation. Looking back on those meetings now, it is fair to say that many of the views and anecdotal evidence presented by college and regional representatives at that time were subsequently to be echoed repeatedly in the voices of the participants of this study.

It is anticipated that this study's findings will have immediate relevance not just to UCU members working in the FE sector but to all staff involved in the process of lesson observation i.e. observers, observees and Senior Management Teams (SMTs). This does not apply solely to FE colleges, but to those involved in adult learning and work-based learning settings. In addition, the findings will be of interest to policy makers at national level, professional bodies and agencies with an involvement in or representation of the FE workforce. Finally, it is hoped the study will also contribute to wider discussions in the field of professional learning and development for staff in FE, but will also be relevant to the schools' and HE sectors.

Overview of the research project and its focus

This project is concerned with investigating the role of lesson observation in the Further Education (FE) sector. It has two interrelated aims, the first of which is to explore and evaluate current models of lesson observation in use and their perceived purposes and effects on FE tutors' professional practice and development. The second seeks to examine and identify those aspects of lesson observation practice that create optimum opportunities for expansive professional learning and development among tutors as well as those that restrict them. It is anticipated that the research findings will generate recommendations that in turn will help to form the basis of a national framework for colleges on how to make best use of lesson observation. This framework will be built on the core principle that observation is at its most effective as a form of intervention when it prioritises the growth of tutors' professional learning and skills and empowers them to become active agents in the construction of their own professional identity, learning and development. The key objectives of this research project can thus be summarised as follows:

- To conduct a national inquiry that examines the current use and impact of lesson observation schemes on UCU members working in the FE sector
- To explore and identify observation models that are considered to reflect 'good' or 'effective' practice
- To identify observation models that are considered professionally enriching and that prioritise the professional learning and development of FE tutors
- To establish a national framework and recognised code of practice that sets the standard for all FE colleges on the most effective use(s) of lesson observation

The project is underpinned by the following interconnected research questions:

- What models of lesson observation are currently in use in FE colleges?
- What are the purported aims for their use and how well do the outcomes match these aims?
- What is the impact of current models of lesson observation on improving standards in teaching and learning?
- How can lesson observation be used most effectively to support FE lecturers' professional learning and development?

3. Literature review of lesson observation: understanding its role in schools, colleges and universities in England¹

Introduction

Lesson observation has a longstanding tradition in the assessment and professional development of new and experienced teachers in England. Over the last two decades it has progressively emerged as an important tool for measuring and improving teacher performance in schools and colleges. This section reviews relevant literature and studies across the three education sectors in England (i.e. schools, Further Education (FE) and Higher Education (HE)) in order to compare and contrast the role of observation. In doing so it discusses the key themes and issues surrounding its use in each sector and identifies common and contrasting patterns. It argues that in schools and FE, observation has become increasingly associated with performance management systems; a dominant yet contested model has emerged that relies on a simplified rating scale to grade professional competence and performance, although the recent introduction of 'lesson study' in schools appears to offer an alternative to such practice. In HE, however, there is limited evidence of observation being linked to the summative assessment of staff, with preferred models being peer-directed and less prescribed, allowing lecturers greater autonomy and control over its use and the opportunity to explore its potential as a means of stimulating critical reflection and professional dialogue about practice among peers. The section concludes with a synopsis of the recurring themes and issues to emerge across all three sectors with a view to establishing their significance to all those involved in the process of lesson observation. These themes and issues will also serve as useful reference points for discussion in later sections, particularly when analysing the study's data.

The schools' sector experience

Much of the existing literature on lesson observation is located in the schools' sector with a particular focus on the *practice* of observation and takes the form of textbooks rather than research-based texts. Besides there being fewer studies in FE and HE, those that do exist have occurred mainly in the last decade. This in itself is significant as it highlights how observation has a longer history in the schools' sector (Grubb 2000; Wilcox and Gray 1996).

As discussed in the previous section, it was following the educational reforms of the 1980s and 1990s that observation materialised as well-established practice in schools. Although it had long been and continues to be a pivotal method of assessment in Initial Teacher Training (ITT) courses, its rise to prominence for qualified teachers was closely linked to wider political reforms at the time, which demanded increased public accountability and an educational reform agenda determined to impose greater control over what teachers did in the classroom (Lowe 2007). Amid concerns

¹ This section is based on a chapter from O'Leary (2013c). Included with permission.

about standards and the quality of teaching, observation emerged as a key method of collecting evidence on which to base subsequent systems of teacher appraisal. Marriott (2001: 3) has argued that:

The long term success of the school depends to a very great extent on the quality of teaching ... It is hard to see how the headteacher and other managers in schools can be fully aware of the quality of work unless they are gaining first-hand information by systematically observing in classes.

In a similar vein, back in the 1990s Wilcox and Gray (1996) referred to the 'dominance' of observation as the main method of collecting data about what went on in classrooms. This dominance was crystallised by the introduction of a cycle of inspection by Ofsted and formal appraisal for teachers as discussed previously.

Wragg's seminal work, *An Introduction to Classroom Observation* (1999), is one of the most widely cited textbooks on the subject to date. It is located in the context of English primary schools and covers a wide range of themes related to observation pedagogy and theory, although the emphasis is largely on the practical application of observation as a pedagogic tool. Given the breadth of coverage in Wragg's work it is helpful to use it as a starting point from which to explore some of the key issues and themes in the field and to relate these to other relevant studies.

At the beginning of his book Wragg raises the issue of the reliability of observation as a form of assessment. He remarks that 'we often "observe" what we want to see' (1999: vii). His comment draws attention to the subjectivity of observation and how events are 'inevitably filtered through the interpretive lens of the observer' (Foster 1996: 14). The subjectivity of observers' interpretations is a common theme in the literature, particularly when discussing notions of *good practice* (e.g. Fawcett 1996; Montgomery 2002; Tilstone 1998). Wragg maintains that 'mostly when we talk about a "good" teacher, an "effective" strategy or a "bad" lesson, we are referring to our own subjective perception (*op. cit.*, p. 60). He exemplifies his argument by recalling a session in which thirty five highly experienced teacher educators were shown a videotape of a student teacher's lesson and were asked to grade it on a scale of A-E (A at the top end of the scale and E at the bottom). Their grades varied from a D at the lowest end to a B+ at the top end of the scale. Such differing judgements illustrate the issue of observer subjectivity and reinforce the unreliability of observation as a sole method of assessment, especially when a grading scale is used to measure performance.

Wragg is critical of hierarchical grading systems as he claims that 'the nature of the levels can still be vague and diffuse, using words like "adequate" or "considerable" that are open to widely differing interpretations (*op. cit.*, p. 103). It cannot be assumed that there is a shared understanding among observers or observees as to

the meaning and interpretation of value-laden terms such as 'good' and 'outstanding', as used by Ofsted (See Brooks 2009; Wolf 1995). Wragg (1999) argues that these terms, together with the assessment criteria that underpin them, need to be carefully defined when used and attempts made to establish a collective understanding. But even when such attempts are made e.g. standardisation exercises for observers in observation systems and associated assessment criteria, the limitations of what they can achieve need to be acknowledged (Montgomery 2002).

In other words, whilst they might be useful in raising collective awareness among observers, it is unrealistic to expect the assessment criteria to be uniformly and consistently applied. Besides, extant research suggests that experienced assessors are likely to judge intuitively, even ignoring published criteria. Though it has to be said that this is not a phenomenon specific to observation as a method of assessment, but reinforces more widely held beliefs among key researchers in the field that 'assessment is not an exact science and we must stop presenting it as such' (Gipps 1994: 167). In her book on competence-based assessment, Wolf (1995) highlighted how 'assessor judgement' was underestimated as a significant variable in the assessment process, based largely on the misguided notion that the more detailed and specific the assessment criteria, the more objective it would make assessor judgement. Wolf goes on to say that this fails to understand the complexity of the assessment process and address the aspect of judgement.

In their two-year monitoring study of teacher appraisal in English primary and secondary schools, Wragg et al (1996) highlighted what they described as a 'snapshot' approach to lesson observation (i.e. one-off observations) as one of the main obstacles to identifying incompetent teachers. The reason for this was because such teachers could deliver the 'rehearsed' lesson as a one-off performance, hence avoiding detection. Marriott (2001: 8) has also highlighted the limitations of a snapshot approach as 'the impact of teaching on learning, and therefore progress, is harder to evaluate in the context of one lesson.' As a means of mitigating such limitations, Wragg et al (1996) advocated the need for a series of observations to be carried out as part of a longitudinal approach to construct a more realistic picture of a teacher's classroom competence. However, in their research into teacher effectiveness, Campbell et al (2004: 133) maintain that 'even successive observations of a teacher will only ever supply a collection of snapshots rather than a full picture of teacher behaviour over the year'. While they acknowledge the importance of observation as a source of evidence for systems of teacher appraisal, they also remark that as a method of data collection 'it is often used with little regard for, or knowledge of, its characteristics' (p.133). What they mean by this comment is that despite its widespread use as a means of gathering data, there is a lack of rigour in its application and insufficient awareness on the part of those carrying out observation of its limitations as a method.

Wragg (1999: 3) succinctly summarises some of the paradoxes involved with observation when comparing the ways in which teachers respond to the different contexts in which it occurs and its application as a multipurpose tool in the following comment:

Skilfully handled, classroom observation can benefit both the observer and the person being observed, serving to inform and enhance the professional skills of both people. Badly handled, however, it becomes counter-productive, at its worst arousing hostility, resistance and suspicion.

The rules of observer-observee engagement are likely to differ according to who is observing whom, in what context and for what purpose. Underpinning Wragg's comment and the observer-observee relationship are the notions of power and authority. As Wragg comments, 'the actual or perceived power relationship between observer and observed is not just a sociological concept, but rather a reality that needs to be recognised' (*op. cit.*, p.62). For example, if a head of department is observing a newly qualified teacher (NQT), to what extent does the teacher feel able to challenge their assessment? Are efforts made to ensure that the observee's voice is heard? How many observation schemes actually choose to tackle this issue? What efforts are made, if any, to address the distribution of power? These are questions that will be explored in later sections but for now it is worth acknowledging that what links them is the degree of ownership and autonomy afforded to the observee.

Ownership and autonomy are identified as key features of successful observation schemes in schools, which are characterised by a move away from authoritarian models where observation is something that is 'done to teachers' to a more egalitarian approach in which ownership of the process is devolved to teachers (e.g. Metcalfe 1999; Tilstone 1998). Tilstone (1998: 59) advocates 'partnership observation', a term which she uses to express a more collaborative, democratic relationship between observer and observee. She argues that 'such partnerships will only work if the [observer] is not regarded as an authoritarian figure and is able to take on the role of facilitator with the teacher in control of direction of the observation and consequent actions' (p. 60).

Metcalfe (1999: 454) reflects on his experience of the use of observation from the perspective of both a researcher and an Ofsted inspector:

What is becoming clear in schools is that classroom observation, as an aspect of monitoring and evaluation, is felt to be most acceptable when it is part of a broader approach ... in which teachers work collaboratively as opposed to a 'bolted on' approach, which is felt to be connected with 'checking up', accountability and control.

These views can be seen to embody a wider, egalitarian philosophy of professional learning and development, sometimes referred to as 'democratic professionalism' (see, for example, Sachs 2001; Whitty 2000). A similar approach to observation is supported by two of the main professional associations for school teachers, the National Union of Teachers (NUT) and the Association of Teachers and Lecturers (ATL) (ATL 2008; NUT 2006). It is their shared belief that observation constitutes an important element of a teacher's professional development and as such should be 'neither a burden for the teacher concerned nor an opportunity to "police" a teacher's performance' (NUT 2006: 5), but 'should be conducted in a manner that equates to a professional dialogue' (ATL 2008: 1).

The advantages of observation programmes that prioritise development over surveillance are well documented. As Wragg argues (1999: 17), 'good classroom observation can lie at the heart of both understanding professional practice and improving its quality'. When it is used insightfully observation can have a profound impact, which 'can lead to a more open climate, greater trust between colleagues, and the development of strong professional relationships' (Marriott 2001: 3). One of the biggest obstacles to the creation of such a climate would appear to be the issue of grading.

Historically, graded observation has been a contentious issue in schools and provoked a resolute response on the part of the two largest professional associations for schoolteachers, the NUT and the National Association of Schoolmasters Union of Women Teachers (NASUWT) to campaign against its use (e.g. NUT 2007). Both unions believe that grading encourages school management to view observation as a surveillance mechanism with which to monitor the quality of teachers' work, instead of seeing it as a valuable means of stimulating professional dialogue. Marriott (2001) maintains that the grade can take on such importance that it threatens to undermine the value of the dialogue and feedback between the observer and the observee. Both parties can 'become hung up on what the grade means rather than how to improve the teaching' (2001: 46). Such is the anxiety surrounding grading that the 'teacher may become over-concerned about whether he or she has "passed"' (*Ibid.*).

One of the most recent developments in the field of lesson observation in schools has been the use of 'lesson study' as a model for improving teaching and learning. In drawing on Stigler and Hiebert (1999), Lieberman (2009: 83) traces the origins of lesson study to Japan, where it has a long and well-documented history and has been used as the most common 'form of teacher professional development in the improvement of mathematics and science education'. Unlike conventional models of observation that tend to be based on an atomistic approach, relying on evidence collected during a single, isolated observation on which to base judgements and formulate follow-up improvement action plans, lesson study 'challenges the status quo of teachers and their classrooms as islands – relatively unaware of events on

other islands – with students floating in between’ (Wang-Iverson 2002: 1). Far from being seen as a corrective mechanism to improve the practice of individual teachers, the planned impact of lesson study is meant to be collaborative and fully inclusive of all an institution’s practitioners. The emphasis is placed on the observation of an entire curriculum unit rather than an isolated lesson, and how those who teach that unit can enhance greater student understanding and achievement. In short, lesson study is broadly based on an action-research approach to studying what goes on in classrooms where teachers work collaboratively as active researchers. Furthermore, one of its unique characteristics is how it seeks to involve the learners in the discussion and analysis of the observed lessons. According to Lieberman (2009), lesson study puts student and teacher learning at the centre of the observation process rather than teacher evaluation.

As part of a pilot project in England co-funded by the Teaching and Learning Research Programme (TLRP), the National College for School Leadership and the Centre for British Teachers (CfBT), teachers from twenty schools (both primary and secondary) across eight local authorities were involved in the trialling of lesson study with the aim of answering the project’s key question: ‘Would Lesson Study work in the UK and if so would it do so in a way which would add value to the range of professional development approaches already in use?’ (Dudley 2007). The research was conducted in two phases: the first phase from 2003-06 and the second from 2007-2010.

One of the findings to emerge from the first phase of the project was that lesson study was found to be a ‘popular, powerful and replicable process for innovating, developing and transferring pedagogic practices’ (*Ibid.*). According to Dudley ‘it was popular with both experienced and less experienced teachers alike’ and had a ‘demonstrable impact on the quality of teaching and on pupil progress and attainment’ (*Ibid.*). It remains difficult to analyse the validity and reliability of these claims given that these were emergent findings from an on-going research project, the methodology of which was not made transparent.

The second phase of the research aimed to explore the ‘critical features of teacher learning in lesson study’ and what distinguished it as a form of classroom inquiry. Some of the key findings indicated that lesson study fostered joint risk taking among teachers, enabled teachers to develop evidence-based practice to inform their professional learning and empowered them to take control of their professional development through their own classrooms (Dudley 2008).

Dudley (2007, 2008) claims that the use of lesson study has subsequently become more widespread across schools in England in recent years, with particular emphasis on mathematics and literacy teaching, though this claim is not supported by any quantitative evidence. Notwithstanding the lack of data to support his claim,

there is evidence of a growth in the use of lesson study as a form of intervention in teacher development in other countries. Lieberman (2009), for example, has reported on the popularity of lesson study increasing in the USA over the last decade. In her research Lieberman (2009) found that lesson study encouraged greater openness among staff, which helped to expose vulnerability as an issue that affects both experienced and novice teachers. Reinforcing some of Dudley's (2007, 2008) findings, Lieberman argues that lesson study has helped to foster a collegial approach to teacher development through peer observation and thus prompted teachers to take more risks in their teaching.

The FE sector experience

There has been comparatively little research regarding the role of observation in FE to date. One of the first studies was Cockburn's (2005). His qualitative research consisted of interviews and focus groups with observers and observees as well as documentary analysis of the chosen institution's observation policy and feedback reports. The research was based in one college and although the number of participants is not specified, one gets the impression that it was a relatively small cohort. The study's aim was to report on the perceptions and attitudes of staff to the use of observation.

Cockburn (2005: 376) provides what he refers to as a 'typology of resistance' of those who expressed negative views about observation. Some of the issues that he lists resonate with other work (e.g. O'Leary 2006, 2011, 2012b, 2013a; Wragg et al 1996). For example, he suggests that there is evidence of 'artificiality' in lessons as a result of being observed, which leads some tutors to adopt an 'orthodox style of delivering lessons' (p. 380) on the basis that there is a 'formula' for effective teaching. Such changes in behaviour are considered symptomatic of the methodological problem of *reactivity* or what is commonly referred to as the 'Hawthorne' effect (see section 5 for further discussion). That is the extent to which the behaviour of the observed environment is influenced by the observer's presence and/or the observee alters their behaviour due to an awareness of being observed. Nevertheless, in a much larger and more recent research study, O'Leary (2011) found that this 'artificiality' and 'orthodox' teaching style were largely due to the high stakes nature of the assessment and what he describes as the 'normalisation of practice', which resulted in many teachers 'playing the game' in order to ensure a successful outcome during graded observations (see below for further discussion).

In Cockburn's study the credibility of the observer also emerged as a contested issue amongst observees, specifically relating to their experience and suitability to perform the role. O'Leary's (2011) research unearthed similar findings. Many of his participants' comments about credibility tended to converge around whether observers were still current practitioners. A popular complaint on the part of observees was that invariably observers were middle managers who had not taught

for some time and thus had 'lost touch with the classroom'. Equally a significant percentage of observers stated that they thought it was essential to still be teaching for them to remain credible in the eyes of their colleagues.

Related to the question of credibility in O'Leary's (2011) research was whether the observer had knowledge and/or experience of teaching the subject area of the observee. There were numerous instances of teachers having been observed by colleagues from curriculum areas that appeared to have little in common with their own. Gavin, for example, was an agricultural studies tutor who was observed teaching tractor driving to a group of teenagers by an IT specialist. Debbie, a special needs teacher, was observed teaching a dance and movement class to a group of young adults with severe learning difficulties by a manager from engineering whose students were mainly HE level.

Like Wragg (1999) above, Cockburn makes reference to the power relationship between observer and observee, arguing that 'the observer is commonly perceived as possessing greater power', which 'is legitimised by organisational arrangements' (2005: 384). This is a phenomenon that has been commented on by other writers in the field particularly regarding how it threatens to undermine the developmental potential of observation. This is accentuated when it is used to fulfil the dual purpose of performance management requirements and the developmental needs/goals of teachers (e.g. Ewens and Orr 2002; O'Leary 2006, 2011).

O'Leary (2006) has argued that an assessment approach to lesson observation, like that employed by Ofsted and internal QA schemes, is ineffective in terms of its impact on improving the standards and quality of classroom teaching and learning. The primary purpose of such approaches 'is not to inform and improve current practice but simply to make a judgement about the quality of teaching and learning being observed' (p. 192). One of the main problems with such approaches according to O'Leary is that they 'place an inequitable proportion of control and decision-making at the behest of the observer, thus limiting the role of the person being observed (the *observee*) to that of a passive recipient rather than an active participant' (*Ibid.*).

O'Leary's (2011) recent research adopted a mixed-methods approach and was carried out in a sample of 10 colleges situated across the West Midlands with a total of 500 participants, consisting of 50 from each college. The sample comprised teaching staff, middle and senior managers. One of the key findings to emerge from the research was how graded observation had become normalised as a performative tool of managerialist systems fixated with measuring teacher performance rather than actually improving it. The vast majority of colleges involved in the study adopted what O'Leary refers to as a 'restrictive approach' to the use of observation, typified by their reliance on the use of the Ofsted 4-point graded scale to measure

performance, prioritising the needs of performance management systems over those of their staff. Yet, in contrast, in those colleges where there was evidence of an 'expansive approach', grading was seen as less important, as the professional development needs of staff underpinned the way in which observation was used.

O'Leary's (2011) research data also uncovered repeated examples of teachers being encouraged to demonstrate normalised models of 'effective practice' based on prescribed notions of 'excellent' or 'good' teaching, often cascaded from senior management, who were understandably keen to promote 'best practice' given the high-stakes nature of such observations particularly during inspections. One of the repercussions of this pressure on teachers to perform was how it encouraged what Ball (2003) refers to as 'inauthenticity' in teacher behaviour and classroom performance during graded observation. This was typically manifested in the delivery of the 'rehearsed' or 'showcase lesson', where the teacher concerned 'played the game' in order to succeed. For example, Terry, an engineering teacher with over twenty five years' experience, provided a candid and detailed account of how he followed a 'checklist' in the planning and delivery of a recent observation to achieve in order to ensure that he achieved 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 hand-outs 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 observation and the need to 'play the game'. 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, often by SMTs and/or external consultants, and subsequently apply them to their own teaching when being observed. This resulted in such practice becoming normalised and adopted as the *default* model for all those striving to achieve a high grade, which itself raises questions concerning the validity and reliability of graded observation as a means of assessing classroom performance, two important factors that are discussed in the project's findings. Such formulaic approaches to achieving

'outstanding' are reinforced by a plethora of publications and short courses currently available across the sector informing staff how to attain a 'grade 1'.

O'Leary's (2011) research drew on and was informed by the work of Foucault (1977, 1980, 2002) as its theoretical backbone, along with concepts relating to theories of new managerialism and performativity. Foucault's work, in particular, provided a useful framework for analysing the phenomenon of lesson observation as some of the key concepts he explored provided a lens through which to examine relationships of teacher agency and structure, as well as a language with which to describe and discuss the phenomenon of observation, for example, the concept of 'normalisation'.

Normalisation is a Foucauldian term that can be defined as the adjustment of behaviour to fall into line with prescribed standards. Perryman (2009: 614) states that:

Normalization is a powerful mechanism of power which is achieved through the hegemonic internalisation of discourses of control. In general, this means that those who are subjects of power internalise expected behaviours and learn these behaviours through acceptance of a discourse.

In the case of O'Leary's (2011, 2013a) research into the use and impact of models of graded observation, normalisation can be seen as a means of conceptualising the process by which teachers operate within the accepted norms of "good practice", a concept largely determined by agencies such as Ofsted. In her research, Perryman (2006: 150) argued that it is the discourse of school effectiveness research that has been appropriated by Ofsted that forms the dominant discourse in the context of inspections, which 'uses performativity and normalization as its mechanisms'. Perryman identifies this as an example of the use of knowledge to convey power.

Foucault (1977: 184) asserted that 'the power of normalization imposes homogeneity; but it individualises by making it possible to measure gaps, to determine levels, to fix specialities and to render the differences useful by fitting them one to another.' The 'homogeneity' that Foucault refers to is imposed by the requirement for all teachers to demonstrate standardised notions of good practice during graded observations. Those that are able to manifest such normalised behaviour become members of 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.*). As discussed further below, graded observation epitomises Foucault's *examination*, where a teacher's performance is categorised and differentiated by the observer according to Ofsted's 4-point scale.

Hardman's research (2007) adopted a case study approach in which the practice of observation was explored in three FE colleges and three HEIs. In seeking to identify differences between the two sectors, she reported that observation was much more heavily associated with Quality Assurance (QA) and performance management systems in FE colleges where there was an emphasis on grading staff. This contrasted with HE where there was a tendency for it to occur in developmental contexts, reflected in the prevalence of informal and ungraded peer observation as the most common model of practice and discussed in the following section. This viewpoint is echoed by Armitage et al (2003: 50) who refer to observation as the 'mainstay of the [FE] institution's quality assurance process', although they are also keen to stress its value as 'the basis of some of the most useful professional reflection you can undertake in order to improve performance' (p. 47).

The three FE colleges in Hardman's study all had observation policies and procedures that sought to combine the purposes of QA requirements for Ofsted, together with internal staff development agendas. Observation schemes are certainly time intensive and expensive for colleges. With limited budgets it is hardly surprising that colleges should attempt to dovetail two different purposes into one scheme. However, as Hardman suggests, the effectiveness of such a strategy is questionable. QA requirements appear to take precedence over the developmental needs of teachers. Furthermore, the use of observation for QA purposes is not without its controversies as has been discussed elsewhere

O'Leary's (2011) research revealed that the prioritisation of the performance management agenda over the developmental led to the nullification of observation as a tool for CPD in many institutions. As a result, teachers have come to experience a growing sense of disempowerment, increased levels of anxiety and general discontent in relation to its use. A recurring theme from O'Leary's data was the perceived lack of benefit of graded observation to teachers. Some said that the college management was the only beneficiary as it provided them with the necessary data to compare levels of performance to national benchmarks. Others referred to it as a 'tick-box' exercise that was more concerned with satisfying Ofsted than their development needs. Equally senior managers were sceptical of the use of such data as they saw it simply as part of the 'evidence trail' required for Ofsted.

Postlethwaite (2007) states that 'observing classes as part of quality assurance procedures has become a contentious matter in many FE colleges' (cited in James and Biesta 2007: 168). This is a viewpoint shared by the main professional association for the sector, the University and College Union (UCU) who report that it has 'become an increasingly common flash point in colleges, triggering local negotiations, and in some places industrial disputes' (UCU 2009: 1). As a result, the union have 'call[ed] for a code of practice over how such work is carried out' (Lee 2007: 1) as, unlike the schools' sector, there is 'currently no national agreement on

lesson observations with the Association of Colleges' (UCU 2009: 2), though legislation in the 2010 white paper for schools (DfE 2010) has since threatened to undermine this agreement in schools. The FE stance came as a response to complaints from members regarding the draconian, 'intimidatory and not supportive' way in which observation was being used by managers in some colleges (Lee 2007). At the centre of the debate was the issue of grading. O'Leary (2011) found in his research that in some cases, teachers are threatened with disciplinary action or denied annual pay increments if they are rated as a grade 4 or even a grade 3 on the Ofsted scale. In recent times the issue of unannounced observations has also become another hotly debated topic. Unions contest that they are being used by some unscrupulous employers to single out and harass specific individuals, though practices seem to vary across the sector with some institutions including them as formal, graded assessments and others as informal, ungraded visits. Interestingly, in the USA, where they have been in use for some years, Downey et al (2004) argue that unannounced observations or 'walk-throughs' as they call them, were originally intended to be separate from any formal teacher evaluation process and to be used strictly as a means of engaging teachers in dialogue and reflection about teaching practices and school-wide goals.

From 2006-09, researchers based at the University of Huddersfield published three separate reports (Burrows 2008; Ollin 2009; Peake 2006) into the use of observation in the context of ITT programmes as part of the 'Huddersfield Post-Compulsory Education and Training (PCET) Consortium'.

The first of these reports (Peake 2006) explored the perceptions of teacher educators and trainee teachers concerning the purposes of observation and sought to identify examples of good practice. The research methods consisted of two survey questionnaires, one for trainee tutors and the other for subject specialist mentors, as well as interviews with eleven teacher educators working at five different centres. In total there were 134 responses to the trainees' questionnaire and only four responses to the mentors' questionnaire, which had been sent to 12 mentors in total. No explanation was provided for why there was such an imbalanced ratio between these two groups, which was surprising given how the report emphasised the importance of subject mentors in the observation process.

Although the research lacked a discernible theoretical framework with which to analyse the data, its key findings revealed some interesting areas of discussion resonating with issues covered in related studies. Below is an adapted summary of these based on the original report:

- Importance of the observer being a subject specialist
- Conflicting purposes of ITT and QA observations

- Concerns regarding the lack of consistency and standardisation of practice between observers across the consortium
- Trainees avoid taking risks in observed sessions
- Value of peer observation in professional development
- Observation is resource intensive – it is time-consuming and expensive

The second report (Burrows 2008) focused on exploring trainees' perceptions of observation as part of the 'new curriculum' i.e. the increase in observation to a minimum of eight hours under the LLUK qualifications (2007), together with the undertaking of mentor observations to support the development of subject specialist pedagogy. The project was underpinned by two aims: 'To identify the perceptions of trainees of observations within the new curriculum [and] to formulate an action plan to improve observations based upon the analysis of the research' (p. 5).

Although the research aimed to explore issues confronting both pre and in-service trainees and their observers during observation, only the former were included as part of the sample. Eighty participants completed a questionnaire similar in focus and design to that used in Peake's (2006) study, followed by four focus groups in which sixty five trainees participated. Like Peake's study, there is a lot of descriptive detail but no theoretical underpinning. Notwithstanding these shortcomings, the key findings summarised below add to the undeveloped field of observation in FE:

- More structured training of observers needed, particularly subject mentors who are unfamiliar with the observation process
- Over fifty per cent of respondents thought that ITE observation should not be graded compared to a third who thought that it should.
- Observation as 'formative' assessment is a key means of supporting professional development and more value should be attached to it than course assignments. Peer observation is highly valued.

The third report (Ollin 2009: 5) was carried out:

[I]n response to the introduction of Ofsted's new grading criteria for inspection of ITE in the [sector] (2009), which state that over fifty per cent of trainees need to be judged 'outstanding' for an ITE provider to achieve the highest inspection grade.

The research explored the implications of introducing the Ofsted scale for teacher educators across the consortium. Ollin remarked that the grading criteria on Huddersfield's programmes 'previously operated on a pass/fail basis' and that the transition to the Ofsted scale 'will influence the way that Certificate in Education/PGCE programmes are developed and delivered' (p. 7). This is a very significant point highlighting the challenges faced by providers as a result of having

to adapt to changing policy, while also protecting the values and beliefs that underpin many ITE programmes i.e. the emphasis being on encouraging teacher development through a combination of formative and summative assessment where the former is prioritised over the latter. Nevertheless, Ollin fails to point out that providers are not obliged to adopt Ofsted's scale as a replacement for their own assessment criteria. It is the choice of an institution to decide to do so or not.

The underpinning aims of Ollin's (2009: 12) research were, 'To develop a working conceptualisation of what constitutes 'outstanding' teaching [and] to use this information to further develop staff and quality systems, taking into account issues of grading of trainees' practical teaching'. The research was 'qualitative and interpretive in nature' and used interviews and observations. These included a small sample size of nine case studies of 'outstanding' and 'weak' trainees during which the ITE tutors were observed carrying out observation and giving feedback, followed by a semi-structured interview 'focussing initially on understandings and judgements of "outstanding" related to the specific observation' (p. 15). Some of the key findings revealed:

- Mixed interpretations as to what constitutes 'outstanding'. The conclusion reached is that 'outstanding teaching is more than the sum of its parts' (p. 30).
- The need to consider the effect of context on the notion of 'outstanding'.
- Tensions between the 'dual identities' of those in-service trainees observed on ITE programmes and as part of internal college QA schemes. In the first instance they are seen as 'students' with 'developmental needs' and in the second as employees with an obligation to prove their professional competence. Similar tensions were revealed by observers who were involved in observation as ITE tutors but also in college observation for QA purposes.
- Resistance on the part of observers to grading based on the premise that it undermined the developmental nature of ITE observation.

Like the two previous reports from the Huddersfield Consortium, this report is descriptive. The research concludes that the Ofsted grading criteria are likely to have a significant impact on future ITE programmes across the Huddersfield consortium and present a challenge in 'balancing the underpinning values related to the learning and development of trainees with increasing demands for standards of teaching to be monitored and assured' (p. 6). It is surprising though that there is no reference to previous studies carried out in this specific field (e.g. Cope et al 2003; Sharp 2006), which have argued that the graded assessment of observation for trainee teachers is unsustainable on the basis that:

There is no published research which confirms that meaningful grading is possible. Attempts to implement grading schemes ignore the lack of support

from research and imply that the assessment of teaching is based on measurement rather than professional judgement (Cope et al 2003: 683).

Finally, Lawson's recent research (2011) has explored the use of an 'observational partnership' between a university education department and three local colleges. After outlining the principles and processes underpinning the partnership, along with a brief summary of its history, he goes on to discuss the findings from his research, which was based on a content analysis of the texts of 924 observation report forms collected between 2002 and 2009. The aims of the analysis were to 'establish an understanding of which practices may be successfully changed through observation of the classroom and which may be more resistant to transformation, as well as giving insights into the process of enacting change in the classroom' (p. 10).

Lawson's analysis identified two areas of teachers' classroom performance that appeared to lend themselves more easily to change ('planning for learning' and 'assessment for learning') and two that were more resistant ('questioning' and 'student involvement'). In relation to the former, the findings revealed no discernible patterns as to why these two aspects seemed more conducive to change than others. With regards to the latter though, Lawson argues that in the case of 'questioning', 'the practice is so complex and nuanced that it is difficult for teachers to develop their practice'. As for 'student involvement', Lawson puts this down to 'deeply ingrained habits and suppositions about teaching' (p. 17). In other words, old habits die hard and it is difficult to get those who have spent their entire career using a teacher-centred approach to embrace a more student-centred philosophy in their teaching, though this is also exacerbated by the pressure faced by many teachers to 'get through the curriculum' so as to ensure as high a level of achievement amongst their learners as possible.

Lawson concludes by stating that 'sustained observation offers a robust way of changing some classroom practices and of making inroads in others'. He attributes the success of the partnership's observation scheme to its collaborative nature and its 'continuity', which has helped to establish a mutual understanding and trust amongst those involved as to what its purpose is i.e. to encourage an open, shared dialogue between observer and observee with a view to further improving teaching and learning. Notwithstanding this, he is mindful of how not all those observees involved were 'open to the possibilities of change' and this was manifested by them 'going through the motions' (p. 18).

The HE sector experience

It is only in recent years that observation has begun to emerge in HEIs. This has been partly fuelled by QA demands for greater accountability but more increasingly as a result of its potential for supporting the CPD of lecturers (Hammersley-Fletcher and Orsmond 2004, 2005; McMahan et al 2007; Shortland 2004). Unlike FE and the

schools' sector, its use is much less commonplace or prescribed. There is less evidence of links to formal, centralised QA systems and it appears to operate mostly on an informal, voluntary and departmental basis. Hardman's research (2007) revealed that it also occurred as part of academic programmes of professional study for staff, such as the postgraduate award in Learning and Teaching in HE, a compulsory qualification for new staff in many post-1992 universities. Even in such contexts though, there seems reluctance to grade performance as the emphasis appears to be on the developmental support of staff rather than an evaluative judgement of them.

The dominant model used in HE would appear to be 'peer observation' (Hammersley-Fletcher and Orsmond 2004, 2005; Peel 2005; Shortland 2004). Shortland (2004: 220) defines peer observation as 'peers observing each other's teaching to enhance teaching quality through reflective practice, thereby aiding professional development'. There are some researchers, however, who contest the generic application of peer observation in HE as an all-encompassing term for observation and instead prefer the label 'third-party observation' (Fullerton 2003; McMahon et al 2007). For them the term peer observation refers to a specific model of observation based on a collaborative partnership between peers, which is underpinned by 'equality between observer and observed' (McMahon et al 2007: 500). This is a legitimate and helpful terminological distinction to make, especially if we are to avoid a blurring of the boundaries between the different models, contexts and purposes of observation.

Extant research reveals a commonality in the key issues, most of which centre on the perceived opportunities and threats associated with the use of peer observation in HE. Peel (2005) is mindful of its potential danger as a surveillance tool on an institutional level. Research carried out among GP teachers revealed opposition to schemes that used peer observation to address the twin aims of teacher development and QA. Such schemes were considered 'unlikely to succeed if seen to be conveying quality assurance in the guise of tutor support' (Adshead et al 2006: 72). The transparency of the aims and objectives of any peer observation scheme in HE is regarded as fundamental to avoid it being viewed with suspicion by lecturers (Hammersley-Fletcher and Orsmond 2005; Peel 2005; Shortland 2004).

In a small-scale qualitative study involving eighteen interviews with lecturers from two academic schools of a post-1992 English university, Hammersley-Fletcher and Orsmond (2005) explored their experiences as participants in a peer observation scheme. Their findings revealed uncertainty regarding the expectations of their roles as both observer and observee. Some lecturers felt uncomfortable about providing critical feedback for their peers, which the writers (2005: 218) recommend 'must be presented in ways that are constructive and will lead to new understandings and improved practice'. The uncertainty and unease expressed by lecturers showed how

a shared understanding of what was meant by the term 'critical feedback' was missing. It also exposed their lack of experience in providing constructive feedback.

In Hammersley-Fletcher and Orsmond's study the success of the peer relationship between observer and observee was seen to be dependent on the notions of trust and confidentiality. These were considered fundamental to facilitating honest reflection. It also emerged as an important issue in other studies (e.g. Gosling 2002; Shortland 2004). Gosling (2002: 2) talks about the need for staff to be seen as 'genuine peers in which there is real mutuality and respect for each of the participants as equal'. He suggests that the process can be undermined if the observer is senior in hierarchy to the observee, although his claim is unsubstantiated. His concerns seem to be based on the premise that such a relationship is likely to result in more senior members of staff taking charge, hence threatening the equality of the interaction.

In an autobiographical study, Peel (2005) reflects on her personal experiences as a new lecturer and examines the arguments for and against peer observation. She avers that it can be a useful means of facilitating reflection as long as it incorporates reflection on wider issues of the teaching and learning process and not just that of the observed lesson. She remarks that it was as a result of engaging in critical reflective thinking triggered by the feedback element that led to her successful CPD rather than discussion centring on the observation itself. Thus, peer observation is being used as a 'lens' to stimulate critical reflection (Brookfield 1995). Similarly, Hammersley-Fletcher and Orsmond (2005: 223) highlight the importance of reflection. Like Peel, they see peer observation as 'a vehicle for encouraging academics to develop their reflective thinking about their role as professional lecturers and to seek and engage in developmental processes as a result'.

Following the Browne (2010) review into university funding, the president of the National Union of Students (NUS), Liam Burns, called for all university lecturers to be subjected to similar training programmes to those of their counterparts in schools and colleges (Boffey 2012). His rationale for the need for the introduction of such qualifications was set against the background of the significant rise in university tuition fees and the need to ensure QA mechanisms that students were being taught by appropriately qualified staff. If the government were to follow up this recommendation, it would suggest that it may not be long before the use of lesson observation is introduced as a form of measuring quality/standards in HE.

Synopsis of key themes and issues across the sectors

There are clearly recurring themes surrounding the use of observation in all three sectors. For example, its value as a means of stimulating reflection on practice by engaging in professional dialogue with colleagues, who act as 'critical mirrors' (Brookfield 1995) seems to be a shared interpretation among researchers and practitioners in all three sectors, albeit with the caveat that specific ground rules

need to be established for this to work successfully i.e. notions of mutual trust, respect, ownership etc. At the same time, there are divergences between the three that partly reflect their historical status and the history of policy in each sector. In FE and schools observation appears to have operated principally to satisfy policy driven agendas of performance management systems, especially in FE. In HE, its role is less prescribed, thus allowing lecturers more autonomy and control over its use, though arguably this can be attributed to the fact that the observation of teaching is not included in Quality Assurance Agency (QAA) inspections of HEIs whereas it is one of the main sources of evidence in Ofsted inspections.

The use of graded observation has triggered debate regarding the reliability of observation as a form of assessment in schools (Wragg 1999) and more so recently in FE (O'Leary 2011). In some ways graded observation is perhaps the single most contentious issue relating to the topic. Its performance focus is something that seems to have provoked strong reactions across both sectors. Brown (in Brown et al 1993: 51) compared such models of observation to the traditional examination in that both teacher and student are required to produce 'peak performance under stressful conditions with little opportunity for dialogue with the examiner and no real chance to gain meaningful feedback on how things are going.' There is a natural link here to Foucault's (1977) notion of the 'examination', where the performance of the teacher becomes subjected to a process of 'objectification' through the system of graded observation. Gipps (1994) refers to such types of assessment as 'high stakes'. In her work on the assessment of high school students, she found that normative grading threatened collaborative learning by provoking unhealthy competition and impacted negatively on levels of motivation. More recently, Coffield (2012) has expressed similar concerns regarding the use of graded observations in FE colleges.

Some of the existing research has highlighted how the performative nature of graded observations has resulted in a decline in the creativity and innovation of teachers' work in the classroom (e.g. Coffield 2012). There is a reluctance to want to 'take risks' for fear of being given a low grade. Teachers are aware of the need to 'play the game', which can result in them following a collective template of 'good practice' during observation. According to Elliott (1990: 83), this is an example of 'management exercising control over performance by preventing teachers from reflexively developing new understandings of the nature of teaching and learning tasks'. In FE, Peake (2006) has illustrated how even in the context of ITE trainees avoid taking risks during observation. Yet recent research into the use of lesson study among qualified teachers in schools seems to suggest a counterbalance to this.

With regard to the use of observation as a formative tool for CPD, there would appear to be a commonality across much of the literature in terms of some of the key concepts discussed i.e. collaboration, equality, autonomy, ownership, trust etc. Much of this work has focused on the use of lesson study in schools and peer observation

in HE. Referring to Ramsden (1992), Jones (in Brown et al 1993: 31) comments that in order for observation to work it needs to be part of a teacher's professional development and not something that is 'done to them':

Ownership of observation needs to be devolved down as much as possible to the participants in the teaching process. The closer the ownership of the process is located to the actual participants, the more likely it is that the aims will be achieved and the outcomes accepted by all concerned (Brown et al 1993: 10).

To conclude, across all three sectors previous studies have revealed that observation is regarded as an important means of evaluating, reflecting on and improving the quality of teaching and learning as well as contributing to a greater understanding of these processes. Whether this occurs as part of QA systems or CPD programmes, the central role that observation has to play in the professional practice of teachers seems incontestable. Where the contestations start to emerge, however, is in relation to the stated aims behind its use, the extent to which the outcomes match these aims and the way in which the process of observation is operationalised.

Wragg (1999) argued that the *purpose* of observation should largely determine how it is used, but evidence above and sections 5 and 6 of this report suggests that the boundaries between different models, contexts and purposes have become blurred and contested. At the heart of these contestations lies a conflict between 'structure' and 'teacher agency', and related notions of power and control that manifests itself in the sometimes paradoxical agendas of policy makers, the institution and its teaching staff. This conflict is epitomised by the way in which the developmental needs of staff and the requirements of performance management systems are forced to compete as they are often conflated into a 'one-size-fits-all' model of observation in schools and colleges, with the latter overshadowing the former.

Summary

This section has reviewed relevant literature and previous studies of lesson observation across the schools', FE and HE sectors. It has identified key themes and issues to emerge in each sector and provided a synopsis of areas of commonality and differences. Many previous studies and publications have focused largely on descriptive accounts of *practice* with limited discussion of the wider contexts and cultures in which it is situated. Where there have been links to context these have often occurred in relation to ITT. Nevertheless, more recent work in FE (e.g. Coffield 2012; O'Leary 2011, 2013a, 2013c) has helped to develop a synthesis between the practice of observation and the contexts in which it occurs. The use of observation in the university sector is still relatively new, with much of the existing research focusing on models of peer observation and its application as a tool for reflection. Finally, the key themes and issues explored in this section will serve as useful reference points for discussion throughout the rest of this report

4. Research methodology

Research design

This research project adopted a mixed-methods approach involving quantitative and qualitative methods of inquiry. Johnson and Onwuegbuzie (2004: 17) define mixed-methods research as:

A class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study.

It was anticipated that the combination of these methods would help to provide the study with a balance between breadth and depth in its data collection and analysis, as well as building on their complementary strengths. 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 as listed in the report's introduction. Some of these questions were of a factual nature and thus leaned themselves to a quantitative method of inquiry. Others sought to explore the experiences and perspectives of practitioners in the form of a narrative and so required a qualitative approach. As Arksey and Knight (1999: 3) have argued, 'perception, memory, emotion and understanding are human constructs, not objective things'. Interviews and focus groups therefore represented an opportunity to gather 'rich data on people's views, attitudes and the meanings that underpin their lives and behaviours' (Gray 2004: 213).

The rationale for a mixed-methods design was pragmatic and principled. It was pragmatic in the sense that developing as thorough an insight into lesson observation as possible was what drove the selection of research methods overall rather than any affiliation to a specific methodological paradigm. As Tashakkori and Teddlie (1998: 21) state:

For most researchers committed to the thorough study of a research problem, method is secondary to the research question itself, and the underlying worldview hardly enters the picture, except in the most abstract sense.

Thus decisions about what data to collect, what were deemed to be the most appropriate and effective means of collecting the data, along with what to do with the data were 'dictated by the research question[s]' (Newman and Benz 1998: 15), the underpinning aims of the study and a commitment to the quality of the research.

The decision to use mixed methods was also principled in the sense that the study was conducted on the basis that neither a qualitative nor a quantitative approach can be considered superior to the other. 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 and Morrison 2011: 22). Both methodological approaches have their strengths and weaknesses, as others have argued (e.g. Punch 2006) and 'even greater strength can come from their appropriate combination' (Gorard and Taylor 2004a: 1).

The quantitative data elicited through the use of an online questionnaire in the first phase of the project was expected to provide a broad, panoramic sketch of the research topic to which more colour and specific detail would be added with the qualitative aspect. For example, some of the survey data provided an overview of current models of lesson observation in use nationally as well as members' views on the impact of these models. Yet to gain an understanding of how lesson observation could be used most effectively to promote lecturers' professional learning and development, more detailed qualitative data were needed and this required the use of interactional methods of inquiry such as interviews and focus groups. Combining these different methods was important in adding to the triangulation of the data.

Triangulation is fundamental to strengthening the rigour of the data collection and analysis process and plays an important part in increasing the overall validity and reliability of the research undertaken and ultimately its findings. Combining different research methods, data and participants allows for the capture of a range of viewpoints, which can lead to a better understanding of the research topic under investigation and a more rounded interpretation. Differing findings and perspectives can be compared and contrasted and consistent trends can help to strengthen the credibility of a particular category of analysis.

Triangulation is often mentioned as one of the main advantages of a mixed-methods approach (Gorard and Taylor 2004b; Tashakkori and Teddlie 1998). Miles and Huberman (1994: 267) see triangulation as a 'way of life' rather than a 'tactic', which becomes an integral part of verifying the research process:

If you self-consciously set out to collect and double-check findings, using multiple sources and modes of evidence, the verification process will largely be built into data collection as you go.

Denzin (1978) is commonly acknowledged as one of the first qualitative researchers to emphasise the importance of triangulation and its relevance to qualitative methods. Denzin discussed four basic types of triangulation:

1. *Data triangulation* (the use of a variety of data sources in a study)
2. *Investigator triangulation* (the use of several different researchers)
3. *Theory triangulation* (the use of multiple perspectives to interpret the results of a study)

4. *Methodological triangulation* (the use of multiple methods to study a research problem)

(Cited in Tashakkori and Teddlie 1998: 41)

This study applied 1, 3 and 4 of Denzin's types of triangulation. The use of an online questionnaire that captured both quantitative and qualitative data as well as qualitative interviews and focus groups ensured the collection of varied data sources. Theory triangulation was evidenced by the application of a theoretical framework that drew on concepts from differing disciplines, which helped to broaden the scope of the interpretation of a complex data set. Methodological triangulation was employed by combining different methods of data collection and analysis. Thus, for example, analysing qualitative data from interviews in conjunction with quantitative data from questionnaires provided a way of mediating diverse interpretations, understandings and meanings in contested contexts and situations that required juxtapositional analysis.

Gorard and Taylor (2004b) claim that 'complementarity' is the defining quality of triangulation. In this study, quantitative and qualitative research methods were selected to complement each other so that their combination would result in the creation of a more complete picture of the research topic. Denzin and Lincoln (2003: 8) maintain that the use of:

[M]ultiple methods, or triangulation, reflects an attempt to secure an in-depth understanding of the phenomenon in question ... the combination of multiple methodological practices, empirical materials, perspectives, and observers in a single study is best understood as a strategy that adds rigor, breadth, complexity, richness and depth to any inquiry.

Research sample

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 on the basis that it 'allows us to choose a case because it illustrates some feature or process in which we are interested' (Silverman 2006: 306). 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.

Another element of the purposive sample was that two of these colleges were identified at a UCU London regional committee meeting in December 2012 as having implemented 'alternative' models of observation. Given that part of the focus of the second phase of the data collection was on exploring those aspects of lesson

observation practice that were considered to create optimum opportunities for expansive professional learning and development among tutors, it was considered important to ensure that colleges that were recognised as having experience in having worked with different models of observation were included in the sample as they might have an important contribution to make to the discussion.

It is acknowledged that the use of a purposive and/or convenience sampling strategy carries with it certain limitations (Cohen, Manion and Morrison 2011; Gorard and Taylor 2004a; Robson 2002). For example, it was recognised that employing such a strategy when selecting the participating colleges might limit the external validity of the research. Inevitably, an element of caution needs to be applied when interpreting the findings and reaching conclusions. Notwithstanding these limitations, the profile of the sample was considered broadly similar to the national picture, though it is not claimed to be representative of all FE provision in England.

All participants of the study were notified that any information they provided would be dealt with confidentially and that every effort would be made to protect their identities at all times, as too would the identity of each participating college. It was anticipated that there might be limitations to maintaining the complete confidentiality of some participants when writing up the findings inasmuch as some might be more readily identifiable than others within their institutions as a result of their respective roles; this was made clear to all participants and their consent was established before the final draft of the report was completed.

Research methods: Online questionnaire

During the first phase of data collection, an online questionnaire was circulated via Survey Monkey to all UCU FE members whose up-to-date email addresses were stored on the union's central database. It was anticipated that the use of an online questionnaire would increase the overall response rate. Harris (1997) argues that one of the advantages of online surveys is that they have the potential to appeal to an increased audience. Added to this is the fact that they supply data quickly and can be completed at a time and place convenient to respondents who prefer to work via an electronic medium (Madge and O'Connor 2002). In total there were 3958 returns, of which 3525 were fully completed questionnaires with 432 partially completed. UCU FE membership was reported to be approximately 32,000 at the time the survey was circulated, thus there was an overall response rate of just over 11%.

In the opening paragraph of the questionnaire participants were assured that their identities would remain protected at all times and their responses treated confidentially. The survey was divided into six sections (see Appendix 1). Section A included a range of questions designed to collect demographic data about the participants in relation to their gender, employment status, teaching experience etc. It also included factually oriented questions relating to the policies and procedures

of observation in the participant's workplace e.g. which models of observation were most commonly used in the participants' workplace.

Sections B-E each contained a list of matrix questions relating to graded, ungraded models of observation, observation feedback and unannounced lesson observations. Participants were asked to respond to the statements via the use of a Likert-type ordinal scale of 1-4, ranging from 'strongly agree' at one end of the scale to 'strongly disagree' at the other and a fifth 'not applicable' box at the end of the scale. The decision to use a four-point as opposed to a five-point ordinal scale with a 'neutral category' was influenced by relevant literature highlighting the 'central tendency' issue whereby respondents opt for the mid-point in a five-point scale as a means of 'sitting on the fence' (e.g. Cohen, Manion and Morrison 2011; Oppenheim 2001). This was justified by the fact that only two respondents commented in the text box at the end of the questionnaire that they would have preferred to have the option of a 'neutral' category.

Section F contained a blank box where participants were able to leave extended qualitative comments. Placing it at the end of the questionnaire provided participants with the choice of whether or not they wished to spend more time on it. Just under half of those respondents who completed the survey wrote comments (n = 1619), reinforcing the idea that lesson observation was a topic of significant interest to them. In fact, some participants' responses ran into several pages in length. There was also a space for respondents to leave an email address and/or contact telephone number at the end of the survey if they wished to give their consent to participate in an interview/focus group in the second stage of data collection. Over a quarter of respondents left their contact details (n = 1012). In order to protect anonymity, all references to the identity of an institution or its employees were removed when reporting on the data and respondents' comments were simply assigned a number.

Research methods: Interviews and focus groups

Interviews were carried out with the person responsible for overseeing lesson observation in each of the participating colleges. This was invariably someone from the senior management team who occupied a position such as the vice principal for teaching and learning, director of quality etc. Focus groups were held with practitioners, sub-divided into groups of observers and observees whenever possible, though several participants straddled these two groups.

The interviews and focus groups followed a semi-structured format. A list of predetermined questions was used to provide a structure and to help guide the discussion (See Appendix 2). At the same time, however, a flexible approach was adopted that allowed me to respond spontaneously, or to tailor the discussion appropriately (Robson 2002). Examples of such changes included the ordering, rewording or rephrasing of questions at times. All interviewees were assured that

their identities would remain protected at all times and their responses treated confidentially. Each interviewee was also informed that their real name would be replaced with a pseudonym, as was the case with the colleges included.

Participants were interviewed with the common aim of exploring their views on those models of lesson observation in use in their workplace and their impact on improving teaching and learning along with their own professional skills and knowledge base. Discussion also centred on whether participants perceived certain models of observation to be more worthwhile than others and what improvements, if any, they considered important to make in order to create optimum opportunities for expansive professional learning and development for all those involved in the observation process. All interviewees were asked a similar set of questions, which helped to strengthen the validity and reliability of the interview schedule by maintaining a consistency across different interviewees i.e. senior managers, observers and observees. What Cohen, Manion and Morrison (2011: 204) refer to as 'face validity' i.e. 'whether the questions asked looked as if they are measuring what they claim to measure' was reflected in the interview schedules (See Appendix 2).

Data analysis

The data analysis process began with the questionnaires, which contained quantitative and qualitative data. Initially these two data sets were analysed separately, mainly due to the practicalities of managing such a large amount of data rather than any fundamental differences in principles to analysing quantitative and qualitative data. There were obviously differences in technique when it came to analysing the two data sets inasmuch as the quantitative data were fed through pre-developed systems for analysis as part of an online survey with Survey Monkey, whereas the coding of the qualitative data was not something that could be predetermined and thus evolved on an on-going basis. Nevertheless, there were similarities in the procedures followed for analysing both sets of data, illustrating the iterative connections between these mixed methods. For example, in both cases I went through a process of reading and re-reading all the data in order to become completely familiar with them. The analysis of all the data collected throughout the project was based on a continuous process of comparing chunks of data (e.g. interview transcriptions and questionnaire data), identifying similarities and differences among them and re-visiting them as the data analysis ensued in order to develop and refine categories.

The approach I adopted in analysing the qualitative data combined steps outlined by Creswell (2003: 191-195) and Miles and Huberman's (1994: 9-12) 'three concurrent flows of activity' of analysis i.e. *data reduction, data display and conclusion drawing/verification*. I started by reading and re-reading all the written comments from the questionnaires to get a feel for the data and to see if any recurring issues or

themes as well as discrepancies emerged from them. The coding of the data into categories involved a broad, 'grounded approach', which was informed by my existing knowledge of the field as well as drawing on relevant literature and theory. Thus the categories were not pre-determined, but emerged from the data.

Following the preliminary stage of coding, twenty-one themes/issues were identified. These were strengthened, re-categorised or discarded as more data were analysed. Interrelationships between emerging themes/issues across participants and data sets were explored to identify key patterns. The second stage of coding involved further re-reading of the data and was concerned with establishing a framework for linking the preliminary set of twenty-one themes/issues, which was subsequently reduced to eighteen. The whole database was reviewed for further evidence to support existing themes and patterns and to identify any new ones. Those that appeared consistently were subsequently incorporated into the key findings. In short, the process of revisiting the data on several occasions led to a refinement of the categories and a better understanding of the relationship between them. The eighteen key themes/issues to emerge from the data were then finally incorporated under four overarching thematic categories, which are discussed in the following section of the report.

5. Findings and discussion

This section presents and discusses the study's key findings, drawing on research data taken from the online questionnaire, semi-structured interviews and focus groups. Where possible the presentation of different data sets is integrated i.e. where thematic links occur naturally, in keeping with the study's mixed methods approach. This is to ensure a varied presentation of the findings and breadth of coverage in what amounted to a very large collection of data.

There were chunks of quantitative data that did not lend themselves to being thematically linked with some of the qualitative data but were, nevertheless, important in their own right to report e.g. demographic data from the sample. So as not to exclude these data and to 'set the scene' of the sample, they are presented at the beginning of this section.

The quantitative data presented in **Figures 1-5** below thus help to provide a descriptive overview of the demographic profile of the sample, covering the areas of gender, employment status and capacity (i.e. position held within the institution), teaching experience and respondent status i.e. if the participant was an observer, observee or both.

Figure 1 below presents the breakdown of respondents according to gender. The ratio of female to male participants, broadly 60-40 per cent, reflected the national profile according to recent statistics on the FE workforce (LSIS 2012).




| 1. What gender are you? | | | |
|-------------------------|---|------------------|----------------|
| | | Response Percent | Response Count |
| Male |  | 42.6% | 1,685 |
| Female |  | 57.4% | 2,269 |
| Transgender |  | 0.1% | 2 |
| answered question | | | 3,956 |
| skipped question | | | 20 |

Figure 1 - Gender



Figure 2 – Employment status

As **Figure 2** above reveals, just under two thirds of respondents were employed on full-time permanent contracts, with the majority of the remaining respondents either on fractional permanent or hourly paid contracts. Less than 2% of respondents were either retired or seeking employment. When these figures were compared across genders, there were twice as many female members of staff employed fractionally or on a part-time basis as their male counterparts. Once again, this statistic is consistent with national trends.

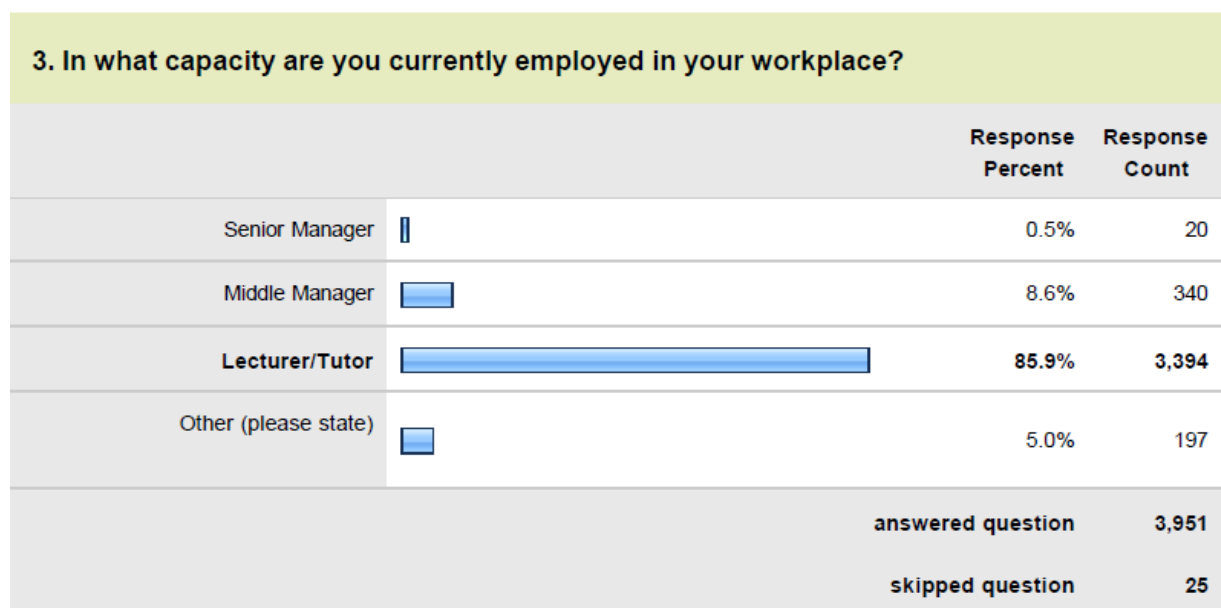


Figure 3 – Employment capacity

Over four fifths of participants were employed principally in a teaching role as shown in **Figure 3** above. Just under a tenth of respondents described themselves as middle (n = 340) or senior managers (n = 20), although the vast majority came under the former category. There were no significant variations across gender groups in their employment capacity. There were, however, some noteworthy differences in responses according to employment capacity in sections A-E of the questionnaire, which are discussed below when comparing cross-tabulations.

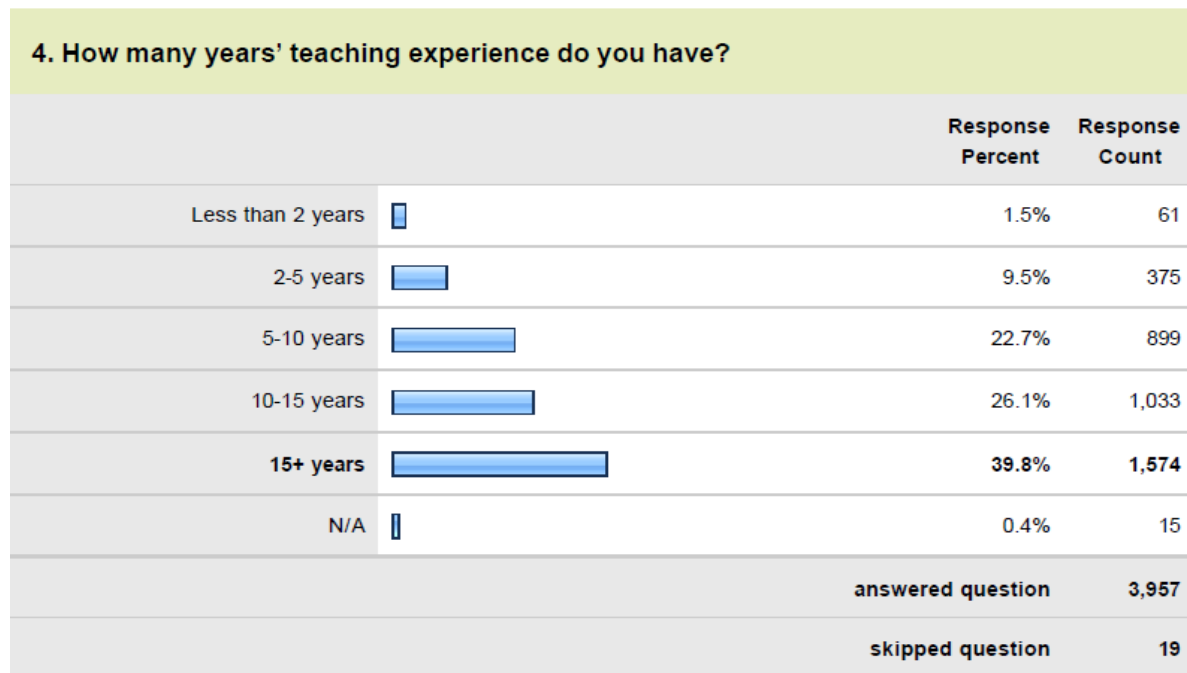


Figure 4 – Years of Teaching Experience

The percentages relating to the years of experience in **Figure 4** above were representative of national trends (LSIS 2012). Well over four fifths (88.6%) had been teaching for at least five years and two thirds had more than ten years' experience. When compared across gender groups, there were no significant differences in years of teaching experience. It was clear that a majority of respondents were able to draw on substantial levels of experience when completing the questionnaire. It is worthwhile noting though that there was evidence of some variation in response to certain questions depending on the years of experience of participants. Discussion relating to **Figure 9** below provides an important contrastive example of this.

Figure 5 below revealed that the overwhelming majority of respondents (four fifths) were involved in the observation process as observees in contrast to those who occupied the dual role of observer and observee, which was less than a fifth and those whose role was solely as an observer (3.6%). Senior and middle managers comprised the latter group, whereas those that straddled both groups (i.e. observer and observee) predominantly occupied middle management roles and as such were

still involved in teaching themselves. This is not an insignificant observation as an issue to emerge from the survey's qualitative comments and the focus groups and interviews centred on the importance of observers remaining active as classroom practitioners, as discussed in more detail below. In addition, cross tabulations revealed some disparities in the responses of observers, observees and both (see below), particularly regarding their views on graded observations and feedback.

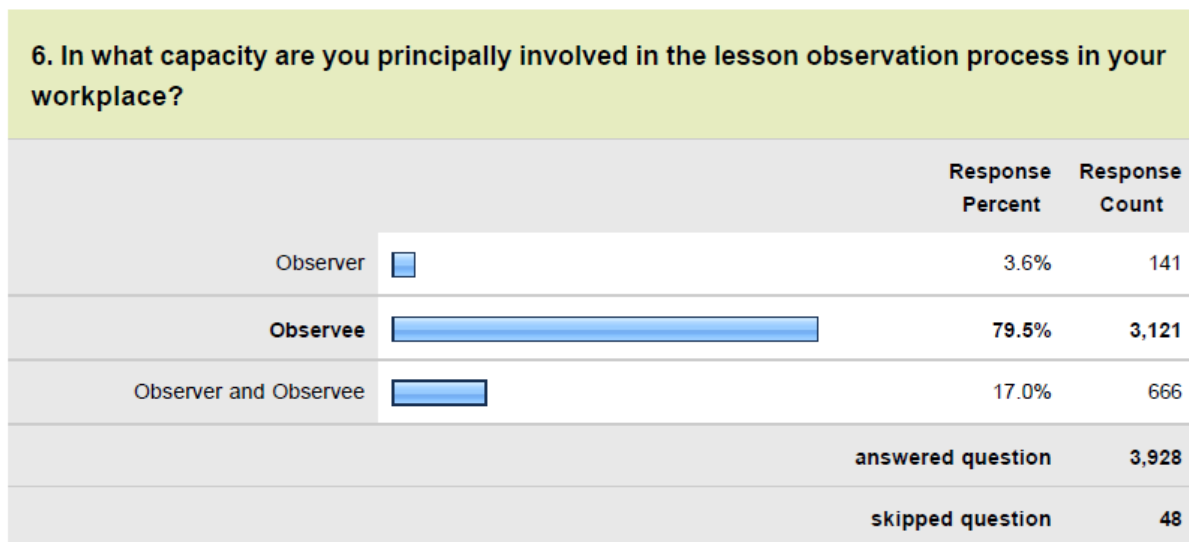


Figure 5 – Observer, observee or both?

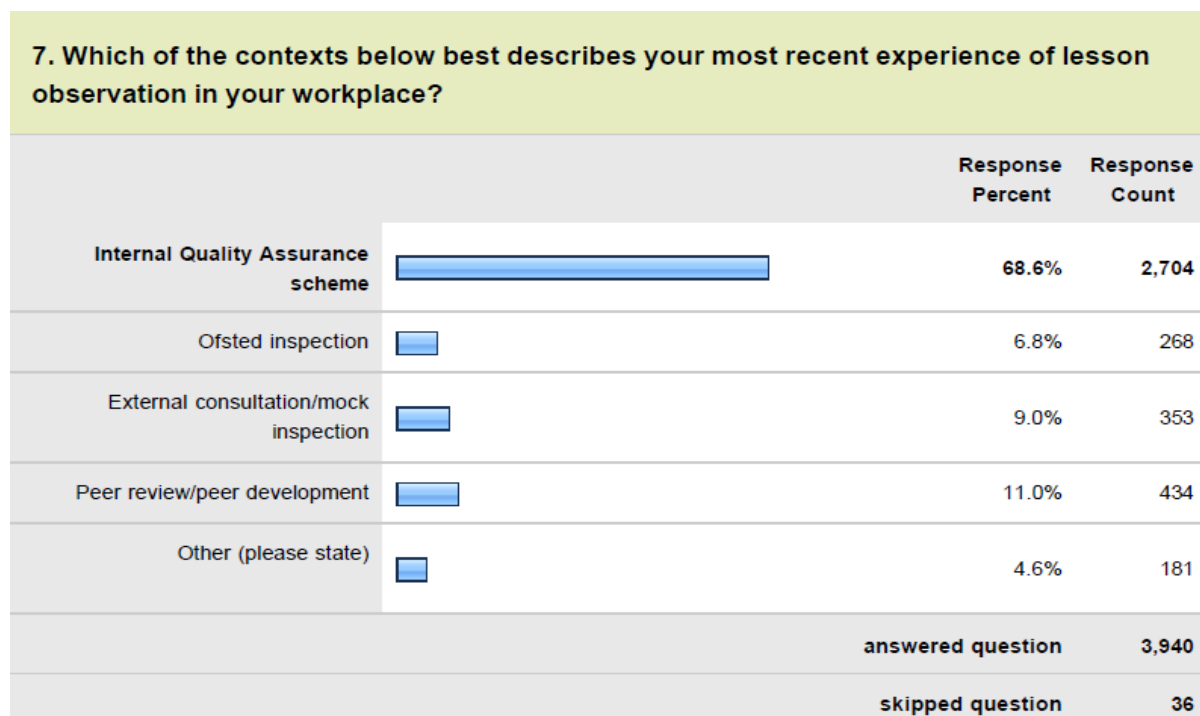


Figure 6 – Contexts of lesson observation

In order to get an overview across the sample of the contexts and purposes for which lesson observation was used, a specific question was included that required respondents to indicate the context that best described their most recent experience of observation. **Figure 6** above presents a summary of the responses to that question.

Just over a tenth of respondents (11%) chose peer review/development, which would suggest that this particular use of observation was relatively marginal in the sector, with much of it taking place in the ITE context. By far 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 where the lesson is evaluated and scored against the 4-point scale (Ofsted 2012). Similarly, the context of ‘external consultation’ also follows the Ofsted model and tends to be used by colleges as a mock inspection, where external consultants are employed to carry out observations across the institution. So when combined, the first three contexts listed in **Figure 6**, all of which adopt a similar QA approach, amounted to over four fifths (84%) of the responses. This statistical return was very similar to the findings of a previous, smaller-scale study carried out in ten colleges across the West Midlands (O’Leary 2011), thus reinforcing the view that lesson observation was predominantly associated with QA and was closely aligned to performance management systems in FE.

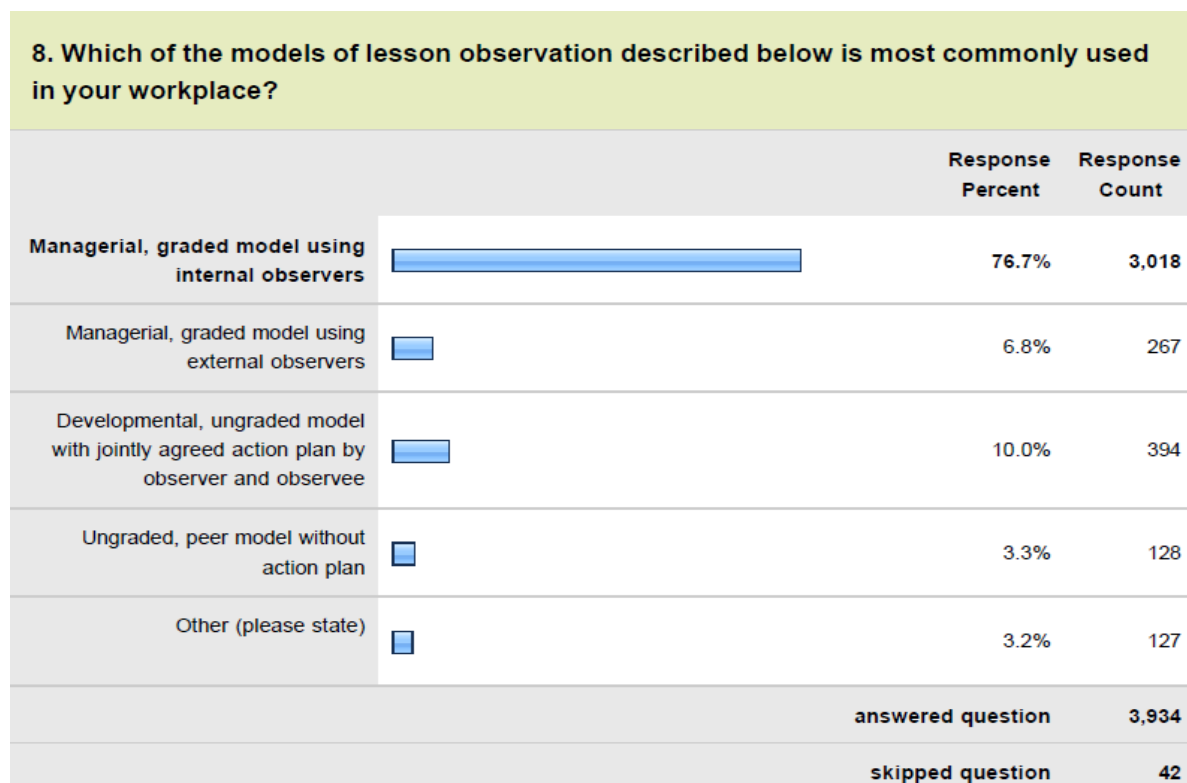


Figure 7 – Models of lesson observation

Following on from identifying the observation contexts, **Figure 7** above sought to categorise the particular models of observation that were most commonly used in FE. Unsurprisingly, there were correlations between the responses in **Figure 7** and those previously discussed in **Figure 6**. Thus, the ‘managerial, graded model’ accounted for over four fifths (83.5%) of responses, whereas ungraded models just over a tenth (13.3%). These trends were reinforced in the qualitative data.

Unannounced observations or ‘learning walks/walk throughs’ (as they are commonly referred to in the sector) were not included as a separate ‘model’, as they were assumed to be an extension of internal QA schemes when the survey was created, an assumption which was largely confirmed across data sets. With hindsight, however, it may have been helpful to differentiate between the QA and peer review/developmental contexts for walk throughs, as it is acknowledged that it cannot be assumed that practice is uniform across all institutions in the sector.

The final question in Section A of the survey asked respondents to indicate whether unannounced observations were in use in their workplace. As discussed previously, unannounced observations are one of the most recent developments in FE and emerged largely as a response to Ofsted’s decision to introduce short notice inspections. **Figure 8** below reveals that just over a third of respondents (36.1%) indicated that they were in use in their workplace at the time of the survey. It would be reasonable to assume, however, that this percentage is likely to rise in the future as more providers adjust their observation schemes to accommodate and reflect the new Ofsted regime.

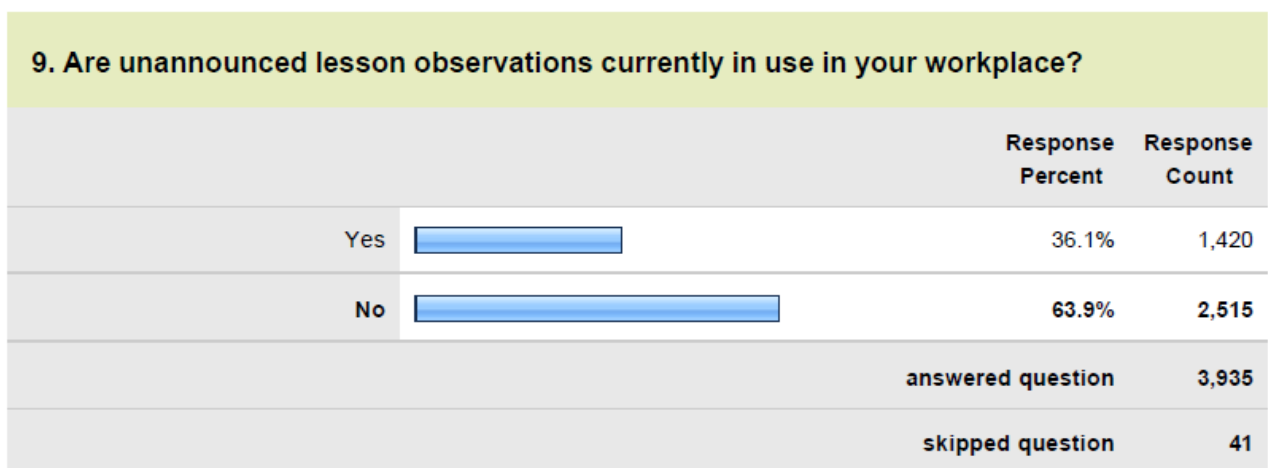


Figure 8 – Use of unannounced lesson observations

Figures 9 and **10** below corresponded to sections B and C of the survey and contained a list of questions relating to graded and ungraded lesson observation respectively (see Appendix 1). Overall, levels of agreement regarding the benefits and effectiveness of graded observation as a method of teacher assessment and

improvement were consistently much lower than disagreement as shown in **Figure 9** below, though there were some noteworthy disparities in cross-tabulations, particularly relating to the responses of senior managers (n = 22), observers (n = 129) and those participants with less than two years' teaching experience (n = 55).

Graded lesson observations

Approximately two fifths (39.7%) of all participants agreed that graded observations were essential for monitoring the quality of teaching and learning compared to just under three fifths (59.5%) who disagreed. As the percentages listed in **Figure 9** below show, this was by far the highest level of agreement recorded for the questions on graded lesson observations across all groups. Nevertheless, it was interesting to note in the cross-tabulations for this particular section on graded observations that participants with less than two years' teaching experience recorded higher levels of agreement overall than their counterparts with more experience². The more experienced practitioners were, the lower the levels of agreement. It is important, of course, to highlight that the number of respondents (n = 55) with less than two years' experience was small in relation to the overall sample. Nevertheless, it raises the question of how such variation might be explained.

One plausible explanation is that participants with less than two years' teaching experience are likely to have had limited exposure to different models of lesson observation and as such may only have experienced graded lesson observations in the workplace. It is not unusual for QA and the monitoring of the quality of teaching and learning to be seen as the only purpose of lesson observation in some instances.

Another explanation as to why there were higher levels of agreement in response to this first question across all groups compared to other questions in Section B might be that as graded observations have become normalised in FE in recent years (O'Leary 2013a), so staff have become increasingly conditioned to expect to be graded on their classroom performance whenever they are observed. Thus, for some, being observed has become synonymous with grading according to the Ofsted scale, particularly for those who have only worked in the sector for the last decade and thus, as argued above, have known nothing else. As Brian, a curriculum coordinator and interviewee in one of the focus groups aptly commented:

We're so used to getting a grade now when we're observed that even if a colleague does something like a peer observation of you, there's a part of you that still wants to know how they'd grade it even though that's not the point.

Nonetheless, two other groups who registered higher levels of agreement in response to the first statement in **Figure 9** were senior managers (n = 22) and

² Please contact UCU directly to request a copy of the cross-tabulations across different variables.

observers (n = 129), with the former registering more than four fifths (81.8%) agreement and the latter just under two thirds (62.8%)³. Although it has to be said that these percentages represented a small fraction of the overall sample size, the disparity between their views and those of practitioners cannot go without comment. It suggests a significant difference in interpretations as to the value attached to graded observations in monitoring the quality of teaching and learning by senior managers, observers and teaching staff. A key reason as to why there was such disparity between the groups relates to the question of what the purpose is of such observations and whose interests are best served by them, a debate which is explored below and further on in this section of the report.

| 10. I believe that graded lesson observations ... | | | | | | |
|---|-----------------------|---------------|----------------------|--------------------------|------------|---------------------|
| | Strongly agree | Agree | Disagree | Strongly disagree | N/A | Rating Count |
| are essential for monitoring the quality of teaching and learning | 7.4% (274) | 32.3% (1,194) | 38.2% (1,412) | 21.3% (788) | 0.7% (24) | 3,692 |
| are essential for improving the quality of teaching and learning | 6.6% (244) | 27.1% (998) | 42.2% (1,556) | 23.5% (868) | 0.7% (24) | 3,688 |
| are essential for the continuing professional development (CPD) of staff | 5.9% (218) | 23.6% (871) | 44.5% (1,644) | 25.4% (938) | 0.6% (21) | 3,692 |
| are the most effective method of assessing staff competence and performance | 1.8% (67) | 12.1% (446) | 46.8% (1,724) | 38.4% (1,414) | 0.9% (34) | 3,685 |
| are a reliable indicator of staff performance | 1.1% (40) | 11.1% (412) | 43.9% (1,626) | 43.0% (1,591) | 0.8% (31) | 3,700 |
| have helped to raise the standards of teaching and learning in my workplace | 2.1% (79) | 17.9% (658) | 44.2% (1,625) | 32.5% (1,195) | 3.3% (121) | 3,678 |
| have helped me to improve as a classroom practitioner | 3.3% (120) | 19.3% (711) | 42.4% (1,564) | 32.4% (1,195) | 2.6% (97) | 3,687 |
| are the fairest way of assessing the competence and performance of staff | 0.9% (32) | 9.7% (357) | 46.7% (1,716) | 41.4% (1,522) | 1.4% (51) | 3,678 |
| are a necessary part of staff appraisal | 2.5% (92) | 23.4% (864) | 42.4% (1,562) | 30.6% (1,129) | 1.1% (40) | 3,687 |
| should no longer be used as a form of assessment | 37.0% (1,365) | 30.4% (1,120) | 23.1% (852) | 7.0% (257) | 2.5% (94) | 3,688 |
| | | | | answered question | | 3,735 |
| | | | | skipped question | | 241 |

Figure 9 – Graded lesson observations

³ Please contact UCU directly to request a copy of the cross-tabulations across different variables.

Two thirds (65.7%) of all respondents disagreed that graded observations were essential for improving the quality of teaching and learning and three quarters (74.8%) disagreed that graded observations had helped them to improve as classroom practitioners. This level of (dis)agreement was similarly reflected in responses to the question relating to whether graded observations had helped to raise the standards of teaching and learning in their workplace. However, levels of agreement amongst senior manager and observer respondents contrasted starkly with practitioners with over four fifths (86.4%) of senior managers and over half of observers (59%) agreeing with the statement that graded observations had helped to raise the standards of teaching and learning in their workplace. This is an interesting contrast in perspectives given that annual graded observations tend to be resource-intensive activities with many providers investing a lot of time and money in them. Such schemes are expected to meet the dual purpose of monitoring and improving the quality of teaching and learning. Yet if there is such a significant level of disagreement across groups as to the benefit of graded observations, one has to question the extent to which they can be seen to satisfy either or indeed both of these purposes. This is a question that will be explored further in the final section of this report, but for now let us turn our attention to the other key finding to emerge from the data presented in **Figure 9**, which concerns the area of teacher assessment.

The highest and most striking levels of disagreement recorded in the responses to the use of graded lesson observations converged around the topic of teacher assessment. Over four fifths (85.2%) of respondents disagreed that graded observations were the most effective method of assessing staff competence and performance. A similarly high level of disagreement was recorded in response to whether they were regarded as a reliable indicator of staff performance. However, the highest level of disagreement (over 88%) of all the questions in this section was the response to whether graded observations were considered the fairest way of assessing the competence and performance of staff. In contrast, only a tenth (10.6%) of all respondents agreed with this statement. When comparing cross tabulations, similarly high levels of disagreement featured in the responses of all practitioners regardless of the length of teaching experience. Over two thirds of observers also disagreed with these two statements, with senior managers being the only group to register greater levels of agreement than disagreement. For example, over three quarters of senior managers (77.2%) considered graded lesson observations to be the fairest way of assessing the competence and performance of staff. This was completely out of sync with all other groups and although the number of senior manager respondents (n = 22) was extremely small, it was, once again, indicative of a significant divide in perceptions between SMTs and the rest of the FE workforce.

It was interesting to note that in response to the final question, two thirds (67.4%) of all respondents agreed that graded lesson observations should no longer be used as a form of teacher assessment, though, once again, there were differences in opinion

between practitioners, senior managers and observers with the latter two strongly in favour of retaining the use of graded observations. Such differences in opinion are not surprising given the importance attached to the collection of this quantitative performance data by senior managers and observers, who themselves invariably occupy senior positions in many institutions.

The lesson observation 'grade profile' (i.e. statistical data sets of how many lessons were graded as a 1, 2, 3 or 4), as it is commonly known, has become custom and practice in FE in recent years and is relied on heavily by senior managers as a key tool with which to measure and compare levels of staff performance internally and against national benchmarks, as part of annual self-assessment exercises and in preparation for external audits such as Ofsted inspections. Thus the compilation and scrutiny of statistical data from annual graded observations is seen as an essential component of the performance management cycle for senior managers in monitoring and assessing the quality of teaching and learning across the institution, despite the scepticism expressed by some senior managers in other studies as to the value of this practice (see O'Leary 2013a). Others have commented on such practice as an example of 'scientific management' or 'neo-Fordism', driven by the agenda for continuous improvement (e.g. Boocock 2013).

Despite the differing views of senior managers, the majority of responses in this section revealed an overwhelming discontent with the use of graded observations for teacher assessment and accountability purposes among practitioners. These views were reinforced in the qualitative data below in **Table 1**, which contains a small sample of randomly selected comments from questionnaire respondents and interviewees, providing a brief glimpse of some of the commonly recurring views to emerge from the project's data concerning graded observations⁴.

These views centred on what are categorised as the 'counterproductive' effects of observation, explored and discussed in more detail later on in this section when grouping the study's qualitative data and the key themes and issues to emerge from that. What is particularly revealing about the excerpts included in **Table 1** below is the adverse effects of graded observations on practitioners, with very sparse evidence of them making any sort of valued contribution to their on-going practice and CPD. Furthermore, despite the positive responses in support of the use of graded observations expressed by senior managers in discussion of the online survey data above, there was a surprising dearth of comments to reinforce this level of support in the qualitative data. Even in those instances where comments were broadly supportive of the use of graded observations, they were consistently accompanied by a conditional statement, as some of the quotes below demonstrate.

⁴ Interviewees are denoted by pseudonyms and questionnaire respondents by identity numbers throughout this report.

Qualitative comments on graded lesson observations

The regime of graded lesson observations is putting unbearable pressure on lecturers. It does not help develop good teaching and learning, which is much better achieved by good on-going CPD and professional development and the opportunity to share good practice in a positive and supportive environment including ungraded peer observation where appropriate (34)

Everybody wants to do a good job; lesson observations, be them graded or ungraded, announced or otherwise, are a key part of the Quality Assurance and a positive experience if the emphasis is on the development of the practitioner and are not used as an excuse for disciplinary or 'capability' procedures (Rose, senior manager)

Everything about my nature tells me that grading is harmful to getting people to reflect on their practice and improve and where we've used non-graded observations they've been quite successful in getting people to focus on what we want them to work on, but whether people have become institutionalised to accepting there's got to be a grade to everything they do means to some people that if we took the grade away they'd say 'well was that outstanding' or 'was that good'? (Sean, senior manager)

I don't see the value in a one-off, one hour graded observation that judges a teacher based on 0.12 % of the work they do (Isabel, senior manager)

The problem is that some teachers 'perform' for graded observations and get better grades than those who produce consistently more interesting lessons (unobserved) but who may not do so well in the graded observation because of 'nerves'. I have witnessed this in my workplace, with the same members of staff getting away with 'poor' quality lessons for many years (181)

I believe that graded lessons can be an excellent opportunity for developing staff but only if the observer is capable of identifying realistic and relevant opportunities for improvement and that they have the skills and experience to coach the observee through the improvement opportunity (199)

Current graded system places undue stress on observee. Seen in many colleges as a management exercise to satisfy external bodies (241)

The problem with any graded observation is that they are essentially non-developmental – because sanctions are attached to the grade, the developmental elements are meaningless (330)

While I feel there is a need for graded lesson observations, I do not feel I have improved as a teacher having been through the process for a number of years. When I ask how to improve my grade, the answer invariably changes with each observer and I am none the wiser (712)

Table 1 – Sample of qualitative comments on graded observations

Ungraded lesson observations

In contrast to **Figure 9** above, the responses to questions about ungraded lesson observations in **Figure 10** below reflected a very different picture. Overall, ungraded observations were viewed more favourably than their graded counterparts, with responses to similar questions registering a much higher level of agreement among practitioners and observers, with senior managers being the only outlier.

Over four fifths of respondents (81.2% and 81.3%) agreed that ungraded lesson observations were a more effective means of improving the quality of teaching and learning and played a more important role in the CPD of staff than graded models. Levels of agreement were slightly lower among observers, though responses still registered at least two thirds. Such high levels of agreement may well be due to the removal of the ‘fear factor’ that is often associated with the high-stakes nature of grading teaching performance, particularly when so much is dependent on the outcome of these annual assessments that capture such a minute part of lecturers’ work, as explored further below.

| 11. I believe that ungraded lesson observations ... | | | | | | |
|--|------------------|------------------|------------------|--------------------------|------------------|--------------|
| | Strongly agree | Agree | Disagree | Strongly disagree | N/A | Rating Count |
| are more effective in improving the quality of teaching and learning than graded lesson observations | 31.0% (1,103) | 50.2% (1,786) | 12.5% (445) | 2.8% (100) | 3.5% (126) | 3,560 |
| play a more important role in the CPD of staff than graded lesson observations | 30.0% (1,064) | 51.3% (1,821) | 12.3% (436) | 2.8% (98) | 3.6% (129) | 3,548 |
| are more effective in assessing staff competence and performance than graded observations | 27.2% (963) | 49.1% (1,738) | 16.3% (578) | 3.3% (117) | 4.1% (147) | 3,543 |
| are a reliable indicator of staff performance | 16.0% (566) | 43.0% (1,520) | 30.5% (1,079) | 6.5% (229) | 4.0% (143) | 3,537 |
| have helped to raise the standards of teaching and learning in my workplace | 11.4% (401) | 33.3% (1,174) | 21.4% (756) | 5.3% (187) | 28.7% (1,012) | 3,530 |
| have helped me to improve as a classroom practitioner | 14.9% (527) | 33.8% (1,195) | 20.3% (717) | 5.4% (192) | 25.6% (906) | 3,537 |
| should replace the use of graded lesson observations | 40.9% (1,450) | 35.7% (1,266) | 13.4% (475) | 4.0% (143) | 6.0% (211) | 3,545 |
| | | | | answered question | | 3,588 |
| | | | | skipped question | | 388 |

Figure 10 – Ungraded lesson observations

Only senior managers disagreed (63.6%) that ungraded lesson observations were more effective than graded ones. Given previous discussion about the importance of collecting quantitative performance data and the reliance of SMTs on the observation grade profile as a measurement tool, such a response is hardly surprising. However, it might also be considered symptomatic of a risk-averse culture amongst SMTs to explore alternative models of observation or assessment for fear of failing to comply with the hegemony of normalised practice in FE. In the following quote, Paul, a senior manager, illustrates this by describing how his college contemplated moving to an ungraded model but with an Ofsted inspection imminent decided against it:

We toyed with a number of models and we had links with another outstanding college and knew they'd decided to scrap grading altogether, although they did this shortly after a successful Ofsted inspection and it's quite interesting when you look at the colleges that do abandon grading, they're almost exclusively colleges who have just been through a successful Ofsted inspection so they're not expecting an inspection team to return for a number of years. We weren't in that position because we were inspected in 2009 so we are expecting to be inspected this year. So we didn't really feel the time was right or it might be advisable to lose grading altogether just before an inspection.

Paul's comments not only reinforced previous arguments about the normalisation of graded observations in FE, but also revealed the apprehension of some providers in choosing to implement alternative and/or ungraded models. Added to this was the conditionality of professional autonomy and how it was linked to inspection performance. For Paul, those recently judged successful were more likely to experiment with new approaches, as 'they're not expecting an inspection team to return for a number of years' and as such were afforded more freedom to do so.

A high number (76.3%) of respondents agreed that ungraded observations were more effective in assessing staff competence and performance, though again senior managers were the only outliers in disagreeing with this. These quantitative findings were reinforced qualitatively in the interviews and focus groups where participants were able to elaborate on some of the key differences between graded, managerial models of observation and ungraded, peer-based models, as discussed below.

Where there was more variance across all groups was in response to statements four, five and six in **Figure 10**. Just under three fifths (59%) agreed that ungraded lesson observations were a reliable indicator of staff performance. This was in sharp contrast to just over a tenth (12%) of respondents who agreed with the same statement in relation to graded observations in **Figure 9** and the two thirds (68%) who disagreed. Where opinion seemed more divided was in the responses to whether ungraded lesson observations had helped to raise the standards of teaching and learning in the workplace and to improve the classroom practice of staff. Just below half of respondents (44.7% and 48.7%) agreed with both of these statements,

compared to a quarter who disagreed. The percentage of 'N/A' responses was noticeable, approximately a quarter of responses for both statements. The main explanation for such a high level of 'N/A' responses is likely to be that ungraded models of observation were not in use in the workplaces of these respondents, thus they were unable to express an opinion on their impact on practice. Finally, it was interesting to note that over three quarters of respondents (76.6%) agreed that ungraded lesson observations should replace graded observations, with senior managers again proving the exception. Furthermore, the high levels of agreement recorded regarding the use of ungraded models of observation were echoed in the qualitative data, as some of the excerpts in **Table 2** below illustrate.

| Qualitative comments on ungraded lesson observations |
|--|
| <i>I can understand that line managers or similar need to check on people's performances but this should take place in a supportive manner, ungraded and learning points picked up/expanded in staff development sessions where everybody could contribute (Steve, tutor)</i> |
| <i>Observations should not be assessed. Instead they should be for the lecturer to receive helpful feedback on their teaching and for self-reflection. All staff should be observed with sufficient notice. Ungraded observations should apply, with only underperforming staff to be identified, and not the draconian 'graded' observation process that currently exists at my college (142)</i> |
| <i>The chance to observe colleagues (ungraded!) has been very useful for me as a first time lecturer (194)</i> |
| <i>I have found the non-judgemental, ungraded approach to observation and feedback incredibly beneficial as observer and observee (336)</i> |
| <i>An ungraded two-way process with sufficient time given to feedback and discuss what has been observed with a direct link to CPD and staff development is what is needed (738)</i> |
| <i>I believe in observations between respected and supportive peers, who know the teaching area, who know the issues amongst the students, and who would provide constructive and supportive feedback to encourage improvement. If we have to have formal observations by managers, then they also need to be properly conducted, with knowledge as above and most definitely ungraded, with constructive feedback (Sarah, curriculum coordinator)</i> |
| <i>The more frequently ungraded observations occur (within reason!), and the more constructively the ensuing feedback is given, the more a culture of embracing peer observation as a tool for professional development without the element of suspicion and fear develops. This can only be a positive thing for students, teachers and colleges (905)</i> |

Table 2 – Sample of qualitative comments on ungraded observations

Lesson observation feedback

| 12. I believe that lesson observation feedback ... | | | | | | |
|--|----------------|---------------|---------------|-------------------|--------------------------|--------------|
| | Strongly agree | Agree | Disagree | Strongly disagree | N/A | Rating Count |
| given during graded observations has identified clear areas of improvement | 5.5% (193) | 31.6% (1,111) | 42.0% (1,476) | 16.4% (576) | 4.6% (162) | 3,518 |
| given during ungraded observations has identified clear areas of improvement | 10.8% (379) | 39.5% (1,382) | 17.4% (608) | 5.8% (202) | 26.5% (925) | 3,498 |
| is well managed in my workplace | 3.1% (108) | 26.8% (937) | 41.5% (1,452) | 25.8% (905) | 2.9% (100) | 3,502 |
| is allocated sufficient time in the observation process in my workplace | 3.5% (122) | 30.3% (1,063) | 41.0% (1,439) | 22.6% (794) | 2.7% (94) | 3,512 |
| is the most important part of the observation process | 25.0% (880) | 46.8% (1,646) | 18.1% (638) | 8.0% (280) | 2.1% (75) | 3,519 |
| | | | | | answered question | 3,549 |
| | | | | | skipped question | 427 |

Figure 11 – Lesson observation feedback

Questions relating to lesson observation feedback in **Figure 11** generated mixed responses with no significantly discernible trends among practitioners, suggesting a range of varied experiences across the sample, with senior managers and observers comprising the two outliers. In comparing the clarity of focus for future areas of improvement, less than two fifths of respondents (37.1%) agreed that clear areas of improvement had been identified during feedback as part of their graded observations, compared to just over half (50.3%) for ungraded observations. This difference in levels of agreement may be due to the grade being seen as an ‘obstacle’, as discussed in section 3 of this report and reiterated in some of the participants’ qualitative comments below, though equally it may reflect the remit and role(s) of observer and observee during feedback in each of these models.

Once again, in the case of ungraded observations, there was a sizeable ‘N/A’ response (26.5%), suggesting that more than a quarter of respondents had no direct experience of them. In contrast to practitioners’ responses, three quarters (73.8%) of observers and over four fifths of senior managers (85%) agreed that clear areas of

improvement had been identified during graded observations, indicating a greater level of confidence in the clarity of feedback and feed forward than observees. Just under three quarters of responses (71.8%) agreed that the feedback stage was the most important part of the observation process. Yet, less than a third (29.9%) agreed that feedback was well managed in their workplace and only a marginally higher percentage (33.8%) stated that sufficient time was allocated to it. Once again, this was not a view shared by senior managers and observers, unsurprisingly perhaps given their leading role in the process. Some of the randomly selected qualitative comments in **Table 3** below add weight to practitioners' views and are developed in later discussion concerning the topic of lesson observation feedback.

| Qualitative comments on lesson observation feedback |
|--|
| <i>In my college the process of observation feedback varies widely. Some of us get feedback within a couple of days, but some wait weeks for feedback. I think feedback given this late would have lost its relevance. It also makes teachers feel that the lesson observation is not taken seriously by some managers. The managers use lesson observations to judge us, but it works both ways. Managers who are slow to give feedback gain a reputation amongst teaching staff for being lazy and we also see it as very bad manners (75)</i> |
| <i>I would like to see a positive model of lesson observation & feedback being introduced that stressed areas of success; acknowledged the teacher's understanding of her class and gave a guarantee of useful follow-up CPD (128)</i> |
| <i>How can people who do not teach give realistic, helpful and formative feedback (133)</i> |
| <i>Feedback is often timetabled way after the observation and squeezed into a short slot when the observer is free. The paperwork is already completed and there is no useful discussion and negotiation for future development (324)</i> |
| <i>Feedback should be informed, constructive and supportive. This is rare in my experience. Often I have been criticized by ignorant and uninformed people who masquerade as managers. When I have received constructive, supportive feedback it has been helpful in informing my progress (500)</i> |
| <i>The focus of the process should be the feedback – feedback sessions should in my mind outlast the observation in terms of duration because two educational professionals are discussing strengths and areas for development and not just looking for the grade! (663)</i> |
| <i>Lesson observations by a non-subject specialist often lead to trivial feedback and misunderstandings. Observations are fine as long as they're formative and therefore not graded but the important thing is to make sure there's enough time for feedback (935)</i> |

Table 3 – Sample of qualitative comments on lesson observation feedback

Unannounced lesson observations

The final section of questions focused on participants' views of unannounced lesson observations. On the whole, there were strong patterns to emerge from the responses presented in **Figure 12** below, many of which pointed to a consensus across groups that unannounced observations were not viewed favourably by FE staff, with senior managers being the only group whose responses were more divided.

Over four fifths (83.2%) of all respondents disagreed that unannounced observations were a welcome addition to the quality improvement process, with opinion almost evenly split among senior managers. An even higher percentage (89.7%) agreed that unannounced observations would lead to increased levels of stress and anxiety amongst staff, a statement with which half of senior managers disagreed. This is a concerning finding given that graded observations have previously been identified as a significant factor in accentuating stress and anxiety among staff in the sector (e.g. Edgington 2013; O'Leary 2011).

| 13. I believe that unannounced lesson observations ... | | | | | | |
|---|----------------------|-------------|----------------------|--------------------------|-----------|--------------|
| | Strongly agree | Agree | Disagree | Strongly disagree | N/A | Rating Count |
| are a welcome addition to the quality improvement process | 3.2% (112) | 11.6% (405) | 33.6% (1,171) | 49.6% (1,731) | 2.0% (71) | 3,490 |
| will lead to increased levels of stress and anxiety amongst staff | 69.8% (2,456) | 19.9% (700) | 5.0% (176) | 4.1% (144) | 1.2% (42) | 3,518 |
| will help to improve the accuracy and reliability of teacher assessment | 4.0% (138) | 16.9% (591) | 36.8% (1,286) | 40.7% (1,421) | 1.5% (54) | 3,490 |
| will help to identify underperforming practitioners more easily | 5.5% (193) | 24.7% (861) | 36.6% (1,278) | 31.5% (1,098) | 1.7% (59) | 3,489 |
| are a positive development in assessing the competence and performance of staff | 3.4% (119) | 12.8% (448) | 38.2% (1,335) | 43.4% (1,515) | 2.1% (74) | 3,491 |
| should be a statutory requirement for assessing teaching and learning in FE | 3.2% (111) | 8.0% (280) | 31.3% (1,091) | 55.6% (1,938) | 1.9% (66) | 3,486 |
| | | | | answered question | | 3,525 |
| | | | | skipped question | | 451 |

Figure 12 – Unannounced lesson observations

The qualitative findings also reinforced some of the unease felt about the introduction of unannounced observations, as the random sample of comments from participants listed in **Table 4** below illustrates and further discussion later on in this section helps to provide more specific detail on how observation impacted on the professional lives of practitioners.

| Qualitative comments on unannounced lesson observations |
|--|
| <i>Unannounced graded observations are more about impressing Ofsted, that the college is following their example rather than about improving teacher performance (74)</i> |
| <i>The introduction of unannounced, ungraded, informal teaching observations in my college has been useful; they are of more use than the graded lessons as the observers see a more realistic version of teaching (when unannounced), and feedback can be constructive and helpful. However the quality of the feedback, and stress caused by it, very much depends on the observer; I have had some very positive, and very negative, experiences of these types of lesson observations with different observers (138)</i> |
| <i>I would welcome someone observing my lesson unannounced if it was part of a supportive confidential process that would make me a better practitioner (Jenny, lecturer)</i> |
| <i>Unannounced graded lesson observations are an appalling idea, an indictment of the system, a violation of collegial trust, and symptomatic of everything that is wrong with UK FE (577)</i> |
| <i>I believe the unannounced observations that we now have put a lot of extra strain on staff who are already overloaded with work and lesson preparation. I do not feel that it is a fair assessment as to the quality of teaching as everyone can have an off day and also badly behaved students (657)</i> |
| <i>I am not against unannounced lesson observations if they remain ungraded. It is clear among most departments in my workplace that peer observation is the most effective at improving performance and showing a better understanding of context (Elizabeth, observer & Head of Department)</i> |
| <i>Unannounced lesson observations are a useful tool to bully staff. The concept of 'learning walks' as it is laughingly called in my institution is a way of maintaining stress levels. As an experienced practitioner of over 25 years, I find it insulting to be snap-inspected by inexperienced management staff, most of whom do no teaching in their role at all! 'Learning Walks' are another farcical example of Ofsted inspired edutainment language (1600)</i> |

Table 4 – Sample of qualitative comments on unannounced observations

As has been discussed previously, Ofsted's shift in policy to short-notice inspections in colleges and no-notice inspections in schools has been the main catalyst for the

introduction of unannounced observations. This policy shift has itself been underpinned by the belief that such an approach is likely to capture a more accurate and realistic picture of practice. Yet when asked whether it would improve the accuracy and reliability of teacher assessment, over three quarters (77.5%) of respondents disagreed compared to a fifth (20.9%) who agreed. In comparison, less than half (41%) of observers agreed and more than two thirds (68.1%) of senior managers agreed with this statement. Over two thirds (68.1%) of all respondents disagreed with the statement that it would help to identify underperforming practitioners more easily, another key reason linked to this policy shift in many institutions. This was in sharp contrast to a similar percentage of senior managers (63.7%) who agreed and just over half of observers (52.5%).

In addition, there were very high levels of disagreement recorded in response to whether unannounced observations should be regarded as a positive step in assessing the competence and performance of staff (81.5% disagreed) and if it should become a statutory requirement for assessing teaching and learning (88.9% disagreed). As much of the qualitative data discussed below highlights, objections to the introduction of unannounced observations centred on debates about notions of professionalism, typically touching on principles of trust, professional autonomy and collegiality and how these jarred with prevailing regimes of surveillance and accountability. This was also evident in some of the qualitative comments in **Table 4** above, where participants seemed to agree that unannounced observations would be viewed more favourably by practitioners if they were part of a supportive and collegial approach rather than an extension of current performative models of observation, which seemed to be the shared experience of many participants whose workplace had introduced unannounced observations.

Key themes to emerge from the project's qualitative data

Figure 13 below illustrates the four main thematic categories under which the key findings from the project's qualitative data were grouped. The use of a pyramid as an illustration is useful as it helps to capture the proportional element of each category. In other words, the largest category – in terms of the most frequently occurring data generated throughout the project relating to that category – appears at the bottom (i.e. counterproductive effects of observation), with the pyramid narrowing up to the smallest/least frequently occurring category at the top (i.e. observation as a formative tool).

The largest proportion of the project's qualitative data converged around the 'counterproductive effects' of observation, highlighting the predominant perception among many of the project's participants that the use of observation in the sector was deemed problematic rather than productive. Admittedly, the headings of each of the four categories in **Figure 13** below are rather broad and all-encompassing, but each thematic category contains a list of specific sub-related themes/issues outlined

in more detail in **Table 5** below. These key themes/issues are summarised in bullet points that form the basis of a series of sub-headings in the remainder of this section, which includes excerpts from interviews, focus groups and textual comments from the questionnaire, along with a summary commentary and discussion.

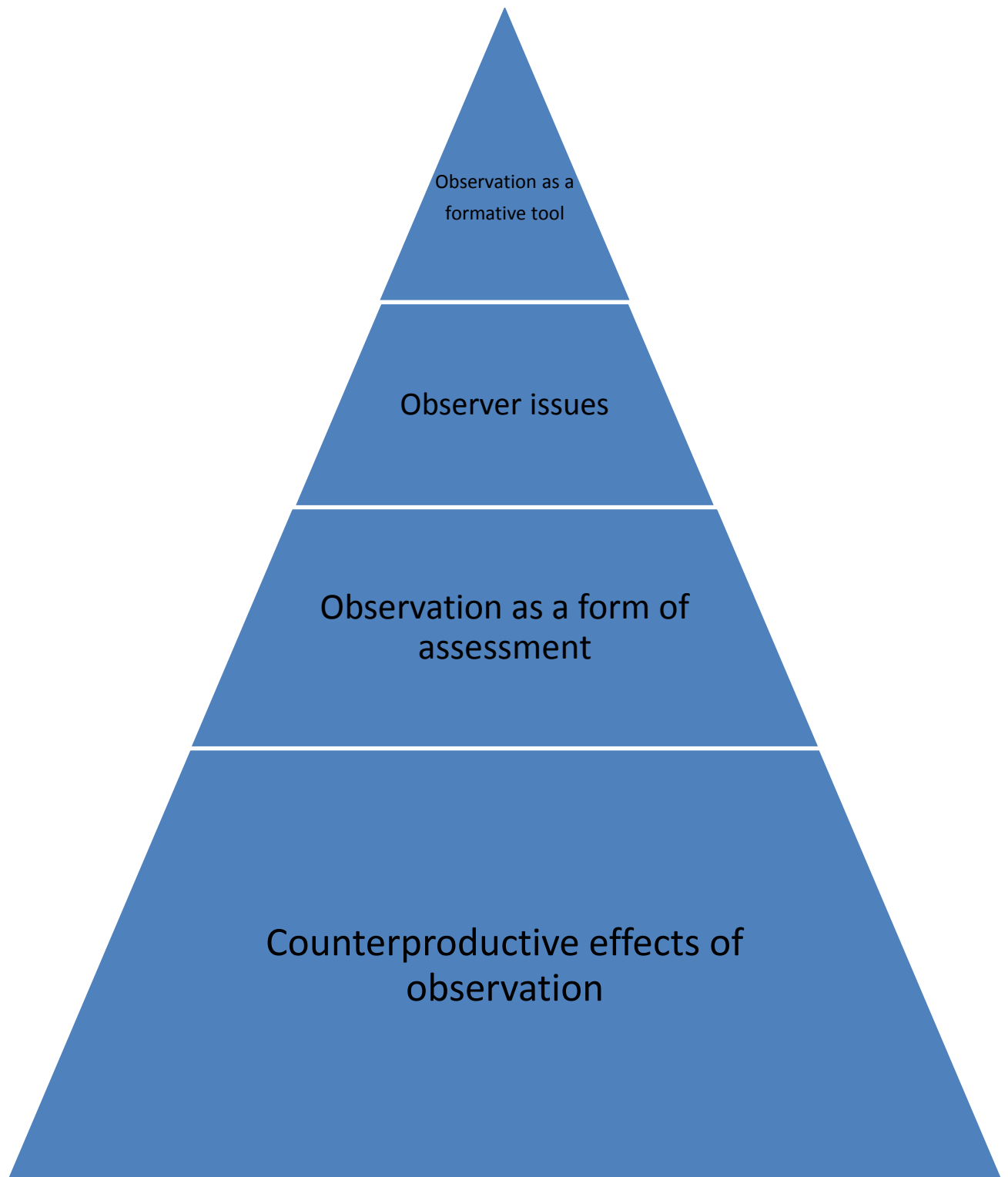


Figure 13 – Key thematic categories of the project’s findings

| Thematic category | Sub-related theme/issue |
|--|--|
| Counterproductive effects of observation | <ul style="list-style-type: none"> • Punitive effect/use of observations (especially graded) are seen as a 'stick' with which to beat staff instead of a tool for CPD/not very helpful or developmental • Graded observations are regarded as 'box-ticking'/'jumping through hoops' exercises • High levels of stress and anxiety caused by current graded observation regimes • Unannounced observations causing increased stress • Too much emphasis on judging and measuring performance rather than concrete support on how to improve Teaching and Learning (T & L) • Lack of trust in professionalism of teaching staff • Time spent preparing for formal, graded observations is incommensurate with the perceived benefits/impact • Focus of observations driven by latest Ofsted priorities rather than genuine interest in excellence in T & L |
| Observation as a form of assessment | <ul style="list-style-type: none"> • Unfair to judge practitioners' capabilities on snapshot observations; they should be more inclusive of other key indicators such as student achievement rates, student evaluations, peer review etc • Concerns regarding the validity and reliability of judgement through lesson observation • Inauthenticity of observations makes them unreliable instruments for judging practitioners' capabilities and identifying underperforming staff |
| Observer issues | <ul style="list-style-type: none"> • Importance of subject specialist observers • Need for observers to demonstrate outstanding, current practice to have professional credibility • Observers need to be fully trained and update their skills continuously • Inconsistency (some good & bad) and subjectivity of observer judgements • Lack of prioritisation and timeliness of the feedback given by some observers |
| Observation as a formative tool | <ul style="list-style-type: none"> • Importance of observation as a 'learning tool' – especially the benefits of ungraded feedback by 'critical friend' • Value of peer observations |

Table 5 – Summary of key themes and issues

Category – Counterproductive effects of observation

| Observations as a 'punitive' rather than a useful tool for CPD |
|--|
| <i>The observation process is used to bully and intimidate members of the teaching staff that are not favoured by the managers (185)</i> |
| <i>In our college two failed observations will link a tutor to the capability disciplinary procedure. Our observers have been encouraged to mark teachers more harshly this year due to poor student results. This will make it easier to identify teachers who could be removed without the need for redundancy pay-outs (223)</i> |
| <i>Poor grades only seek to destroy a teacher's confidence and grading can be a subjective process. The devastation and demoralization teachers feel following a bad observation needs to be investigated further (Sian, lecturer)</i> |
| <i>Lesson observations, where I am employed, are now being used as a way of reducing staff morale and encouraging staff to leave (275)</i> |
| <i>Our college seems to be adopting a punitive rather than developmental attitude to graded observations, made worse by the use of some external observers whose professional competence is questionable, and who have little or no recent experience of FE teaching - a disaster this year (709)</i> |
| <i>The impression is that lesson observations are being used to "weed out" those of us deemed less desirable. So my recent experience of lesson observation has felt punitive, not helpful or developmental in any way (Richard, lecturer)</i> |
| <i>In principle I accept that observations are helpful but the current climate has created a lot of fear and anxiety which is sometimes being exploited unfairly by management. This is why I would currently be against observations being graded as I am not sure that they would be assessed in a fair and objective way (1234)</i> |

Table 6 – Sample of qualitative comments on observation as a punitive tool

In policy terms, many participants talked of links between the outcomes of graded observations and formal capability/disciplinary procedures in their workplace. Data from a document compiled by UCU representatives of the FE London regional committee outlining the observation policies and procedures of the 21 colleges included in the region in December 2012, revealed that in more than half (n = 13), the award of a grade 4 triggered capability procedures. In three colleges this occurred in the event of two successive grade 4 observations and in one college this led to an immediate termination of contract. In professional terms, the detrimental impact of such policies on practitioners' self-esteem and confidence was a noticeable trend to emerge from the qualitative data as some of the comments above indicate.

But not all colleges necessarily adopted a punitive approach to those staff whose performance was judged to be below the required standard in their assessed observations. Lucy, a senior manager, explains how in her workplace, lecturers were encouraged to take advantage of a number of support systems to help them to improve their practice before being re-observed:

We have a procedure for staff that might get a Grade 3 or 4. If they do get a Grade 3 or 4 (and we haven't had many, only about 8% of our staff) I meet them, they have a professional discussion with me and first of all I ask them how they felt about the feedback actually to see whether or not they agree, as that's the first thing we've got to establish and generally their reaction is 'no, I can see what was wrong with that session' and we talk through an action plan with them and it's by mutual agreement of a period of time to do some staff development events and further reflection and I am very keen to pair them up with other lecturers and that they do a peer observation and it helps them improve in areas that were identified as being weak, and then we do a re-observation. So far this year, two thirds of the staff that fell into this group improved their grade. Two went up to Grade 1 when re-observed and the other to Grade 2. There are three teachers who haven't improved and two of them again really by mutual agreement, they have self-selected out and decided that perhaps they're not cut out for teaching and we have directed them into other duties or support roles or more administration work. I've got one teacher who hasn't improved so far and the action plan with him was that he'd been putting off doing any teacher training at all, so finally this galvanised him to go and do his PTLLS, which he completed two weeks ago and also I have been observing him on a frequent basis, formative observations, I'm not grading him, I am trying to work with him on trying to improve those areas of practice.

The approach adopted in Lucy's workplace to dealing with tutors awarded grades 3 and 4 certainly differed from those institutions that adopt a 'restrictive' approach to observation, where 'underperformance' is often regarded as the 'problem' of the individual practitioner. In such circumstances, the onus is on the tutor alone to improve their practice and demonstrate this accordingly when re-observed, with little in the way of professional support provided by the institution to help them achieve this (O'Leary 2013d). In contrast, Lucy's description of how her college went about dealing with such cases seemed to be mindful of striking a balance between being sensitive to and supportive of the developmental needs of lecturers, whilst also ensuring that the quality of the learning experience was not compromised and learners were taught by competent and suitably qualified staff. Lucy's reference to the 'professional discussion' and the importance of encouraging staff to take responsibility for and ownership of their own practice by reflecting on it and subsequently engaging in collegial dialogue was indicative of how her institution seemed committed to a collaborative approach to helping staff develop through observations, but at the same time did not shy away from having frank discussions with those who perhaps were not wholly suited to teaching.

Graded observations are regarded as a 'box-ticking' exercise

It has become a tick box exercise to indicate internal quality by management and a lever to pressurize staff into covering up issues that are often resource based or funding strategies that do not add to a student's learning (144)

Lesson observations in my college are a box ticking exercise by management. They serve no useful purpose in helping me improve my delivery (152)

Box ticking exercise used to justify why not to award annual pay increases (175)

It's a jumping through hoops exercise and the negative comments live with you for years. If this was the way you treated a student by undermining their confidence and questioning their ability you'd be on disciplinary. However, my employer is allowed to do it. Why?(342)

I do not feel that Lesson Observation, as carried out in my workplace, is in any way supportive of staff, and often appears to be more of a tick box (or not!) exercise - it often feels like just another hoop to jump through and it is beginning to feel as though the hoops are getting smaller (Richard, lecturer)

We are subject to never-ending quality claptrap and observation cycles; spending more time planning how to jump through hoops and producing acres of paperwork to satisfy their appetites for meaningless data. A waste of time generally. Rarely helping to support staff and genuinely help them to improve the standards of their teaching (891)

They are a ritualised, tick-box exercise, which often detract from more important work - which is often deferred to type up lessons plans etc - and cause needless stress and anxiety. There must be a better way. For example, why not insist on better training and qualifications in the first place before allowing someone to teach in a classroom? (1314)

Table 7 – Sample of qualitative comments on observation as a 'box-ticking' exercise

The comments in **Table 7** above not only highlight how QA observations have become normalised practice across the sector, but also call into question the value of such practice. As has been discussed elsewhere (e.g. Boocock 2013; O'Leary 2013c), observation schemes are typically intended to fulfil multiple purposes simultaneously, with the monitoring and improvement of standards of teaching being at the forefront of most institutions' schemes. Arguably such schemes might claim to be successful in achieving the former of these two purposes, though as discussion regarding the use of observation as a form of assessment below highlights, this is contested among staff at all levels in the sector. With regards to the matter of improving standards of teaching via formal observation schemes, there was a groundswell of opinion among the project's participants that the impact of such practice was at best negligible and at worst detrimental to the professionalism

of practitioners, as echoed by some of the comments above and in the tables that follow under this category. As Boocock (2013: 488) argued in his study:

Senior managers, in seeking legitimisation from Ofsted, did not regard lecturer professionalism or tacit knowledge as an important ingredient within the OTL process. This, from the lecturer perspective, led to an OTL experience which was neo-Fordist in nature rather than developmental for lecturers observed.

In the following extract, Beverley, a senior manager in a large college, expresses her concerns about the 'tick box approach to observations'. She suggests that this is influenced and exacerbated by the inspection process, particularly the way in which Ofsted's national priorities impact on the focus of observations:

I definitely am concerned about the tick box approach to observations and the hardest thing is when you have an external body like Ofsted come in and say in the new framework their 'flavours of the month' that have become new themes and when they write reports that are critical about the way you address those themes, you are then obliged to check whether they are being addressed in future observations. Now the two for us that came out from our recent inspection were the promotion of equality and diversity and the development of English and Maths skills. So we're having to make that more of a focus in our observations next year so that staff will again have to demonstrate this in their classes. The worry is they will pay lip service to that and not engage with it properly in the spirit that it's meant. So I think the whole danger of the observation system is that it can be that again it's preparing staff to jump through hoops rather than really engaging with the spirit of what makes outstanding teaching, it's giving them another focus that takes their mind off the fundamental part of the job that we want them to do and that's a shame that we have to go along with the system which distracts them from what we really would like to be doing.

In Ofsted's defence, it does not prescribe to providers that the only medium through which its national priorities should be addressed and/or evidenced is that of the lesson observation process, although it stresses that 'observations are key sources of this evidence' (Ofsted 2013: 18). The reality is that lesson observations have become a catch-all, multi-purpose mechanism for many institutions across the sector in preparation for inspection and as part of on-going QA audits, adding further weight to the argument that they predominantly serve the performance management agendas of the institution rather than the professional needs of individual practitioners. As Beverley suggests above, in her workplace one of the repercussions of this focus is that staff come to view observations as something they 'pay lip service to' in order to 'jump through hoops' rather than engage with it as a meaningful opportunity for improving practice.

High levels of stress and anxiety caused by current observation regimes

Having to perform for the observer and try to present, in most cases, an unrealistic lesson that has all the bells, whistles and sparkle that you could shake a stick at causes increased stress and anxiety, reduction in immune system and illness; ultimately time off in the future for many staff (Suneeta, lecturer)

Except for formative observations and informal peer assessments, which have been extremely helpful, other observations have had a negative impact and most of my colleagues would agree put extra stress on lecturers who are teaching too many hours a week and working flat out. Morale is always very low after observation cycles and a recent mock Ofsted left many lecturers devastated and considering leaving the profession including myself (482)

We have been under the 'Ofsted can arrive anytime!!!!' fear-factor since September. The ONLY result of this is stress and our 'Senior Leadership Team' ramping up the 'perfect lesson' deluge of paperwork, flowcharts and buzzwords etc ... This whole issue is a destructive cancer. It's appalling and everyone in my work room of 70+ experienced lecturers is sick to their back teeth. ENOUGH!!! (769)

Lesson observations cause me massive stress. They make me ill & destroy my quality of life, upset my work life balance. I am now considered disabled under the Equality Act 2010 due to the stress inducing regime adopted at the college where I work that caused me to be absent long term. A fear culture has been created, I used to love my job now I hate it (895)

Lesson observations put extremely large amounts of added stress and pressure on staff (Maureen, Head of Department)

In my workplace observations feel threatening and cause stress and anxiety. They are used as part of a scoring process if redundancies are taking place. They do not feel supportive or helpful (1535)

I find the process hugely stressful and while Ofsted are usually only around for three days the internal observations can be any time within a week which is simply unfair (1592)

Table 8 – Sample of qualitative comments on stress and anxiety associated with current observation regimes

The word 'stress' appeared repeatedly in the qualitative data in reference to formal (graded) observations. Taking the textual comments from the questionnaire as an example, over a quarter of the 1619 responses included the word 'stress', often in conjunction with other terms such as 'anxiety' and/or 'pressure'. Significant numbers of participants associated the whole experience of lesson observation, particularly for performance management purposes, with a set of predominantly negative emotions. This was something that was not restricted to the act of being observed

but occurred in the lead-up and post-observation period, and in some cases had more longer-lasting consequences.

Too much emphasis on judging and measuring performance rather than supporting improvements in Teaching and Learning

The way graded observations are more of a 'punishment' rather than incentive for improvement. For example the emphasis is on what tutors do wrong and not on what they do right! There should be a balance (1)

The observer's role needs to be much more supportive than it is at present in my workplace (88)

Lesson observations should be done in a spirit of professional collaboration, rather than being imposed in a judgemental, managerialist framework (512)

As an observer, it is very, very easy to pick apart a lesson (or teacher) and focus on negative aspects of a lesson. This is a situation that can definitely be abused by some line managers. The critical consideration is that, whatever regime is preferred, is implemented within a supportive college culture (Michelle, observer)

There is too much emphasis on lesson observations and the grading thereof in my workplace. There is a significant lack of support for any member of staff who is seen to be 'underperforming' in terms of observation grades. Instead, these people are subjected to more observation and scrutiny (1157)

There is too much obsession with the grade and no worthwhile support following the observation. The problems are pointed out but no clear support path is put in place to overcome these problems. It's a case of 'this is what you're doing wrong, go and sort it out'. If it were that easy I wouldn't be doing it wrong in the first place! (1169)

The process at our college is a deficit process where they wait for you to fail without telling you what the requirements are in the first place. There is no training on the current requirements of a good lesson and to say it's on our web site is not good enough. There is no training on how to improve lessons or how to make them more engaging or interesting (1198)

Table 9 – Sample of qualitative comments on emphasis on judging and measuring rather than supporting improvements in teaching and learning

In some institutions it appeared there was a lack of support as to how practitioners might go about improving their practice following an assessed observation. The emphasis seemed to be on the diagnosis rather than the cure, with the latter sometimes left to the observee to resolve. Prioritising the measurement or judgement of the observation was not unusual, particularly as the 'grade profile', as discussed previously, was often seen as an important source of data for audit and inspection

purposes. So much of the time spent doing observations was consumed with the collation and completion of an accompanying paper trail both on the part of observer and observee, with the result that the time available for professional dialogue between the two was often squeezed. This was not helped by insufficient time being allocated to the post-observation phase in many institutions, or none at all in some cases.

The issue of what happens after the observation and how the developmental needs and areas for improvement of observees are dealt with featured in some of the focus groups and interviews with senior managers. Sandra, a senior manager with overall responsibility for quality and improvement in her college talked about how the follow-up to the observation is crucial for future development to occur:

I think too many people talk about observation in a vacuum. The process is only ever going to be any good if you do anything with the information that comes out of the observations afterwards and how your CPD is set up to support and take forward the things that come out of the observations.

In response to a question regarding what models of observation were in use in his workplace, Sean, a senior manager, explained how the 'feed forward' element of the observation process was a particular conundrum that his college were grappling with:

The predominant model is a graded observation with feedback and our concern I suppose is that there's an action plan that's drawn up following the observation and we're not convinced that the actions are followed through, that there's a wish list of this is what you could do to improve but we want to be sure that some action has been taken to address that.

In a separate interview, Paul, the head of quality and professional development at another college also talked about the need to ensure that reflection remained at the centre of the observation process and how a review of the observation scheme in his workplace highlighted concerns around this:

What we found was that we were still using a model where basically a member of the college observation team (they may or may not by that point be the line manager), they watch an hour of you teaching and then this judgement falls out of heaven and my concern was, as somebody who came into quality management through teacher training, my concern was to what extent is that encouraging teachers to reflect properly on what they're doing and to what extent is that actually provoking professional development.

Concerns about the effectiveness of the observation scheme in Paul’s workplace were initially raised by the vice principal for teaching and learning, who believed that they needed to address the prevailing view amongst many staff at the college that observation was something that was ‘done to them’. Such a view is indicative of ‘top-down’ approaches to observation that are often characterised by hierarchical delineations between observers and observees, resulting in the former controlling the agenda with minimal input and sense of ownership of the process on the part of the latter. Paul’s workplace was intent on transforming the way in which it employed observation into a more inclusive activity that staff felt they had an active role in discussions and decision making. The latter part of this section explores how they went about achieving that in practice.

| The amount of time spent preparing for formal, assessed observations is incommensurate with the benefits/impact |
|--|
| <i>The whole paperwork issue needs to be addressed. I don’t know anyone who doesn’t end up giving massive personal time to preparation for these observations (54)</i> |
| <i>In my institution the current OFSTED model is used for managerial observations where we are told the Thursday morning of obs that could take place any time during the following week. This is more onerous and stressful than OFSTED. We have examples of where lecturers have had to ‘be prepared’ for obs for 5 days of the week. Where obs haven’t taken place until Friday afternoon lecturers are absolutely exhausted and mega stressed! (237)</i> |
| <i>This is still the most stressful part of the job, so much work goes into getting prepared for observation that on-going teaching and learning gets neglected (317)</i> |
| <i>If over the last 5 years my college had put the staff/time/money resources that it has put and continues to put into graded observations, instead into ungraded developmental observations and feedback, I am convinced that teaching and learning standards would have improved far more significantly than they have (653)</i> |
| <i>They are a complete waste of time causing additional unnecessary stress, anxiety and paperwork for staff. They undermine the professionalism of the teaching staff who have too many ‘additional’ tasks to undertake these days outside of the actual teaching (957)</i> |

Table 10 – Sample of qualitative comments on preparation time for observations

The amount of time spent preparing for formal, assessed observations was regarded as incommensurate with its impact on developing the professional knowledge and skills base of those involved. If, for example, as seemed to be the policy in many institutions, staff were notified of the week when internal observations were due to take place, then for many this required them to produce a set of paperwork for each of the classes they were timetabled to teach during that week as they could be

observed at any time. This paperwork often consisted of a detailed lesson plan for the session (and sometimes proof of previous lesson plans), attendance records for the group, a pen portrait of the learners, an updated scheme of work, a sample of learners' assessed work with feedback etc. The preparation of this documentation was a time consuming activity and, as one of the participants commented above, 'more onerous and stressful than Ofsted', as staff were expected to have it ready for all of their classes during observation week. Furthermore, the fact that tutors were required to produce copious amounts of paperwork for classes that might/might not be observed was for some a matter of professional trust.

| Need for more trust in the professionalism of teaching staff |
|---|
| <i>As a lecturer for over 30 years in Adult and FE sectors, I think practitioners are able to self and peer manage performance in a fairer, less stressful way that would enhance teaching and learning and CPD. I also think that experienced teachers/lecturers should be trusted to perform their duties after initial mentoring (384)</i> |
| <i>Although other jobs have appraisal systems to monitor performance etc (as do we as teachers) they don't have the constant scrutiny that teachers have. If the government consider us as "professionals" how come they don't believe we can self-regulate and yet other areas outside teaching can? (416)</i> |
| <i>I hate being observed. 30 years in teaching and still can't be trusted to provide my students with excellent teaching standards. The attitude seems to be 'You are only as good as your last observation'. What other profession requires continual monitoring on this scale? Name one! (501)</i> |
| <i>The constant scrutiny is demoralising and demotivating. We should be left as professionals to get on with our roles instead of jumping through hoops for management and OfSTED (635)</i> |
| <i>Graded lesson observations add to a climate of mistrust and reduce professionalism. In our institution we have peer reviews for HE lessons and graded observations for FE lessons. This leads to an ethos where FE staff feel that their professionalism is not respected (717)</i> |

Table 11 – Sample of qualitative comments on need for trust in the professionalism of teaching staff

The comments in **Table 11** above centred on the notion of professional trust and how current observation regimes seemed to militate against this according to some practitioners. There was the suggestion that the increase in the frequency with which mechanisms of accountability and surveillance such as graded and/or unannounced observations were used 'leads to an ethos where FE staff feel that their professionalism is not respected' and that they could not be trusted to do their jobs without being subjected to 'constant scrutiny'. There were calls for a greater level of

self-regulation, which resonates with aspects of UCU's on-going debate on professionalism in the sector and raises some interesting questions for SMTs, employer bodies such as the Association for Colleges (AoC), the recently formed Education and Training Foundation (ETF), along with other professional bodies and teaching unions to consider.

Summary

Many of the views expressed by participants in this category pointed to how the use of performance-driven models of observation seemed to give rise to a web of interconnected, counterproductive consequences for practitioners. Although there was some evidence of perceived benefits, these came principally from those who occupied senior management positions, highlighting a significant divide between the perceptions of the two and the value attached to this activity by different groups. As discussed previously, for SMTs, the 'grade profile' generated by colleges' annual observation schemes was an important benchmark for both internal and external purposes. Internally, the statistical data produced as a result of graded observations was viewed as an important yardstick for SMTs to compare performance across different curriculum areas within the institution, particularly with a view to identifying those areas that consistently performed well or poorly. Externally, the data were often used by agencies such as Ofsted during inspections to test out their reliability and whether or not the self-assessment systems of the institution could be regarded as rigorous and robust, along with providing an indication of its overall quality of provision.

The next sub-section moves on to discuss the qualitative data relating to the use of observation as a form of assessment.

Category – Observation as a form of assessment

One of the most resounding themes to emerge across multiple data sets and participant groups was how the current reliance on annual graded observations as a means of measuring a practitioner's professional competence was considered an inequitable and reductive practice. There was a collective consensus among lecturers, middle and even some senior managers that such snapshot models of assessment were extremely limited as a source of evidence and that there was a need to look to harness other data sources to supplement them. As some of the randomly sampled qualitative comments in **Table 12** below illustrate, a recurring issue for many was the fact that there are other important sources of data that could be drawn on to inform and supplement evidence of teacher performance gathered during lesson observations (e.g. student achievement rates, student feedback, peer review). However, the annual one-off observation seemed to take precedence above everything else.

Reliance on snapshot observations is unfair in judging professional competence – need for a more fully inclusive model

This is the only industry where you'll be judged on your performance during the 50min observation and not the final outcome i.e. the students' achievement/success, which is the final result and hence a better indicator of your sustained efforts over 800 hrs (47)

A mixture of responses should be used: 1) Observations that are ungraded; 2) Focus groups - ask the students how the staff are doing - they are going to be honest! 3) Survey - again ask the students; 4) Walk-throughs - where managers pop in but not to check paperwork, to see how the class is running and if students are happy; 5) Student achievement - surely this is an important issue! (Leanne, senior manager)

One-off observations are something I dread, you have one chance in one lesson in which to impress your manager and prove your worth. If something goes wrong, learners misbehave or there is a technology fault, that's it, you're graded for the entire year based on that particular session. Other factors should be taken into account when deciding if a practitioner is underperforming, face-to-face teaching is only a fraction of the job (168)

Graded lesson observations should NOT be used as the only method of assessing staff. Other indicators are appraisal sessions, student results, staff and student feedback (233)

Some teachers are more prone to "exam anxiety" than others. These teachers will continue to suffer from lower grades as long as observations form the main part of teacher assessment. Why not combine observations with assessed discussions, file audits and conversations with students outside the classroom etc to see what else teachers are capable of? (Jacqui, lecturer)

Current lesson observations do not give a true indication of a teacher's performance, it should be an observation over a set period of time and feedback should be more constructive. Tutors should be observed not only in the classroom but also in their other roles and responsibilities, their contribution to departments, success rates etc. Consideration should also be taken into account of the type of student groups being managed at the time of observation (801)

The lesson observation currently only takes a snapshot of a person's performance and this is not a true indication of the teaching standards, considering a lesson observation is 1-2 hours out of 828 hrs in a full time post (802)

Table 12 – Sample of qualitative comments on unfair reliance on observations to judge professional competence

Richard, a middle manager working in the quality department in his college, put the case forward for a broader model of observation in the following interview excerpt:

There are others ways to performance manage staff other than attaching a number to a lesson observation ... There's a wider picture of teaching other than the 50 minute

session. I mean the best observers will pick up on what went on last week and how it impacted on this week. I have just done one now and it's a broader look at learning. I think we should have shorter ungraded observations, more often with less paperwork attached with a focus on the key priorities. I think the improvement side should be done by peer observations and thirdly we should be looking at the broader parts of the teacher's delivery of learning and not just focusing on what happens in the observations but the other things they do on a daily basis.

In referring to the 'wider picture of teaching' and the 'broader parts of the teacher's delivery', Richard sought to acknowledge the breadth of roles and responsibilities undertaken by tutors and how current performative models of observation not only failed to capture that breadth but were, by definition, reductive in nature and thus could only ever provide a limited snapshot of practice.

As with the statistical data, what emerged from the qualitative data was an overwhelming discontent amongst UCU practitioners regarding the reliance on annual graded observations as the main or even sole form of evidence for assessing teacher competence and performance. As some of the comments in **Table 12** above indicate, annual observations provided only a minuscule insight into practice. Other important sources of evidence needed to be incorporated into the assessment process to make a more accurate, rounded judgement. **Figure 14** below attempts to capture some of these other sources of evidence as part of a 'multi-dimensional model'.



Figure 14 – Multi-dimensional model of teacher appraisal

How feasible the operationalization of such a multi-dimensional model of teacher appraisal might be remains unclear at present and would no doubt benefit from further research. What was clear from participants' comments, however, was the need to move away from a system that relied on a narrow evidence base.

| Concerns about the validity and reliability of judgement through lesson observations |
|--|
| <i>They are so subjective, and hence are most unproductive, particularly to young enthusiastic lecturers. One observer may award a higher grade, whereas another one would give a lower grade, despite the delivery and the lesson plan being very similar (92)</i> |
| <i>I recently had an unannounced obs and was graded a 4 by my team leader who I feel has strong personal issues with me, which therefore reflected on my grade. I totally challenged this with credible evidence and was given a totally different feedback attitude, which backtracked on the original written evidence (132)</i> |
| <i>As an observer, there appears to be no agreed standard that all observers work to which makes the whole process a complete lottery. Some observers start from a default of 4 and you have to impress them in order to work your way up the grading scale while others seem to start in the middle and need less impressing in order to achieve an acceptable grade. These evident handicaps need addressing to get a truly equal observation system (Gerry, lecturer)</i> |
| <i>Many of our internal observers are seeking career advancement and assess against a tick list derived from their idea of OFSTED criteria – starters, game activity, ICT, plenary etc (Teach by numbers we call it!) I was graded 4 (lowest) when I proceeded with an information-giving session for access students. I was re-observed with a 1/2 grade and invited to join the teacher/expert team or whatever they call it 'rubbish to excellent within 3 weeks'! (567)</i> |
| <i>I have been graded a 1 and a 3 for the same lesson by different observers (1026)</i> |
| <i>Observations are too subjective to be of any real use. Staff can be unfairly targeted by management with scores to settle. if colleges wish to practice continuous improvement they need to adopt the full industry model and put time and effort into training and support and loose the culture of blame and failure (1434)</i> |
| <i>I resent being expected to perform like a trained chimpanzee to a given criteria and then being graded subjectively on this one-off occasion. Competence in the classroom is not about a performance that matches specific criteria, it is about the quality of the output, achievement and success of the students over the whole year not just one isolated occasion when the teacher often does a 'textbook' lesson engineered specifically for the observation which is often totally out of context and bears no real reflection on what happens the rest of the year (1478)</i> |

Table 13 – Sample of qualitative comments on concerns about the validity and reliability of judgement

Validity and reliability are fundamental concepts in any discussion about the use of classroom observation as a method of assessment. Validity is generally concerned with the extent to which an assessment covers the knowledge, skills, attitudes etc that it claims to measure. To put it simply, 'does the assessment measure what it says it measures?' Validity is inextricably linked with reliability which is concerned with the consistency and accuracy of results. As Black (1998) says, 'good quality assessment is inevitably the child of a union between reliability and validity' (p. 54).

There were two issues in particular to emerge from the qualitative data that seemed to call into question the validity and reliability of judgements made during observations. The first and most commonly occurring issue centred on the subjectivity of observer judgements. As some of the comments in **Table 13** above suggest, there were inconsistencies in the interpretation and application of the assessment criteria by some observers, as evidenced by variations in the awarding of grades, even when observing the same lesson in one case. This is nothing new in the field of lesson observation and highlights some of the challenges faced by observers in trying to ensure that their judgements are as valid and reliable as possible. The following two excerpts from interviews/focus groups help to provide a brief insight from the observer's perspective:

There are an awful lot of times where you sit in a lesson and you're observing somebody's teaching and you're not quite sure where it sits and whether it's a Grade 2 or 3 so you watch a bit more to see if it gets a bit better, but then you have so much to juggle in your head before you arrive at that final decision. It's a lot of pressure you know because if you give it a 3 then that has particular consequences for that member of staff who, let's face it, is still a colleague (Trisha, observer).

We had a very interesting discussion in one of the observer training courses I did last year between two observers who were arguing about the amount of weight that should be attached to some issues connected with equality and diversity and one observer clearly attached a great deal of weight to this, so much so that she was prepared to grade this particular session a grade below. So they do attach different weights to things (Paul, senior manager).

Observer judgement was a complex and thorny issue that presented challenges for individuals and institutions alike. Although many institutions engaged in on-going standardisation and training for observers, this did not necessarily mean the issue was any less problematic. Some of the other confounding variables are included in the sub-section entitled 'observer issues' that follows this one.

The second issue related to the performative element and how this impacted on what practitioners did, especially given the high-stakes nature of these observations in some institutions. This is explored in the discussion following **Table 14** below.

Inauthenticity of graded observations makes them unreliable instruments

I now always achieve a 'grade 1' observation. This is more due to having cracked the 'code' of how to produce a lesson plan and a lesson which fits the Ofsted 'boxes' and I find that I have the ability to be able to 'switch on' and 'perform' whilst being watched (84)

No one performs at their norm during these observations; I have carried out peer observations and seen great teaching and learning, only to hear this teacher is graded a 3 or 4 on a graded observation. Other times I have witnessed appalling teaching and learning, but this teacher 'pulled out all the stops' for a graded observation and is hailed a mentor! (237)

Graded observations cannot even be said to provide snapshots of a practitioners daily methods since most people under such observation will be delivering a contrived lesson lacking in spontaneity and flexibility (255)

Some staff only perform when graded lesson observations take place and are very good at it. Others who perform generally well in class can go to pieces during observations. This can give an unreliable opinion of these tutors. There are a lot of good practices that go on throughout the year and these go unnoticed (281)

They are a huge pain in the backside, distorting all activity for many days as you try and "perform" in a totally artificial way - bit like a driving test (430)

Staff being observed either plan an uncharacteristic lesson or become anxious so do not perform as they would normally. This is a major issue that has never been properly addressed (437)

A single lesson observed is never an indicator of what actually takes place in the classroom as teachers can put on a show for observation and revert to their bad ways once the observation is over. There is a teacher in our college who is constantly complained about by students, yet during and inspection he was commended by the inspector (525)

Table 14 – Sample of qualitative comments on inauthenticity of observations

Table 14 above captures comments relating to the 'inauthenticity' of graded observations and the way in which the performative element can distort the observed lesson, what observees do and how they can behave under such conditions. This is commonly known as the 'Hawthorne effect'.

The Hawthorne effect is a psychological term used to describe the extent to which the observed environment is influenced by the observer's presence. In other words, to what extent is a teacher's performance or behaviour in the classroom affected, consciously or not, by being observed? It can be argued that teachers' behaviour is affected by the mere act of being observed and this can have both positive and

negative repercussions for some. In an excerpt from an interview with Sean, a senior manager, he provides us with a situated example from someone with over twenty years' experience as an observer of the ways in which the Hawthorne effect can manifest itself:

It's amazing the amount of teachers you observe who the students say later, 'Oh, you were different later when the observer came in' and you're suspicious in a sense. That makes you think, 'Have they done something differently or tried something they normally wouldn't do, or have they put more effort into it?' But nine times out of ten it's that the pressure of being observed has caused the teacher's normal personality to change to accommodate the observer ... part of their natural rapport and charisma that they have, they stifle that and the students miss it you know, that spark in their personality perhaps wasn't there because they were self-conscious about being observed, it can be either way ... The natural relaxed working atmosphere that normally happens, it's very difficult to replicate that in a stressful situation like an observation.

The 'pressure' that Sean refers to is all too apparent for many practitioners in annual graded observations where the stakes are high and the outcome becomes all the more important, thus encouraging the observee to try even harder than they might under 'normal circumstances' and produce a performance that belies their everyday practice. This is often referred to as the 'showcase' or 'all singing and dancing' lesson alluded to by some of the comments in **Table 14**, and there is the suggestion that it is not an effective strategy to identify underperforming or 'inadequate' tutors, as some will rise to the occasion accordingly. Equally, some consistently good tutors find the pressure too much and uncharacteristically underperform. This is a good example of how graded observations can be considered the epitome of Foucault's (1977) 'examination', which serves as 'a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish. It establishes over individuals a visibility through which one differentiates and judges them' (p. 184).

Summary

This sub-section has focused on the key issues relating to the use of observation as a method of assessment. The reliance of the FE sector on graded lesson observation as the main or even sole source of evidence on which to base judgements about professional competence and performance has been called into question. Many of the concerns expressed converged around the issues of validity and reliability of assessment, with participants emphasising the need to move towards a more fully-inclusive and multi-dimensional model.

Category – Observer issues

| Importance of subject specialist & fully trained observers |
|--|
| <p><i>I believe that observations can be an excellent opportunity for developing staff but only if the observer is capable of identifying realistic and relevant opportunities for improvement and that they have the skills and experience to coach the observee through the improvement opportunity. Furthermore, where graded observations are conducted by a number of observers within a college, all observers should have gone through a robust standardisation and training process; something that is distinctly lacking in my experience (199)</i></p> |
| <p><i>There should be a recognised qualification and a decent length of teaching experience for individuals to achieve before they are allowed to conduct teacher observations. Simply being a line manager and doing a short employer organised training (on the required paperwork!) should not be acceptable! (204)</i></p> |
| <p><i>The observers would have more respect if they taught the same subject as the observee and were still teaching in class 23 hours a week! (308)</i></p> |
| <p><i>The main problem with graded observations in my area is that they have been performed by staff who have neither the experience nor training to give credible feedback, and with observers who can't provide effective feedback on the skills that less experienced and qualified teachers lack, they are unable to improve (314)</i></p> |
| <p><i>Lesson observations are a necessary part of quality processes within the modern teaching environment. However, it is important to understand that not all subjects fall into the mould and style of teaching that perhaps is understood by the traditional or academic subjects. It is therefore important that those managers who are tasked with the observation and assessment of their colleagues are fully aware of the needs and requirements of the subjects they observe (389)</i></p> |
| <p><i>The quality of any observation scheme is dependent on the quality of the observers. People doing the observing should be suitably qualified to do the job and should be a subject specialist in the subject being observed. I have had bad experiences of an observer with no subject specialist knowledge telling me (the subject specialist) how to teach my subject that they know little or nothing about (412)</i></p> |
| <p><i>I think the observer should be fully trained in classroom observation practice and hold a relevant qualification for this role (457)</i></p> |

Table 15 – Sample of qualitative comments on importance of subject specialist & fully trained observers

Of the key issues to emerge from the qualitative data about observers, the first concerned the extent to which they were appropriately trained to carry out their role and the second whether they had experience and/or knowledge of the subject areas they were responsible for observing.

The importance of the observer having an understanding of the subject specialism was something that applied across all curriculum areas, but seemed a particularly contentious matter in vocational subjects that had links to industry such as engineering. Julia, a director of learning, put forward the argument in the following extract that observation systems that used a 'one-size fits all' approach to assessment often disadvantaged those tutors working in vocational areas:

The engineering workshops I have been in, I have seen authentic work and students with employer engagements or coming in and working on real stuff and state-of-the-art equipment, but when an observer comes in they don't recognise those authentic skills of the students, they just wonder why tutors aren't questioning them and why they haven't got a three part lesson plan. So they get graded down and then that de-skills and demotivates the engineering tutor who then hot foots it back to industry, so you get a cycle of that engineering department going down and down whereas somebody could actually go in a little bit more sophisticated and say 'What are you students actually doing?' and getting the teacher to come up with criteria that they want to be judged against and if the teacher says 'these are my successes and indicators and this is what I want you to come in and look at'.

Julia's comments raise important questions regarding the appropriateness of using a uniform set of assessment criteria for lesson observations, regardless of the subject area and the contexts in which these subjects are taught and learnt. In other words, should observation assessment criteria be differentiated to reflect different curriculum areas, learning contexts and cultures instead of using a one-size fits all model? This was an issue that provoked debate among some participants, with vocational lecturers, in particular, of the view that assessment frameworks for observations were designed primarily for 'traditional' and/or 'academic' subjects taught in conventional classrooms rather than those that relied more heavily on work-based learning/practice-based workshops.

The second issue to emerge from **Table 15** above concerned the training provided for observers in preparation for their role. The perception of some participants was that this needed to be more standardised and even formalised, with some suggesting that observers should hold a recognised qualification. Given previous discussion regarding the validity and reliability of observer judgements, this would seem like a concrete and sensible suggestion to take forward. A senior manager from one college involved in interviews/focus groups talked about how all the observers in her workplace were required to undertake a specialist programme of study with a local

university before embarking on their role. Once they had completed this, they were then expected to attend on-going briefings and meetings:

We've always had at the beginning of the academic year observation briefing, training and updating ... and then what I've tended to do is along the lines of much more differentiated CPD is to put on sessions for writing reports and getting feedback, so that is for all observers, but what I was interested in was we needed something midway to say, 'Hey, how's it all going with everyone' with all the observers together. So that wasn't really training it was kind of coming together and 'Here are the things I have picked up' and I am quite close to the teachers because I did observations myself so I try to make sure I am aware of what's on the ground but equally I know all of it.

Need for observers to demonstrate outstanding practice & still teach to have professional credibility

Lesson observations can be useful if carried out consistently and the observer is clearly more advanced in teaching practice than the observee and more qualified. The observer has to earn respect from the people they observe professionally not just because they got the job (69)

I have no confidence in the ability of assessors from management who generally do a minimum of teaching themselves nor of the validity of any comments they make on my performance (208)

The quality and usefulness of the observation depends heavily on the competency of the observer; such competency is often lacking (210)

The internal management that are let loose over 2 days to observe any teacher they choose or feel to, have often not 'taught' a single lesson in their managerial career which questions the competence of them as a judge to a range of teachers (218)

All too often the observers are people who couldn't wait to stop teaching. Ask them to show you how to change something instead of telling you it's wrong and they can't do it! (308)

Lesson observations are being undertaken by staff who have NEVER taught...This provides an issue of credibility for the observer and makes any criticism of lessons difficult to accept for experienced observees (811)

I feel that observations can be really helpful in developing skills provided they are done by someone who you respect who has excellent teaching skills themselves. When they are done by managers who have little understanding of teaching they are meaningless (1045)

Table 16 – Sample of qualitative comments on professional credibility of observers

A lot of lecturers commented that in order for observers to maintain professional credibility among their peers, they should still be actively teaching and be able to demonstrate excellence in their own classroom practice. The fact that many observers occupied management roles meant there was less likelihood of them still teaching, which only served to reinforce the assumptions of some practitioners that they were out of touch with current practice and unaware of the challenges and difficulties faced by lecturers.

Category – Observation as a formative tool

| Importance of observation as a tool for professional learning |
|---|
| <i>I agree lesson observations with feedback, are essential for helping new staff achieve competence and confidence in teaching. There will always be new styles and things to include that become popular that an older teacher may not use and that would be beneficial for their teaching (446)</i> |
| <i>Observations have a useful place in helping teachers improve, but at the moment they are a tool which resembles more a weapon in management's hands, a weapon that causes teachers a great deal of unnecessary stress, than something genuinely helpful and constructive (563)</i> |
| <i>Lesson observations are useful, but only when they are used to develop teachers with useful and informative feedback rather than place them in a box with a grade that does not say very much or help them (787)</i> |
| <i>I'm all for observations to help develop practitioners with their skills set if done in a positive and constructive environment. Normally observations are a very stressful and sometimes demotivating for staff. I have been very fortunate with my observations, receiving constructive feedback that has helped me develop as a lecturer (862)</i> |
| <i>I think feedback from lesson observations is helpful if it is constructive. I think a few more informal lesson observations would be better than one formal graded one (1088)</i> |
| <i>The most useful observation system I have found is supportive in nature where the focus is on the feedback and development (1362)</i> |
| <i>Lesson observation would be useful if; 1) they were seen as a developmental and supportive; 2) they were ungraded and that teaching staff felt that they could go to observers if they were having problems within a class, without repercussions relating to their ability and job security; 3) were in the hands of teachers not managers (1505)</i> |

Table 17 - Sample of qualitative comments on importance of observation as a tool for professional learning

A unifying theme among the comments in **Table 17** above was how instrumental observation could be in helping to encourage professional learning. Here the emphasis was placed on the importance of feedback being 'constructive' and 'supportive'. In a focus group of observers, Neelam made reference to how the grade could be an obstacle to constructive dialogue between observer and observee:

Neelam: *As an observer, it would be easier if you didn't have to grade as you would have a better dialogue without it.*

Researcher: *Is that something which other observers find to be the case?*

Group of 5 observers: (Collective response in unison) *Oh, absolutely!*

The notion that the summative element can impinge on the formative feedback is well documented in the field of assessment. In the context of observations, the general claim is that observees' engagement with the developmental feedback is compromised as the grade forms a barrier between observer and observee. The grade can take on such importance that it threatens to undermine the value of feedback and the professional dialogue, hence the suggestion by Neelam and her observer colleagues that not having to grade would result in 'a better dialogue'.

The final set of comments included in **Table 18** below related to the value that many practitioners seemed to attach to peer observation. Although peer observation was commonplace in participating institutions, it was still a relatively marginalised practice and invariably occurred in the ITE context. In some workplaces it operated on an informal and voluntary basis, whereas in others it was embedded into formal support systems within the institution. As some of the comments in **Table 18** illustrate, not all practitioners were engaged in peer observation in their workplace, but many expressed an interest in having the opportunity to become involved given the reciprocal benefits associated with this model of observation.

The reciprocal nature of the relationship between observer and observee in peer observation offers the potential to redress some of the power imbalance commonly associated with top-down, performance-driven models discussed earlier on in this section and lay the foundations for encouraging a genuine 'sharing of practice' among peers, as Trisha, an experienced observer suggests below:

I don't know the actual ins and outs of it but the value is surely twofold that if you've got two peers, peers watching each other, not only am I getting something out of watching you I might even see something I don't like, but at the same time I might think 'I can do that' and then also when I feedback to you I can say 'this was good' increasing your confidence or secondly 'I do this in my lessons why don't you try that' so you've got two people who are improving from that one process.

Value of peer observation

In a collegiate atmosphere ungraded peer observation is the way forward, where honesty is fostered because there is no implicit threat to the process. It's about sharing best practice and thus improving standards, rather than raising the threat of a poor grade. Teachers do not perform at their best in such circumstances, thus graded or unannounced observation cannot give a true picture of that teacher's capabilities (Suzanne, lecturer)

I think that observations have their place in our teaching practice, but it is stressful, we have a lot of work to do and this does not make it easier. Peer assessments on the other hand is more useful and you learn from because you are not under the same amount of stress (921)

Would welcome peer assessment/honest feedback from peers who have no axe to grind or whose own performance hinges on what they say to me. I find my colleagues' input the most instructive I have had to date and would welcome more opportunities to be involved in this. I learn so much from observing others myself, and if I agree to cover for a colleague will, whenever possible, the class prior to taking over, to take advantage of this opportunity. Not only for a seamless continuity, but also because I have previously learnt so much in this capacity about both the groups and students I am taking on but also, my own teaching and how things can work differently, or better, if approached from a different angle (1013)

I use peer review and it is an excellent method of identifying areas for development which can then feed the appraisal/CPD process. The absence of grading removes a lot of anxiety and you see the lecturer operating in the context in which they would operate in any session. I would like to see a combination of announced (internal school/department) and unannounced (cross-school/department) peer review sessions being used so that a better picture ascertained as opposed to a snapshot but retaining the ungraded element (Peter, observer)

Wouldn't it be amazing to work somewhere with a culture of peer observation where we all felt safe enough to be observed and to observe. How much better might our teaching be if that was our workplace culture rather than the one of fear of a low grade from our line manager as at present (1114)

I'd like to have an opportunity to observe and be observed by peers to allow us to swap good practice and support each other. For example, teaching a group I teach but struggle with. (1169)

Ungraded peer observation is used in my HE post which is much more developmental, professional, dialogic and generally more useful. If we wish practitioners to reflect and develop, I believe ungraded observations with detailed feedback are the best way (1322)

Table 18 – Sample of qualitative comments on value of peer observation

It is interesting to note the discourse used in some of the comments above to describe/refer to peer observation and how this was indicative of a very different

attitude among practitioners regarding this model of observation compared to its performative counterparts. Practitioners talked about it being 'less stressful' and feeling 'safe enough to be observed and to observe', as the emphasis was on 'sharing best practice' and 'learning from observing others'. In addition, peer observation for assessment was regarded as less contrived and thus a more authentic experience. This last point could have significant repercussions for future uses of observation as a form of assessment as it would suggest that the low-stakes nature of peer observations means that they are more conducive to capturing reliable snapshots of practice than high-stakes graded observations.

The benefits of peer observation and its potential to act as a springboard for substantive professional dialogue are neatly captured in the extracts from interviews with Barry and Vera, two experienced observers below:

The best I think we can achieve is to uncover more about good practice than we currently do by providing more opportunities for teachers to see each other at work in their classrooms and to talk to one another about the practices they use, the experiences they have had and the theories and values that underpin their work (Barry, observer).

The most effective staff development happens when teachers talk to and work with other teachers, not managers or observers, but peers. Those of us who observe know how much we gain and learn from watching others teach, and I really do think that the best model is one where the power balance is equal and both partners in the process share common aims (Vera, observer).

The final part of this section moves on to looking at two brief vignettes of other models of observation in use in two of the colleges featured in the study as an example of how there are alternatives to normalised models that currently dominate practice across the sector.

Alternative models of observation

The case of 'dual grading' at Sunnyside College

Sunnyside College was rated 'outstanding' during its last inspection in 2009. Although it had plans to introduce an ungraded model of observation, it was expecting an Ofsted inspection during 2013 and thus the decision was taken to postpone its introduction until after the inspection had taken place. As an interim measure, Sunnyside chose to implement a model of dual assessment on the part of observer and observee. What this meant in practice was that the observation grade would be agreed jointly between the two, but only after a 'professional discussion' in which each had shared their reflections on the observed lesson. In the case of disagreements, the feedback and self-evaluations were sent to the head of the quality and professional development unit, who acted as a mediator. In cases where it was

impossible to reach an agreement then the observee was entitled to be re-observed by a different observer. Paul, a senior manager, summarises this model of dual grading below:

We created an observation format where the teacher now self-assesses as part of the overall process, so what we have basically is both the observer and the teacher write a reflective report, change the name of the observation form to a reflective report and we wanted to emphasise the fact that this model was placing emphasis on genuine reflection and opportunities for staff development, we get both the teacher and the observer to write a report and they both propose a grade.

Paul went on to explain how the creation of this new model was seen as a 'big risk', as it could have resulted in a flood of disagreements and disputes between observer and observee. Yet, as he points out, it turned out to be a worthwhile risk:

The system we've now got is that both teachers and observers make a report of their own and they both propose a grade. They then meet for something called 'professional discussion' and that is where the teacher will put forward their feeling about the session, what worked and what didn't work and why they would propose a particular grade, and the observer will do likewise and put forward what they felt worked and what didn't and propose a grade. The big risk here, of course, was if we were overwhelmed with disputes about grades, it could make the system very unworkable and we would have ended up rather than the observation system supporting genuine professional development, we could have ended up with something that was a potential industrial relations problem. The really, really interesting thing was that we very quickly established, as we were into the first few weeks of it, I noticed we weren't getting disputes, we were getting a very, very high consensus between observers and teachers about how the teaching, learning and assessment framework should be interpreted and used. The number of actual 'disputes' that we've had, we've done 318 observations I think in the current year and the number of disputes we've had, and by 'dispute' I mean I have been contacted either by the observer or the teacher with an email saying we've had the professional discussion and we can't reach an agreement about the grade, you can count on the fingers of one hand. I mean it's been literally something like 1% if that.

Instead of triggering confrontation and disputes between observer and observee, this model of dual assessment seemed to foster increased collaboration and a greater sense of shared engagement on the part of both with the college's observation assessment criteria.

The model introduced by Paul seemed to be underpinned by some of the core attributes associated with 'expansive approaches' to the use of observation (O'Leary 2013d). For example, the decision to empower observees to grade themselves and for

them to then discuss their self-assessment with their observer with a view to reaching a joint judgement seemed like a genuine attempt to reduce the power differential between observer and observee and make it a more collaborative process, whilst simultaneously creating a greater sense of ownership and involvement in the observation process for the observee. In the final excerpt, Paul explains the rationale for this new approach and its intended outcomes:

We have tried to develop a system that puts the emphasis on articulating and working out the meaning we attach to professional judgement and what teachers see in their own work and what observers see in that work and try to marry that up with meeting the needs of the learner and what the college's expectations are etc ... and maybe for some reason we have been able to get the mix right.

Arguably one of the key factors contributing to the success of this new approach was its attempt to be genuinely inclusive, which, according to Paul, increased its legitimacy for practitioners and made them more likely to embrace it. Some might also wonder why the college did not remove the graded element altogether, though the expectation of an imminent Ofsted inspection was the main reason for this, as discussed earlier. Putting the grade to one side, Paul's final comment emphasised how the move to a more inclusive approach had heralded a more candid and open interaction between observer and observee, acknowledging that the task of improving teaching and learning was a collaborative and reciprocal one that did not rest solely on the knowledge and interpretation of the observer, but required a collective and collegial response:

They appreciate being involved in the grading decision, that's the phrase that's been used to me and coming back to the comment the VP made last summer about it's something that's done to them, we have perhaps been able to work up something where now they do feel it's more inclusive ... Another comment that was made was acknowledging that observers don't necessarily have all the answers, a lot of our teachers are dealing with very challenging groups and extremely difficult learners working under quite difficult situations with an ever shrinking unit of resource to support that and we don't necessarily have all the answers, but what we want is to provide support to facilitate improvements with that but we all have to work together to solve these problems. In our observation system I'm not expecting observers to come out with all the answers. I want them to facilitate the situation where the teacher comes up with all the answers.

This emphasis on the importance of collaboration and using observation as a catalyst for wider professional dialogue also reinforces the idea that observation is more widely accepted by teachers when it is used as a collaborative tool rather than one of surveillance and accountability (e.g. Metcalfe 1999; Wragg 1999).

The ungraded 'ESCP' model at Rainbow College

As part of a new institutional policy designed to 'move from teaching to learning', Rainbow College implemented an ungraded model of observation, which they referred to as the ESCP (Engage, Support, Challenge, Progress) model. The ESCP model appears to have its origins in US state schools in New York. In many respects, it shares similar principles to that of 'lesson study' in terms of putting student and teacher learning at the centre of the observation process rather than teacher evaluation (e.g. Lewis et al 2006; Lieberman 2009). However, it seems less resource intensive than lesson study. The following list of bullet points, taken from an internal document of Rainbow College, summarises the 'key principles' of their ESCP approach:

- Focus on engaging, supporting, challenging and progressing all learners
- Focus on learning rather than teaching
- Focus on a holistic view of the learning experience
- Develop learning conversations
- Embed quality improvement practice
- Focus on joint practice development with all members of the college community learning from each other
- Embed observation and visits within general practice rather than as one-off performances
- Being a supportive and developmental process, no formal capability process shall be initiated against any staff as a result of this process

These key principles were to form the foundation for the college's new approach to observation and pave the way for making the transition from the previous system. Penny, the college's Head of Quality, commented that the previous system had focused heavily on assessing and judging individual teacher performance. As a result of feedback from staff focus groups, staff evaluations and discussions with the college's team of observers, the college decided that it was time to review and reform its approach to lesson observation. Penny and colleagues were determined to act on the feedback from college staff and move towards a new system where the focus would be switched from individual teacher performance to evaluating the learners' experience and exploring the impact of teaching on that experience.

In the following interview extract, Penny describes how the ESCP model worked in practice at Rainbow College:

There are four separate 'events' if you like. There's a pre-observation meeting, observation, traditional feedback and then a follow up, which is kind of time consuming, but what we have said now is that the pre-observation meeting can be by phone if it's too difficult to meet. The feedback might be short and what we do hope is that it's something that's valuable and it's about what we do. I do think we spend too

much time doing forms rather than discussions and that's something we wanted to change, you know, put the emphasis on the discussions rather than form filling ...

In Penny's description of Rainbow's ESCP model, she emphasised the importance of dialogue between observer and observee. This was reflected in the prominent role it played in the different stages of the model, with three of the 'four events' consisting of discussions between observer and observee. What was also noticeable was the inclusion of a pre-observation meeting and a follow-up discussion to the observation feedback. The former is not common practice across the sector, but is considered an important element in making the observation process more collaborative and increasing teacher ownership, as has been commented in recent studies:

The inclusion of a pre-observation meeting is another important aspect of increasing teacher ownership of the process. With most assessment models of observation, the pre-observation meeting is a rare occurrence. Not only does this provide both observer and observee with an opportunity to discuss the focus of the lesson and for the latter to provide a rationale for their choice, but also enables them to negotiate a set of shared goals that takes into account the needs of the individual and the institution (O'Leary 2013c: 120).

This view of the pre-observation meeting was confirmed by Penny in the following description of how she envisaged these meetings:

They would arrange to meet a teacher before the observation, what we call the pre-observation discussion and talk about the teaching and what the teacher would like them to observe which is mutually decided and basically it's to be more focused on innovation and development so you would choose to try something out with your observer. You would also identify what particular strand you wanted feedback on, so the teacher being observed plays a key role in deciding the focus of the observation.

The ESCP model thus appears to offer a means of framing the discussion between observer and observee, as well as providing them with a set of discourse and phenomena to reflect on and self-evaluate their chosen area(s) of practice. What was also noteworthy of this model was the focus on 'innovation and development', where observees were encouraged to 'try something out'. This willingness to experiment and take risks in one's teaching is fundamental to the CPD of tutors (e.g. IfL 2012), yet opportunities to do so often depend upon the extent to which an institution embraces and actively seeks to promote an expansive approach to professional learning amongst its staff (O'Leary 2013d).

In the following extract, Penny uses the metaphor of a multi-layered 'onion' to refer to the four ESCP categories and how each was accompanied by its own set of criteria

and contextualised examples to help both observer and observee in their (self) evaluation and reflective discussion:

If you think of those four elements, it's like an onion. You've got those four on the front [i.e. Engage, Support, Challenge and Progress], then some mixed sub-criteria under that and then under that layer you've got a kind of grid that shows what does effective practice in 'engaging' learners look like. So you could be having a discussion and I might say to you I think I am really good at 'engaging' students but I don't think I 'challenge' them all. So we'd then look at what comes under the onion of challenge and what does the classroom look like that's very effective and what does a less effective one look like and look at those descriptions etc. And often I think for me one of the most exciting bits of the process is you can start talking about where teachers felt they needed to develop before you observed them, rather than going through this whole process of going along and doing an observation with a one-size fits all set of criteria.

Penny's comments reinforce the importance of including a pre-observation meeting and how that can play a pro-active role in tutors' professional development. The final point she makes about how the ESCP model moves away from a 'one-size fits all set of criteria' and seeks to focus on the identified needs of individual practitioners is also highly significant as it highlights some of the flaws of such normalised models of observation previously discussed. In short, Rainbow College's ESCP model would appear to share many of the characteristics associated with expansive approaches to the use of observation (O'Leary 2013d).

Finally, when asked how the ESCP model had been received by staff at the college, Penny replied that *'the feedback from staff has been overwhelmingly positive and a lot of people did genuinely experiment with something new'*. However, she was keen to stress that its success was not purely as a result of changing the model of observation itself, but how that model was implemented and the way in which staff engaged with it:

It's not just about the model it's about how a model is enacted and for me one of the things we still need to work on and I think it is continuous is developing the observers and their approach to the observations and the perception of some staff and working on the communication of it. So I think the model to me if you had more time is spend a bit more longer in your peer observation discussion and I think that's a valuable element to keep but it is time consuming so it's how you make that most effective.

Penny's comment about needing to develop the 'perception' and 'approach' of observers and observees' to engaging with observation is an apt one with which to conclude the report's findings. For it underlines how sustainable improvement is underpinned by an on-going commitment to and investment in transforming the teaching and learning cultures of an institution and not just the introduction of a new model or initiative as a quick-fix solution (e.g. James and Biesta 2007).

6. Conclusions and recommendations

In order to draw together and briefly summarise the main findings from the previous section, the project's key research questions are re-visited and used as sub-headings in the concluding section of this report. This is then followed by a concise list of ten recommendations of what needs to happen to ensure that the FE sector makes effective use of lesson observation as a form of intervention in teacher assessment and development in the future.

- *What models of lesson observation are currently in use in FE colleges?*

The report's findings have clearly illustrated that the QA model of graded lesson observation, typically carried out at least once a year, remains the dominant model in use across FE. This model is largely underpinned by a performative focus and tends to employ the Ofsted 4-point scale in its assessment criteria. Following recent changes to the Common Inspection Framework (CIF), the sector has witnessed the increasing use of unannounced or 'walk through' observations, with some institutions choosing to grade performance either individually or across curriculum areas and others choosing not to grade these particular observations at all.

Whilst there was evidence of 'alternative models' in practice across the sector, 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 part of teacher education programmes or as an informal, unaccredited activity that staff undertook on a voluntary basis. These alternative models were rarely viewed by senior managers 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.

- *What are the purported aims for their use and how well do the outcomes match these aims?*

The findings have highlighted how many QA observation schemes are often designed to serve multiple aims simultaneously. Among the most common aims cited in institutional policies on lesson observation are the monitoring, measurement and improvement of teaching and learning. As the previous section discussed in detail, the extent to which the outcomes matched these aims was contested across participant groups. For example, there was a significant difference in interpretations as to the value attached to graded observations in monitoring the quality of teaching

and learning by practitioners and senior managers, though the latter comprised a very small portion (n = 20) of the overall sample. This disparity was illustrative of how, for those on the receiving end of these observations i.e. observees, the outcomes they experienced did not equate with the purported aims of those who were responsible for overseeing and implementing them i.e. senior managers and observers.

Despite the differing views of senior managers, the majority of responses also revealed an overwhelming discontent with the use of graded observations for teacher assessment and accountability purposes among practitioners. Senior managers were the only group who believed that graded observations were the fairest way of assessing the competence and performance of staff. As far as practitioners were concerned, performative models of observation were of little relevance to their professional needs or improved their practice and existed principally to furnish SMTs and external agencies with statistical data that could be used to measure and benchmark performance.

One of the particular concerns expressed by a sizeable proportion of participants about current models of observation was how many institutions had come to rely on one-off, annual graded observations as the basis on which overarching decisions about lecturers' professional capability were made. The reality is that lesson observations have become a catch-all, multi-purpose mechanism for many institutions across the sector in preparation for inspection and as part of on-going QA audits, adding further weight to the argument that they predominantly serve the performance management agendas of the institution rather than the professional needs of individual practitioners. I have referred to this in other work as the 'fetishisation' of the observed lesson (O'Leary 2013d). What I mean by this is that the obsession with the observed lesson has resulted in it becoming a high-stakes crucible in which the on-going validation of teaching staff has become concentrated. Yet the validity and reliability of judgements made during these snapshot, episodic observations has been repeatedly called into question in this and previous studies.

- *What is the impact of current models of lesson observation on improving standards in teaching and learning?*

It must be an issue of concern to all those working within the sector that much of the study's qualitative data converged around the counterproductive effects of observation. With regards to the matter of improving standards of teaching via formal observation schemes, for example, there was a groundswell of opinion among the study's participants that the impact of such practice was at best negligible and at worst detrimental to the professionalism of practitioners.

Over three quarters of the study's participants disagreed that graded observations had helped to improve them as classroom practitioners or raise the standard of teaching and learning in their workplace, though this contrasted with the views of senior managers and observers. The fact that there was such a significant level of disagreement across groups as to the benefit of such models of observation raises questions as to whose outcomes are being met. This contrasted noticeably with views of ungraded models, which were regarded much more favourably and seen to be more effective.

The success of observation schemes in use seemed to centre on five key factors:

1. The clarity of purpose and outcomes of the observation (agreed in collaboration/consultation with observees as far as possible)
2. The preparation and training of observers and how well briefed they were on their role(s)
3. The opportunity for observees to engage in substantive professional dialogue
4. The quality of the feedback
5. The allocation of sufficient time to the observation process (to include time for a pre- and post-observation meeting as well as the observation of the lesson itself)

Graded models of observation were repeatedly criticised by a significant majority of participants for being little more than a 'box-ticking' exercise and, in some instances, a 'disciplinary stick' with which 'to beat staff'. They were also identified as a major cause of increased levels of stress and anxiety amongst teaching staff. In contrast, there was a consensus amongst participants that low-stakes, peer-based models of observation were most conducive to sustainable change and professional learning and thus should be at the forefront of most providers' use of observation and wider CPD strategy. However, the importance of these models seemed undermined by the on-going external demand for statistical performance data.

- *How can lesson observation be used most effectively to support FE lecturers' professional learning and development?*

The report's findings have revealed an appetite for change among the majority of practitioners as to how observation is currently used as well as the ways in which their performance is assessed and managed. In the first instance, there is a need for observation to be exploited as a more supportive intervention with the emphasis on helping lecturers to improve their practice rather than the current deficit model that focuses mainly on measuring and judging it.

Secondly, there is a need 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. The findings have highlighted an overwhelming demand to make the process of teacher assessment more inclusive by extending it beyond the lens of lesson observation and drawing on other sources of evidence such as student feedback, peer review, student achievement data etc.

Thirdly, more peer-based models of observation are needed; these would offer the potential to redress some of the power imbalance associated with top-down, deficit models of observation and encourage a greater sharing of practice and professional dialogue that can be mutually beneficial for observer and observee.

Finally, it is fitting that the last responses to the question above are from two of the study's participants, Barry and Vera, both of whom were experienced observers:

I suggest that we must sever the metaphorical umbilical cord that currently exists between performance management/appraisal and lesson observation. Lesson observation is not an essential component of performance management/appraisal nor is the evaluation of performance an essential element of lesson observation. Each is very capable of surviving without the other. Lesson observation may be useful in some instances for gathering data in relation to the achievement of some particular personal goals but it is an inappropriate method for mandatory use in all performance management/appraisal situations ... The most worthwhile lesson observation models acknowledge that teaching is highly individualised and contextualised. The best lesson observation schemes recognise the complexities of teaching and are not in a rush to reach judgements about the effectiveness of particular practice. In the better models the focus is more one of discovery and illumination whereby the nuances of practice which occurred at a particular point in time are brought in to the open and explored and deliberated in a collegiate way (Barry, observer).

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)

Recommendations

- i. **Alternative approaches to the use of observation:** There is a need to explore alternative approaches to the current, dominant model of graded lesson observation. Such alternatives should seek to combine elements of existing practices but also make use of recent advances in the research of observation as a

mechanism for professional learning. It is recommended that these alternative approaches include features such as differentiated observation, self-assessment (i.e. observee) as well as observer assessment, peer reviews etc.

- ii. **Prioritising improvements in teaching:** Practitioners need more support with how to improve their teaching and less emphasis on measuring their performance. Thus any future use of observation should seek to prioritise the professional development needs of staff rather than the production of statistical data to serve performance management systems.
- iii. **Formal allocation of timetabled hours for ‘feedback’ and ‘feed-forward’:** In those institutions where expansive approaches to the use of observation were evident, the importance of the feedback and feed-forward stages of the process was acknowledged by the formal allocation of timetabled hours to this activity. This was certainly not common practice across many institutions but it is recommended that in order for the outcomes of observation to have an impact on CPD and the on-going improvement of practice, these important stages need to be accredited with sufficient time that is incorporated into staff timetables at the beginning of the academic year. It is therefore recommended that any observation must include a pre-observation meeting between observer and observee and a feed-forward meeting.
- iv. **Multi-dimensional model of assessment:** Graded observations should no longer be relied on as the main source of evidence on which to judge tutors’ professional competence and performance. They are reductive in nature and used in isolation cannot be seen as a valid and/or reliable indicator. Such judgements need to be based on a multi-dimensional model of teacher assessment as discussed above, encompassing a varied portfolio of evidence (e.g. student achievement, student feedback/evaluations, peer review, self-assessment, external verification), so as to ensure a more triangulated and reliable evidence base for assessment.
- v. **Review of observation assessment criteria⁵:** Providers are recommended to carry out regular reviews of their observation assessment criteria and consider the extent to which they cater for the diverse contexts and curriculum areas offered. For example, are the same assessment criteria used across all curriculum areas? If so, do these criteria capture the relevant subject specialisms and any associated pedagogy appropriately?
- vi. **Observers’ qualifications and training:** All observers with a responsibility for carrying out formal, assessed observations must obtain a recognised qualification

⁵ For practical suggestions on how an institution might go about carrying out a review of its observation policy, see O’Leary (2013c: 77-79).

before embarking on their role. Given the complexity of the role and the skills required to ensure that their judgements remain as valid and reliable as they possibly can be, it is also essential that observers regularly update their knowledge and skills with suitable training and standardisation activities.

- vii. **Input from teacher educators in creating an observation policy:** When creating an institutional policy for lesson observations, senior management needs to ensure that it draws on the experience and expertise of those involved in teacher education/training programmes to inform discussion, as these practitioners are immersed in the domains of teacher assessment, development and the area of lesson observation on an on-going basis and thus are well-positioned to make valuable contributions. Not only would this help to create an observation framework underpinned by informed and current practice, but it might also lessen the likelihood of future disputes if teacher educators were to act as mediators, along with union representatives, between practitioners and senior managers in drawing up such policy.
- viii. **Sever links between formal observations and capability procedures:** The outcome of formal lesson observations, whether they are graded or ungraded, should not be linked directly to an institution's capability/disciplinary policy. Given the misgivings surrounding the validity and reliability of observation as a method of assessment discussed at length in this report, this study recommends that any institution's capability policy needs to reflect this by severing any formal links between the two accordingly.
- ix. **Support for underperforming tutors:** In the case of those tutors whose classroom performance is deemed to be below standard or considered a cause for concern, appropriate systems need to be put in place to ensure that they are given the relevant professional support in order to enable them to improve their knowledge base and skills before any conclusive decision is made on their capabilities. Timescales for improvement should be agreed between all parties. In such cases, a temporary reduction in their teaching load should be agreed so that they can undertake the necessary training and support to equip them with the knowledge and skills to improve future performance.
- x. **Observee empowerment:** There is a need to empower observees with the opportunity to play an active role in the focus of their observation and the ability to decide and prioritise key areas for development in collaboration with their observers. Thus action plans following on from observations need to be negotiated and mutually agreed between the observer and observee. In cases where it is impossible to reach an agreement, a third party may be involved in mediating.

References

- Adshead, L., White, P. T., & Stephenson, A. (2006) Introducing peer observation of teaching to GP teachers: a questionnaire study. *Medical Teacher*, 28, 68-73.
- Arksey, H. and Knight, P. (1999) *Interviewing for Social Scientists*. London: Sage.
- Armitage, A. Byrant, R. Dunnill, R. Hammersley, M. Hayes, D. Hudson, A. & Lawless, S. (2003) *Teaching and training in post-compulsory education*. 2nd ed., Buckingham: OUP.
- Association of Teachers and Lecturers (ATL) (2008) *ATL Advice – Classroom Observation*. Available online at: <http://www.atl.org.uk/Images/ADV19%20classroom%20observation.pdf>. Date accessed 23/3/2010.
- Bennett, S. & Barp, D. (2008) 'Peer observation - a case for doing it online'. *Teaching in Higher Education*, 13(5), 559 – 570.
- Black, P. (1998) *Testing: friend or foe? Theory and practice of assessment and testing* (London: Falmer).
- Boffey, D. (2012) Lecturers should need a teaching qualification, says NUS president. *The Observer*, 22nd April 2012. Available online at: <http://www.guardian.co.uk/education/2012/apr/22/liam-burn-nus-academics-lecturers> Date accessed 28/7/2012.
- Boocock, A. (2013) Observation of teaching and learning: teacher development or micropolitics and neo-Fordism. *Journal of Further and Higher Education*, 37(4), 482-503.
- Brookfield, S. D. (1995) *Becoming a Critically Reflective Teacher*. San Francisco, CA: Jossey-Bass.
- Brooks, V. (2009) 'Marking as judgment', *Research Papers in Education*, First published on: 11 November 2009 (iFirst).
- Brown, S., Jones, G. & Rawnsley, S. (1993) *Observing Teaching – SEDA Paper 79*. Birmingham: SEDA.
- Burrows, J. (2008) *Trainee Perceptions of Observation*. Huddersfield: Huddersfield Consortium.
- Campbell, J., Kyriakides, L., Muijs, D & Robinson, W. (2004) *Assessing Teacher Effectiveness – Developing a Differentiated Model*. London:Routledge Falmer.
- Cockburn, J. (2005) Perspectives and politics of classroom observation. *Research in Post-Compulsory Education*, 10(3), 373 – 388.
- Coffield, F. (2012) To grade or not to grade. *Adults Learning*, Summer 2012, pp. 38-39, NIACE. Available at: http://www.niace.org.uk/sites/default/files/documents/adults-learning/AdultsLearning_2012_04-pages38-39.pdf
- Cohen, L., Manion, L. and Morrison, K. (2011) *Research Methods in Education – 7th Edition*. London: Routledge.
- Cope, P., Bruce, A., McNally, J. & Wilson, G. (2003) 'Grading the practice of teaching: an unholy union of incompatibles'. *Assessment & Evaluation in Higher Education*, Vol. 28(6), 673 – 684.
- Creswell, J. (2003) *Research design: qualitative, quantitative, and mixed approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.

- Denzin N. K. (1978) *The Research Act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- Department for Education (DfE) (2010) *The importance of teaching – Schools' White Paper*. London: DfE.
- Downey, C.J., English, F.W., Frase, L.E., Poston, W.K., and Steffy, B.E. (2004). *The three-minute classroom walkthrough: Changing school supervisory practice one teacher at a time*. Thousand Oaks, CA: Corwin Press.
- Dudley, P. (2007) *Lessons for Learning: Using lesson study to innovate, develop and transfer pedagogic approaches and metapedagogy*. London: TLRP. Available online at: <http://www.bera.ac.uk/lesson-study/>. Date Accessed: 4/10/2011.
- Dudley, P. (2008) *Lesson Study Development in England: practice to policy*, paper presented at the World Association of Lesson Studies Annual Conference, Hong Kong, December 2008. Available at: http://www.tlrp-archive.org/tlrp/upload/assets/1231760840_PDWALS08.pdf Date Accessed: 4/10/2011
- Edgington, U. (2013) *Performativity and Affectivity: Lesson Observation in England's Further Education Colleges*, unpublished PhD Thesis, Canterbury Christ Church University, November 2013.
- Ewens, D. & Orr, S. (2002) *Tensions between evaluation and peer review models: Lessons from the HE/FE border*. London: LTSN Generic Centre.
- Fawcett, M. (1996) *Learning Through Child Observation*. London: Jessica Kingsley.
- Foster, P. (1996) *Observing Schools – A Methodological Guide*. London: Paul Chapman Publishing.
- 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.
- Freire, P. (1972) *Pedagogy of the Oppressed*. Harmondsworth: Penguin.
- Fullerton, H. (2003) Observation of teaching, in Fry, H., Ketteridge, S. & Marshall, S. (Eds) *A handbook for teaching and learning in higher education* (2nd edition). London: Kogan-Page.
- Gipps, C. (1994) *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer Press.
- Gorard, S. & Taylor, C. (2004a) *Combining Methods in Educational and Social Research*. Berkshire: Open University Press.
- Gorard, S. & Taylor, C. (2004b) What is 'triangulation'? *Building Research Capacity*.
- Gosling, D. (2002) *Models of Peer Observation of Teaching*. London: LTSN Generic Centre.
- Gray, D. (2004) *Doing research in the real world*. London: Sage.
- Grierson, A. & Gallagher, T.L. (2009) Seeing is believing: Creating a catalyst to teacher change through a demonstration classroom initiative. *Professional Development in Education*, 35(4), 567-584.

- Grubb, W. N. (2000) Opening Classrooms and Improving Teaching: Lessons from School Inspections in England. *Teachers College Record*, 102(4), 696-723.
- Hammersley-Fletcher, L. & Orsmond, P. (2004) Evaluating our peers: is peer observation a meaningful process? *Studies in Higher Education*, 29(4), 489-503.
- Hammersley-Fletcher, L. & Orsmond, P. (2005) Reflecting on reflective practices within peer observation. *Studies in Higher Education*, 30(2), 213-224.
- Hardman, J. (2007) *The use of teaching observation in Higher Education: An exploration of the relationship between teacher observation for quality assurance and quality improvement in teaching in higher education, in the light of further education sector experience*. Report produced for Escalate August 2007. Available online at: <http://escalate.ac.uk/3311>. Date accessed 20/1/2010.
- Harris, C. (1997) Developing online market research methods and tools, Paper presented to ESOMAR Worldwide Internet Seminar, Lisbon, July 1997.
- Institute for Learning (IfL) (2012) *Leading learning and letting go: building expansive learning environments in FE*. London: Institute for Learning.
- James, D. & Biesta, G. eds. (2007) *Improving learning cultures in Further Education*. London: Routledge.
- Johnson, R.B. and Onwuegbuzie, A.J. (2004) Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), pp. 14-26.
- Kincheloe, J. L. (2004) The Knowledges of Teacher Education: Developing a Critical Complex Epistemology, *Teacher Education Quarterly*, 31(1), 49-66.
- Lave, J. & Wenger, E. (1991) *Situated Learning – Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lawson, T. (2011) 'Sustained classroom observation: what does it reveal about changing teaching practices?' *Journal of Further and Higher Education*, First published on 25 March 2011 (iFirst): <http://dx.doi.org/10.1080/0309877X.2011.558891>.
- Learning and Skills Improvement Service (LSIS) (2012) *Further Education College Workforce Data for England – An analysis of the Staff Individualised Record (SIR) 2010-2011*. Coventry: UK.
- Lee, J. (2007) *Are you being observed?* London: Times Educational Supplement, 15 June 2007.
- Lewis, C., Perry, R. & Murata, A. (2006) How should research contribute to instructional improvement? The case of lesson study, *Educational Researcher*, 35(3), 3–14.
- Lieberman, J. (2009) Reinventing teacher professional norms and identities: The role of lesson study and learning communities. *Professional Development in Education*, 35(1), 83–99.
- Lowe, R. (2007) *The Death of Progressive Education: How teachers lost control of the classroom*. Oxford: Routledge.
- Madge, C. & O'Connor, H. (2002) On-line with e-mums: exploring the Internet as a medium for research, *Area*, 34(1), pp. 92-102.
- Marriott, G. (2001) *Observing Teachers at Work*. Oxford: Heinemann.

- McMahon, T., Barrett, T., & O'Neill, G. (2007) Using observation of teaching to improve quality: Finding your way through the muddle of competing conceptions, confusion of practice and mutually exclusive intentions. *Teaching in Higher Education* 12(4), 499-511.
- Metcalfe, C. (1999.) Developmental Classroom Observation as a Component of Monitoring and Evaluating the Work of Subject Departments in Secondary Schools. *Journal of In-service Education*, 25(3), 447-459.
- Miles, M. B. & Huberman, A.M. (1994) *Qualitative Data Analysis*, 2nd Edition. Newbury Park, CA: Sage.
- Montgomery, D. (2002) *Helping Teachers Develop through Classroom Observation – Second Edition*. London: David Fulton Publishers.
- National Union of Teachers (NUT) (2006) *A Classroom Observation Protocol: Guidelines for NUT School Representatives*. Available online at: <http://www.teachers.org.uk/node/1335> Date accessed: 22/1/2010.
- National Union of Teachers (NUT) (2007) Classroom observation: No to grading of lessons. *NUT News*, 17, September 2007. Available online at: <http://www.teachers.org.uk/node/6298> Date accessed: 22/1/2010.
- Newman, I. & Benz, C. (1998) *Qualitative-Quantitative Research Methodology – Exploring the Interactive Continuum*. Carbondale, IL: Southern Illinois University Press.
- Ofsted (2012) *A good education for all*. 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(2), 191 – 198.
- O'Leary, M. (2011) *The Role of Lesson Observation in Shaping Professional Identity, Learning and Development in Further Education Colleges in the West Midlands*, unpublished PhD Thesis, University of Warwick, September 2011.
- O'Leary, M. (2012a) Exploring the role of lesson observation in the English education system: a review of methods, models and meanings. *Professional Development in Education*, Vol. 38(5), pp. 791-810. Link at: <http://www.tandfonline.com/doi/abs/10.1080/19415257.2012.693119>
- O'Leary, M. (2012b) 'Time to turn worthless lesson observation into a powerful tool for improving teaching and learning'. *CPD Matters/InTuition – IfL*, Issue 9, Summer 2012, pp. 16-18. Available at: http://www.ifl.ac.uk/_data/assets/pdf_file/0008/27890/InTuition-issue-9-Summer-2012.pdf
- O'Leary, M. (2013a) Surveillance, performativity and normalised practice: the use and impact of graded lesson observations in Further Education Colleges. *Journal of Further and Higher Education*, Vol. 37(5), pp. 694-714. Available online at: <http://www.tandfonline.com/doi/abs/10.1080/0309877X.2012.684036>
- O'Leary, M. (2013b) 'Embracing expansive approaches to the use of lesson observation'. *CPD Matters/InTuition – IfL*, Issue 8, Summer 2013, pp. 21-22. Available at: <http://www.ifl.ac.uk/publications/intuition>.

- O'Leary, M. (2013c) *Classroom observation: a guide to the effective observation of teaching and learning*. London: Routledge.
- O'Leary, M. (2013d) Expansive and restrictive approaches to professionalism in FE colleges: the observation of teaching and learning as a case in point. *Research in Post-Compulsory Education*, 18(4), pp. (Due for publication in Special Issue, Winter 2013).
- Ollin, R. (2009) *The grading of teaching observations: implications for teacher educators in Higher Education partnerships*. Huddersfield. Huddersfield Consortium.
- Oppenheim, A. (2001) *Questionnaire Design, Interviewing and Attitude Measurement*. London: Pinter.
- 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(4), 489-504.
- Punch, K. (2006) *Developing Effective Research Proposals – Second Edition*. London: SAGE Publications.
- Ramsden, P. (1992) *Learning to Teach in Higher Education*. London: Routledge.
- Robson, C. (2002) *Real World Research – Second Edition*. Oxford: Blackwell.
- Sachs, J. (2001) Teacher Professional Identity: Competing Discourses, Competing Outcomes. *Journal of Educational Policy*, 16(2), 149-161.
- Sadler, D.R. (2005) Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education* 30, 175–94.
- Sharp, S. (2006) [The grading of placement in initial teacher education in Scotland](http://www.scotedreview.org.uk/pdf/208.pdf). Available online at: <http://www.scotedreview.org.uk/pdf/208.pdf> . Date accessed: 17/5/2010.
- Shortland, S. (2004) Peer Observation: a tool for staff development or compliance? *Journal of Further and Higher Education*, 28(2), 219-228.
- Silverman, D. (2006) *Interpreting Qualitative Data – Third Edition*. London: Sage.
- Simons, H. & Elliott, J. eds. (1990) *Rethinking Appraisal and Assessment*. Milton Keynes: Open University Press.
- Stigler, J. & Hiebert, J. (1999) *The teaching gap: best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.
- Tashakkori, A. & Teddlie, C. (1998) *Mixed Methodology – Combining Qualitative and Quantitative Approaches*. London: Sage.
- Tilstone, C. (1998) *Observing Teaching and Learning – Principles and Practice*. London: David Fulton.
- University and College Union (UCU) (2009) *Lesson Observation: UCU Guidelines*. [Accessed 22/02/2010] Available at: <http://www.ucu.org.uk/index.cfm?articleid=2969>.
- UCU (2012) *UCU FE News – May 2012*. [Accessed 19/06/2012] Available via: <http://www.ucu.org.uk/>
- Wang-Iverson, P. (2002) 'What is Lesson Study?' Research for better schools. *Currents*. Volume V, No. 2, pp. 1-2. [Accessed 20th July 2012]. Available online at: http://www.rbs.org/SiteData/doc/currents_0502/320e53d8a9347dcbe2243a74532a6a41/currents_0502.pdf.

- Whitty, G. (2000) Teacher professionalism in new times, *Journal of In-Service Education*, 26(2), 281-295.
- Wilcox, B. and Gray, J. (1996) *Inspecting schools: holding schools to account and helping schools to improve*. Buckingham: Open University Press.
- Wragg, E. C., Wikeley, F. J., Wragg, C. M. & Haynes, G. S. (1996) *Teacher Appraisal Observed*. London: Routledge.
- Wragg, E.C. (1999) *An Introduction to Classroom Observation – 2nd Edition*. London: Routledge.

Appendix 1

Lesson Observation Questionnaire

Dear member,

This questionnaire is concerned with the important topic of lesson observation. Its purpose is to gather your views and experiences of how observation is currently used in the FE sector and its impact on members. It should take you no more than 10-15 minutes to complete. You are assured that this questionnaire is anonymous and all responses will be held securely and treated confidentially.

Section A.

Please tick only **ONE** box for each of the questions below.

1. What gender are you?

- Male Female Transgender

2. What is your current employment status?

- Full-time permanent Fractional permanent Part-time hourly paid Seeking employment Retired

3. In what capacity are you currently employed in your workplace?

- Senior Manager Middle Manager Lecturer/Tutor Other (please state)

4. How many years' teaching experience do you have?

- Less than 2 years 2-5 years 5-10 years 10-15 years 15+ years N/A

5. In what capacity are you principally involved in the lesson observation process in your workplace?
 Observer Observee Observer and Observee Other (please state)

6. Which of the contexts below best describes your most recent experience of lesson observation in your workplace?
 Internal Quality Assurance scheme Ofsted inspection External consultation/mock inspection
 Peer review/peer development Other (please state)

7. Which of the models of lesson observation described below is most commonly used in your workplace?
 Managerial, graded (1-4) model using internal observers Managerial, graded (1-4) model using external observers
 Developmental, ungraded model with jointly agreed action plan by observer and observee
Ungraded, peer model without action plan Other (please state)

8. Are unannounced lesson observations currently in use in your workplace?
 Yes No

9. How much notice is given before a lesson observation is carried out in your workplace?
 No notice Less than 2 days 2-5 days A week More than a week N/A

10. How much notice would you consider to be acceptable before an observation is carried out in your workplace?
 No notice Less than 2 days 2-5 days A week More than a week N/A

Section B: Each statement below relates to the use of **graded lesson observations**. Please tick only **ONE** box for each of the statements.

| I believe that graded lesson observations ... | Strongly agree | Agree | Disagree | Strongly disagree | N/A |
|---|----------------|-------|----------|-------------------|-----|
| 11. are essential for monitoring the quality of teaching and learning | | | | | |
| 12. are essential for improving the quality of teaching and learning | | | | | |
| 13. are essential for the continuing professional development (CPD) of staff | | | | | |
| 14. are the most effective method of assessing staff competence and performance | | | | | |
| 15. are a reliable indicator of staff performance | | | | | |
| 16. have helped to raise the standards of teaching and learning in my workplace | | | | | |
| 17. have helped me to improve as a classroom practitioner | | | | | |
| 18. are the fairest way of assessing the competence and performance of staff | | | | | |
| 19. are a necessary part of staff appraisal | | | | | |
| 20. should no longer be used as a form of assessment in the FE sector | | | | | |

Section C: Each statement below relates to the use of **ungraded lesson observations**. Please tick only **ONE** box for each of the statements.

| I believe that ungraded lesson observations ... | Strongly agree | Agree | Disagree | Strongly disagree | N/A |
|--|----------------|-------|----------|-------------------|-----|
| 21. are more effective in improving the quality of teaching and learning than graded lesson observations | | | | | |
| 22. play a more important role in the CPD of staff than graded lesson observations | | | | | |
| 23. are more effective in assessing staff than graded observations | | | | | |
| 24. are a reliable indicator of staff performance | | | | | |
| 25. have helped to raise the standards of teaching and learning in my workplace | | | | | |
| 26. have helped me to improve as a classroom practitioner | | | | | |
| 27. should replace the use of graded lesson observations in the FE sector | | | | | |

Section D: Each statement below relates to **lesson observation feedback**. Please tick only **ONE** box for each of the statements.

| I believe that lesson observation feedback ... | Strongly agree | Agree | Disagree | Strongly disagree | N/A |
|--|----------------|-------|----------|-------------------|-----|
| 28. given during graded observations has identified clear areas of improvement | | | | | |
| 29. given during ungraded observations has identified clear areas of improvement | | | | | |
| 30. is fair and consistent in all observations carried out in my workplace | | | | | |
| 31. is well managed in my workplace | | | | | |
| 32. could be improved in my workplace | | | | | |
| 33. is allocated sufficient time in the observation process in my workplace | | | | | |
| 34. is the most important part of the observation process | | | | | |

Section E: Each statement below relates to **unannounced lesson observations**. Please tick only **ONE** box for each of the statements.

| I believe that unannounced lesson observations ... | Strongly agree | Agree | Disagree | Strongly disagree | N/A |
|---|----------------|-------|----------|-------------------|-----|
| 35. are a welcome addition to the quality improvement process | | | | | |
| 36. will lead to increased levels of stress and anxiety amongst staff | | | | | |
| 37. will help to improve the accuracy and reliability of teacher assessment | | | | | |

| | | | | | |
|---|--|--|--|--|--|
| 38. will help to identify underperforming practitioners more easily | | | | | |
| 39. are a positive development in assessing the competence and performance of staff | | | | | |
| 40. should be a statutory requirement for assessing teaching and learning in FE | | | | | |

Section F: Additional Comments

If there are any other comments that you wish to make about lesson observation, then please type them in the box below.

Thank you for completing this questionnaire. If you wish to participate in a short interview/focus group in the second phase of this research then please leave an email address or contact telephone number below. Please rest assured that your contact details and any other information you provide in this questionnaire will be treated in the strictest confidence. Your identity will remain protected at all times.

Email: Contact Telephone Number:

Appendix 2

Interview & Focus Group Questions for Phase 2 of Data Collection – July 2013

1. What models of lesson observation exist in this college?
2. Is there more than one model in use and if so how do they differ in their aims and focus?
3. What do you think the impact is of these models of observation on improving standards in teaching and learning?
4. Do you think particular models of observation are more worthwhile than others?
5. If so, are you able to say what the defining features are of these models?
6. How can the use of lesson observation be improved to maximise its benefits to FE lecturers' professional learning and development?
7. Are there any features of lesson observation that you would like to see included or indeed removed from the way in which it is currently used in your college?