A CRITICAL EVALUATION OF CRITERIA-BASED ASSESSMENT OF SUBJECT KNOWLEDGE AND OTHER COMPETENCIES OF TEACHERS IN TRAINING

RICHARD JAMES TYNAN

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of PhD by publication

PART ONE

A PORTFOLIO OF PUBLISHED PAPERS

August 2021

Insights from a subject knowledge enhancement course for preparing new chemistry and physics teachers

Michael Inglis, Andrea Mallaburn, Richard Tynan, Ken Clays and Robert Bryn Jones

ABSTRACT A recent Government response to shortages of new physics and chemistry teachers is the extended subject knowledge enhancement (SKE) course. Graduates without a physics or chemistry bachelor degree are prepared by an SKE course to enter a Postgraduate Certificate in Education (PGCE) programme to become science teachers with a physics or chemistry specialism. SKE courses challenge common assumptions about the nature of subject knowledge for teaching and who should teach it: school science educators or scientists? This article shares the SKE course model developed and taught by the Science Education team at Liverpool John Moores University, and some early insights into supporting subject knowledge development.

Setting the scene

A variety of initial teacher education (ITE) routes for prospective secondary chemistry or physics teachers exist. The most common ITE route is a 1 year Postgraduate Certificate in Education (PGCE) course, aimed at graduates with a bachelor degree in a science subject. PGCE courses are usually led by a university in partnership with schools. To be awarded qualified teacher status (QTS) in England and Wales, a student teacher must demonstrate they have met a series of QTS Standards (revised for September 2012 onwards). Standard 3 states that a teacher must '*demonstrate good subject and curriculum knowledge*' (Department for Education, 2012), and specifically must:

● *have a secure knowledge of the relevant subject(s) and curriculum areas, foster and maintain pupils' interest in the subject, and address misunderstandings*
● *demonstrate a critical understanding of developments in the subject and curriculum areas, and promote the value of scholarship*

The Government requires PGCE courses to provide 120 days of school-based development, typically leaving 60 days of university-based development. The university-based days commonly focus on aspects of science education

pedagogy and critical engagement with educational literature, with limited time available for in-depth exploration and development of subject knowledge. The Government has stated an intention that new teachers should hold at least a 2:2 class bachelor degree (Department for Education, 2010). This policy suggests that degree classification is seen by policy makers to be a good indicator of potential to be an effective teacher. Difficulties with recruiting sufficient teachers has led to the creation by the Government of extended subject knowledge enhancement (SKE) courses, aimed at graduates with insufficient chemistry or physics at bachelor degree level to enter confidently into a PGCE course (Department for Education, 2013). This has resulted in an ITE route that is presented officially as 6 months to 1 year SKE (i.e. developing knowledge of the subject) followed by 1 year PGCE (i.e. developing knowledge of how to teach the subject).

Why is this important?

The SKE course concept challenges the conventional wisdom that a bachelor degree in a subject is a prerequisite to being an effective teacher of that subject. It also raises questions about to what extent science teachers are, or should be, primarily teachers of science or teachers of biology, chemistry or physics. At

least one of the authors of this article expressed scepticism that it would be possible for someone without a physics degree to develop the physics understanding needed to thrive on a PGCE course, and the idea can polarise opinions among educators. Analysis of research has shown that a science teacher's bachelor degree classification (i.e. the awarded *grade* of achievement) has a less significant effect on pupils' learning outcomes than some educational policy makers appear to expect, with factors such as quality of relationships with pupils appearing to exert a more significant influence (Hattie, 2009). What is clear is that, regardless of level of academic qualification, the quality of a teacher's understanding of fundamental concepts in a subject plays an important role in enabling a teacher to deal effectively with pupils' misconceptions (van Driel, Verloop and de Vos, 1998), and this understanding cannot be achieved solely through a subject knowledge audit-driven approach (Lock, Salt and Soares, 2011).

Since SKE courses started at Liverpool John Moores University (LJMU), 67 students have gained QTS having gone through the SKE-then-PGCE route. This article aims to share with science educators how the SKE course model has evolved at one university and to stimulate thinking about the nature and development of science subject knowledge for teaching. A future article will share research by LJMU course tutors into SKE students' experiences of developing subject knowledge and how this has informed their practice in schools.

What is teacher subject knowledge?

Even a cursory look at the literature about teacher subject knowledge reveals that the answer to this question is complex and contentious. What follows is by necessity a brief and selective guided tour that is intended to provide some food for thought. The authors referred to are by no means the only people researching and writing on this topic, but we suggest that they form a starting point for a deeper exploration of thinking about subject knowledge.

Much of the discussion over the last 20 years about the nature of subject knowledge for teaching takes as a starting point that there is a clear difference between scholarly science knowledge (the realm of the scientist) and school science knowledge (the realm of the teacher). Perhaps

the most commonly cited example is a model proposed by Lee Shulman, in which he referred to subject knowledge as *content knowledge* (Shulman, 1986). Shulman proposed that content knowledge can be divided into three categories:

● **Subject matter content knowledge (SMCK)** consists of the '*amount and organization of knowledge per se in the mind of the teacher*' (p. 9). This involves not just knowing the 'facts' of science, but also understanding the rules and principles by which these 'facts' are organised and amended. So, chemistry SMCK is the area of chemistry subject knowledge that is common to both a scholar of chemistry and a teacher of chemistry.

● **Pedagogical content knowledge (PCK)** '*goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge* for teaching … *it embodies the aspects of content most germane to its teachability*' (p. 9) and is the area of knowledge distinctive to teachers, such as knowledge of appropriate analogies and demonstrations, and what makes understanding of particular concepts challenging for learners.

● **Curricular knowledge (CK)** includes knowledge of the programmes and routes that can be followed, and the resources that can be utilised. Shulman described it as '*the pharmacopeia from which the teacher draws those tools of teaching that present or exemplify particular content and remediate or evaluate the adequacy of the student accomplishments*' (p. 10).

The concept of *PCK* is often used by educators. The authors of this article have participated in various meetings and conferences with teachers where PCK is referred to in an uncritical way and with some variation in meaning. Among those who use Shulman's language there appears to be consensus around two specific points (van Driel *et al.*, 1998):

● PCK is subject-specific and therefore different from knowledge of general pedagogy;
● PCK is about how particular topics can be *taught* rather than 'pure' subject knowledge in itself.

In the language of Shulman (1986), the SKE-then-PGCE route can be interpreted as development of SMCK (SKE course) and then development of PCK and CK (PGCE course), an interpretation implied by much of the language used to market these courses.

There are criticisms of the lack of evidence proposed by Shulman to support the idea of PCK

as a distinct category of knowledge, with some instead suggesting that PCK emerges through a process of blending general pedagogy with content.

Banks, Leach and Moon (1999) propose that teachers' subject understanding can also be divided into three categories:

- **Subject knowledge** (i.e. scholarly knowledge);
- **School knowledge** which is created through a process of *transposition* of scholarly knowledge into a restructured and linear form that can be accessible by, and taught to, children;
- **Pedagogic knowledge** which is knowledge of teaching and learning, and an understanding of the relationship between subject knowledge and school knowledge.

These three categories interact dynamically along with the *personal constructs* of the teacher, which emerge from the teacher's prior experiences, beliefs and knowledge of teaching. It is interesting to compare the linear structure of *science school knowledge* created by teachers with the messiness and non-linearity of how scientists actually work. One of the criticisms that can be made about school science is the neat and tidy picture it presents of scientific progress taking place in an inevitable, planned and orderly way (a picture some scientists themselves have been happy to foster!). As with Shulman (1986), the Banks *et al.* (1999) model claims that there is something distinctive about the subject understanding held by science teachers, although it is unclear what the criteria are for deciding whether an idea or model is an example of *subject* or of *school* knowledge.

Models of teacher subject knowledge such as those of Shulman (1986) and Banks *et al.* (1999) are criticised by some for ignoring the effect of interaction with learners on the development of subject understanding. These models imply that subject knowledge and its development resides with individuals and can be 'boosted', 'enhanced' or 'audited' (i.e. it is *objective*). This *objectivist* view of knowledge is criticised by those who adopt social models of learning (e.g. *socio-constructivist* models) and for ignoring the essentially pedagogic nature of language. Claiming to understand a scientific concept requires the use of language, which is an act of communication, i.e. a pedagogical act. For example, when a scientist explains to colleagues some aspect of her

research into the electrical properties of graphene, her explanation is tailored according to who and what the explanation is intended for: explanations always have someone else in mind. As McEwan and Bull (1991: 324) put it:

[s]cholars must be concerned with the comprehensibility and teachability of their assertions, that is, with whether those 'representations' can find a meaningful place in others' webs of belief ... the justification of scholarly knowledge is inherently a pedagogical task ...

This suggests that what is commonly called *subject knowledge for teaching* may be more to do with how experienced and adept someone is at formulating effective explanations for particular groups of learners (e.g. secondary pupils) rather than a body of knowledge that is unique to people labelled as teachers. This experience comes from interacting with others and will be driven by context. Some views of teacher subject knowledge regard these factors as central to how teachers develop their knowledge. Ellis (2007: 447) proposes that subject knowledge should be treated as '*complex, dynamic and as situated as other categories of teachers' professional knowledge*'. Ellis highlights the significance of new teachers participating in what Wenger (1998) described as a *community of practice*, where subject understanding is constructed (or *negotiated*) through interaction with peers and with learners. This is a long way from the simple model of teacher subject knowledge as something that can be taught to individuals as a distinct subject in a context far removed from the one they will be using it in.

These models of subject knowledge development (and others) continue to challenge our thinking about the SKE and PGCE courses and how to support student teachers effectively.

Designing the SKE course

When we first started to plan the course, one of the options considered was to base some of it in the Science Faculty in LJMU, with some supervision and coordination to be provided by the Education Faculty. This idea was discounted owing mainly to concerns about making sure the course content and teaching approaches used would be suitable to support students to prepare them to become science teachers rather than scientists. Looking back at those early discussions,

it is striking how this important decision was based on experience and professional judgement as school science education practitioners, rather than careful consideration of research evidence or other literature. In our experience, this decision has been vindicated and recent studies have shown that students prefer SKE courses to be run by ITE tutors (Lock *et al*., 2011). If subject knowledge for teaching really can be categorised in the way suggested by Shulman (1986) then choosing to delegate subject knowledge development to the relevant academic university departments should be reasonable, with appropriate guidance given to the departments concerned about what to cover. Experience and feedback from other universities that have taken this approach suggest otherwise. At an early stage it was decided to create a course that would be validated by the university as a graduate diploma, which would be worth 120 credits at National Qualifications Framework Level 6 (in the NQF, Level 6 is equivalent to the final year of a BSc). What made the course Level 6 was the emphasis on critical engagement with subject knowledge and the level of independent learning required. The level of chemistry or physics covered was aimed at Levels 2–4, that is GCSE-level understanding of key concepts and building up to A-level and beyond.

The SKE course ran for the first time at LJMU in 2008/9, with a course structure requiring all of the students to study both chemistry and physics during the first semester (September to December) and then choosing one subject to specialise in for the second semester (January to May). Experience and student feedback informed the decision to develop the course further for 2009/10 so that students opted to be chemistry or physics specialists from the start of the course but they still studied both sciences during the first semester. For the 2010/11 academic year onwards, both sciences were treated as separate routes from the beginning to enable students to have more time and support to develop their subject understanding. The numbers of students recruited each year is shown in Table 1.

SKE course students start off as candidates for a PGCE course and all have at least one science A-level. During the PGCE selection day, the course team makes an assessment about candidates' subject experience (a process that leads to much discussion within the course team about how to judge meaningfully someone's

Table 1 Numbers of SKE students at LJMU by route

Academic year	Number on chemistry route	Number on physics route
2008/9	9	7
2009/10	12	12
2010/11	19	11
2011/12	25	12

subject knowledge during a selection process). An offer of a PGCE place may then be made, conditional on successfully completing an extended SKE course first. During 2008/9 and 2009/10, a large proportion of candidates entered the course with psychology or sports sciences backgrounds. In the last 2 years we have seen an increasing number of students enter the SKE course with biological sciences backgrounds, in response to the reduction in number of ITE places available for biology.

Course evaluation feedback from the first two cohorts of SKE students was consistent about the desire to integrate pedagogical considerations into the course and for more support with the amount of independent learning required. The version of the SKE course that has run for the last 2 years incorporates tutor and student feedback and experience and is summarised in Figure 1.

Different modules are phased in at different times so that the students have a chance to develop confidence and to focus on the basic skills and understanding needed for later in the course.

Essential Chem/Phys Concepts and *Further Chem/ Phys Concepts*

The *Essential Chem/Phys Concepts* and *Further Chem/Phys Concepts* modules form the backbone of the course and cover a range of key concepts and ideas in the relevant subject (see Table 2).

The assessment involves an end-of-module examination and a series of assignments where students work independently on questions or

Table 2 Example topics for the main physics/chemistry modules

Physics	Chemistry
Forces and Motion	Atomic structure
Energy	The Periodic Table
Wave and particle models	Equilibrium
Electricity and electromagnetism	Chemistry of carbon
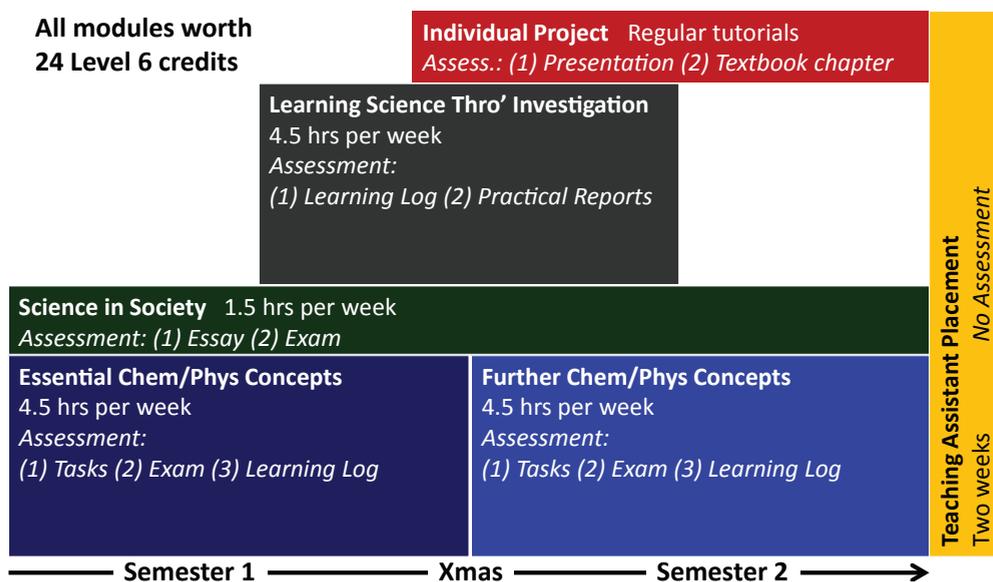
Figure 1 The current extended SKE course structure at LJMU

activities and submit a written piece of work. A key component of the Level 6 aspect of the course is a learning journal, kept by the students, that is assessed at various points in different ways in both modules and in the *Learning Science Through Investigation* module. This includes submission of sections or digests of the journal and critical incidents tasks, which demonstrate the students' ability to reflect on their learning.

Science in Society

The tutor team decided at an early stage that the SKE students should explore science and society issues alongside learning the chemistry or physics content. This module explores some of the current issues in science that are likely to have a major impact on society, while also providing an opportunity to learn aspects of physics or chemistry concepts that might not be covered in other modules. In addition, this module provides the opportunity to develop some subject understanding of biology. Example topics explored include:

● development and use of smart materials;
● genetically modified organisms;
● evaluating risk and the tentative nature of scientific knowledge in decision-making;
● the Standard Model and the Large Hadron Collider at CERN.

Assessment is through an extended essay and an end-of-module examination.

Learning Science Through Investigation

Practical work forms the core of this module, which provides students with an opportunity to develop scientific and pedagogical understanding, and practical skills, through a series of extended laboratory-based physics or chemistry practical tasks. Assessment is through writing of practical reports and extracts from the ongoing learning journal.

Individual Project

To support the students in synthesising their subject knowledge and applying it in a pedagogical context, the *Individual Project* module involves the students individually choosing a scientific context and then developing a textbook chapter. The chapter is aimed at high-attaining GCSE/AS-level students and would support learners to develop knowledge of physics or chemistry applicable to that context. Example topics from recent cohorts include:

● scuba diving;
● amusement park rides;
● using nanotechnology in cancer diagnosis and treatment;
● poisons and pharmaceuticals.

The first assessment task requires the students to present their chapter approach and rationale to their peers. This allows the students to collect peer feedback on their chapter structure and how they

have chosen to break down the science topics, which they can use to produce their final printed chapter a couple of weeks later.

Teaching Assistant Placement

At the end of the course, the students spend 2 weeks in a school science department in a learning assistant role. This activity is not assessed (although a report is completed by the school) and allows the students to develop insights into the work of science teachers, and to relate their own learning to the misconceptions and difficulties experienced by pupils. With the support of the school-based mentor, some students may have the opportunity to teach part of a whole lesson (although this is not a requirement of the placement) and all students have the opportunity to use the experience to gather evidence towards achieving QTS during the PGCE year.

Lessons learned

The last 4 years have involved a learning journey for the LJMU Science Education team as we have engaged with the wide-ranging needs of SKE students. Some of the key lessons we have learned are:

● The students report that they see themselves as student *teachers* on a 2 year journey to achieving QTS, and not simply as students of physics or chemistry in preparation for subsequently learning to teach. We perhaps approached the first year of the SKE course as focusing on SMCK (in the language of Shulman (1986)) or *subject knowledge* (in the language of Banks *et al.* (1999)) and did not appreciate the extent to which the students wanted to learn *how* to teach as much as learning *what* to teach. One student explained that he was very aware of what he described as the 'oncoming train' of the PGCE course and having to help children learn the physics concepts he himself was struggling with.

● Developing subject knowledge *for teaching* requires integral consideration of pedagogical issues. The key additional step is to make these pedagogical issues *explicit*. We have found that the misconceptions held by SKE students are the same misconceptions held by pupils and non-SKE PGCE students. Articulating and analysing your own misconceptions is a vital aspect of developing your scientific understanding, and simultaneously requires critical consideration of how others might learn. For the lead author,

working with SKE students has brought home the full implications of what McEwan and Bull (1991) claimed about the pedagogic nature of subject knowledge and has led us to question how models such as those of Shulman (1986) and Banks *et al.* (1999) can be used to inform ITE practice.

● Student feedback shows that one of the most effective teaching and learning approaches used appears to be peer-teaching. This is carried out in various ways and will be explored in more depth in a later article. As the team has gained experience with each year of the SKE course, there has been more emphasis placed on the tutor and student group as a *community of practice* (Wenger, 1998). Activities where students, individually and in groups, take it in turns to teach peers have a powerful effect on students' understanding of fundamental science concepts through requiring them to focus on how to explain concepts to others. Some students have reported that they struggle to make effective use of their independent learning time and peer-teaching has helped some of them to address this. We have found through peer-teaching that many students prefer to focus on developing their understanding of a topic when they are also required to use it 'for real' in a teaching situation. (Lock *et al.* (2011) report that this seems to apply in general to most ITE students.) Peer-teaching appears to be effective partly because it leads to questions from peer-learners that reveal unanticipated misconceptions and enables the peer-teachers to practise explanations in a 'safe' and supportive setting. Supporting students to work in this way requires tutors to reflect on their role in scaffolding activities, modelling good practice explicitly and supporting students to reflect on how they learn.

● An increasing number of our student teachers are career-changers and some find it daunting to return to education after a period of employment. Formal evaluation evidence suggests that the SKE course allows students, who have valuable prior experiences to bring to the teaching profession, to boost their confidence through rediscovering and practising effective learning approaches.

● Comparison of final grades awarded for QTS Standard Q14 (the 'subject knowledge' Standard replaced by Standard 3 from September 2012) at the end of the PGCE course shows that SKE students' Q14 attainment is consistent with that of non-SKE PGCE students. For the 2011/12

academic year, over 80% of SKE and non-SKE PGCE students at LJMU were graded by schools as '1' against Ofsted criteria for Q14, with no students from either group receiving a '4'. We are not claiming that assessment of Q14 has been an in-depth analysis of a student teacher's subject understanding. However, Q14 was intended to be the threshold to be crossed to be awarded QTS as far as subject knowledge is concerned, and based on this criterion, school-based mentors do not generally identify SKE-route PGCE students as lacking sufficient subject knowledge. The extent to which this is dependent on the nature of the schools that accept SKE-route PGCE students on school-based placements, and the quality of

school-based mentoring they experience, is an aspect to be researched in depth by the LJMU team soon.

What next?

This article has been intended to provide some food for thought about science teacher subject knowledge and SKE courses through sharing one approach taken in one institution. In a future article we intend to discuss in more depth our experiences of using context-based and peer-teaching approaches, and to report the results of a research project to evaluate the effectiveness of the course and the progression of former SKE students in the teaching profession.

References

Banks, F., Leach, J. and Moon, B. (1999) New understandings of teachers' pedagogic knowledge. In *Learners and Pedagogy*, ed. Leach, J. and Moon, B. London: SAGE.

Department for Education (2010) *The Importance of Teaching: The Schools White Paper 2010*. Chapter 2. Available at: www.gov.uk/government/publications/the-importance-of-teaching-the-schools-white-paper-2010.

Department for Education (2012) *Teachers' Standards*. Available at: www.education.gov.uk/publications/standard/publicationDetail/Page1/DFE-00066-2011.

Department for Education (2013) Subject knowledge enhancement courses: www.education.gov.uk/get-into-teaching/subjects-age-groups/age-groups/teaching-secondary/boost-subject-knowledge/ske-courses.aspx?sc_lang=en-GB.

Ellis, V. (2007) Taking subject knowledge seriously: from professional knowledge recipes to complex conceptualizations of teacher development. *The*

Curriculum Journal, **18**(4), 447–462.

Hattie, J. (2009) *Visible Learning*. Abingdon: Routledge.

Lock, R., Salt, D. and Soares, A. (2011) *Acquisition of Science Subject Knowledge and Pedagogy in Initial Teacher Training*. London: Wellcome Trust. Available at: www.wellcome.ac.uk/About-us/Publications/Reports/Education/WTVM053203.htm.

McEwan, H. and Bull, B. (1991) The pedagogic nature of subject matter knowledge. *American Educational Research Journal*, **28**(2), 316–334.

Shulman, L. S. (1986) Those who understand: knowledge growth in teaching. *Educational Researcher*, **15**(2), 4–14.

van Driel, J. H., Verloop, N. and de Vos, W. (1998) Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, **35**(6), 673–695.

Wenger, E. (1998) *Communities of Practice*. New York: Cambridge University Press.

The authors all taught sciences in secondary schools before becoming teacher educators at Liverpool John Moores University. Michael Inglis is now based at the University of Leeds. Biographical and contact information can be found at:

Michael Inglis www.education.leeds.ac.uk/people/staff/academic/inglis

Andrea Mallaburn www.ljmu.ac.uk/ECL/122264.htm

Richard Tynan www.ljmu.ac.uk/ECL/122303.htm

Ken Clays www.ljmu.ac.uk/ECL/122443.htm

Robert Bryn Jones www.ljmu.ac.uk/ECL/122305.htm

Correspondence about this article should go to Michael Inglis at m.inglis@leeds.ac.uk.

TIMSTAR

WE WANT YOUR IDEAS

We at Timstar are seeking innovative product ideas from science teachers and technicians.

For more information contact our marketing department
on: science@timstar.co.uk

Subject knowledge enhancement (SKE) courses for creating new chemistry and physics teachers: do they work?

Richard Tynan, Andrea Mallaburn, Robert Bryn Jones and Ken Clays

ABSTRACT During extended subject knowledge enhancement (SKE) courses, graduates without chemistry or physics bachelor degrees prepared to enter a Postgraduate Certificate in Education (PGCE) programme to become chemistry or physics teachers. Data were gathered from the exit survey returned by Liverpool John Moores University SKE students about to start their science PGCE course. Lesson analysis and final report forms from the PGCE course and an early survey of first destinations were also analysed. Findings suggest that the 2011–12 SKE students valued their course highly. Many issues encourage caution when interpreting PGCE assessment information but, on summative assessment of subject knowledge and overall teaching, there was no statistically significant difference between the frequency of grades awarded to 2011–12 PGCE trainees who had followed a SKE route and those who entered the science PGCE directly. Early indications were that their employment rates in teaching were also similar.

It is difficult to question the notion that appropriate subject knowledge is fundamental to a person's capacity to teach. Its importance has been underlined in the training section of the guidance for the Government's new School Direct programme (Teaching Agency, 2012) where schools and providers are charged with a shared responsibility for developing the subject knowledge of trainee teachers.

An urgent subject knowledge issue for school science is the worrying downward trend in the number of university departments and/or the number of students enrolled on undergraduate courses associated with physics and chemistry (Breuer, 2002; Institute of Physics, 2012). This has contributed to difficulties in teacher recruitment in secondary schools and an unavoidable increase in the number of non-specialists teaching these subjects. One response to severe teacher shortages in key subjects such as the physical sciences has been to retrain graduates to teach shortage subjects. Since September 2008, Liverpool John Moores University (LJMU) has provided one-year full-time chemistry and physics SKE courses. Candidates suitable for teaching but with a need to improve their subject knowledge in chemistry

or physics were offered a place on a one-year SKE course leading to a Graduate Diploma. Successful completion of the SKE course was a condition for taking up the deferred place on our science PGCE the following year. Students on LJMU SKE courses in chemistry and physics who did not complete the course, or completed it but did not meet the assessment requirements, could not take up their PGCE chemistry or physics places. However, they could still apply to other PGCE providers on the basis of their first-degree qualifications that year. They could not re-apply to LJMU for a PGCE place in their first-degree subject until the next year because of the conditional nature of their original offer.

The purpose of this article is to examine data available from course documentation concerning the impact of the chemistry and physics SKE courses at LJMU and so to begin to evaluate their effectiveness. The questions it seeks to answer are:

- How did SKE students perceive their course as a preparation for PGCE?
- What did the judgements of teachers and mentors recorded on lesson analysis and final report forms indicate about the quality of subject knowledge demonstrated by SKE trainees?

- Were newly qualified teachers (NQTs) who followed the SKE route as employable as NQTs with a first degree in their subject specialism?

This small-scale quantitative study cannot claim to answer these questions other than within the context of the providing institution. However, by looking at numerical data generated by LJMU PGCE and SKE course documentation during the 2011–2012 academic year, interesting questions are raised that indicate some possible directions for future course development and for further qualitative investigation. The future of long SKE courses in subject shortage areas and their mode of delivery are under continued Government policy review. This study's findings are consistent with the most recent large-scale evaluations of SKE courses (Gibson *et al.*, 2013) and add to the debate on their future.

Initial teacher training/education in England

The current model for initial teacher training/ education (ITT/E) provision began in 1992 when partnerships between higher education institutions (HEIs) and schools became the established norm. Partnerships and collaborations among schools without HEI involvement have always been an acceptable alternative, with the first school-centred initial teacher training (SCITT) scheme starting in September 1993. These schemes signalled a strong motivation by the Conservative Government of the day to bypass HEIs in favour of on-the-job training for teachers and initiated the development of an assessment system that focused on specified teacher competencies (Barton *et al*, 1994).

The move towards the assessment of teacher competencies led to the qualified teacher status (QTS) standards and ITT/E requirements that apply to all ITT/E programmes. These arose out of the Education Act (HM Government, 2002) in 2003. Trainee teachers must meet all of the standards set down by the Government, including those concerning subject knowledge, and provide evidence that they have achieved a satisfactory level of performance to gain QTS. The latest revision of the *Teachers' Standards* took effect from 1 September 2012 and now applies to trainee teachers and qualified teachers throughout their careers (Department for Education, 2011).

Trainees currently spend the majority of their PGCE time in school (a recommended 120 out of 180 days) and must, at least, pass two training placements in different schools. School-based assessment of trainees against the *Teachers' Standards* is the responsibility of school-based tutors and professional mentors, although final outcomes are subject to quality assurance and moderation either internally or supported by university tutors if an HEI is in partnership with the school. Lesson observations are a crucial assessment mechanism to generate evidence of competence during school placements. Another purpose of assessing trainees during lesson observations is to be able to give accurate focused feedback. During developmental phases, formative feedback informs the training programme to facilitate trainee progress. At review points and at the end of training, assessment informs the trainers and eventually the QTS award body (at the time of writing the Teaching Agency) concerning the trainee's capacity to teach.

Since the start of school partnerships with HEI providers and the development of mentoring in schools, various challenges have been identified associated with assessing competencies and standards (Kerry and Shelton Mayes, 1995). Assessors need to be confident that their assessments are fully fit for purpose. Their assessments must support the inferences made from them, with all the associated outcomes and implications for pupils' learning and trainees' career progression that may arise. Assessment of trainee subject knowledge and overall teaching performance during school placement can be a highly contentious issue. Although assessment is evidence based, it relies on teacher, mentor, trainee and liaison tutor judgements of teacher competencies that are unavoidably subjective and use descriptors for standards that are open to interpretation. When disagreements occur, they can take time to resolve.

Stevens (2010) carried out a small-scale study examining course documentation from five PGCE courses as well as conducting student teacher interviews. The main study focus was the variety and timing of transitions experienced by trainees developing into teachers. A finding that has implications for those making overall judgements of competence highlighted occasions where better performance in one area was associated

with less effectiveness in another. For instance, greater confidence in managing the classroom and lesson delivery was often accompanied by less willingness to be innovative and take risks in the classroom. His interviews also highlighted the restricted range of reflections on subject knowledge made by trainees when evaluating their progress in this area.

Hager and Butler's (1996) model for the progress of members of occupations and professions describes three developmental levels or stages. For any profession or occupation, there are the knowledge, attitudes and skills that have to be acquired. Performance can then be simulated in practice domains, leading, eventually, to personal competence in real situations. Hager and Butler's arguments are then extended to discuss the characteristics and uses of the Scientific Measurement and the Judgemental Models of Assessment to gauge performance at these stages. These two models can be seen to form opposite ends of an assessment continuum. In their analysis, both can be demonstrated to satisfy, in different ways, general principles for assessment in higher education such as validity and reliability. Martin and Cloke (2000) later applied these two models to the assessment of QTS standards during ITT/E.

The Scientific Measurement Model uses assessment tools that measure attainments quantitatively. It emphasises objectivity, validity and reliability, and focuses on examinations taken under controlled conditions. On the other hand, the Judgemental Model seeks to infer competence through a qualitative approach. Using a variety of assessment events simulating life situations, it draws on multiple sources of direct evidence and emphasises avoiding bias through triangulation and the use of informed judgement (Hager and Butler, 1996).

Subject knowledge for teachers fits into this analysis as the first criterion for credible teaching and is an entirely appropriate target for assessment using a Scientific Measurement Model and traditional assessment tools. Trainees offered a place on a science PGCE are deemed to have sufficient prerequisite subject knowledge following analysis of their existing qualifications in science-based subjects where the assessment structures almost invariably have formal examinations as a major component. The purpose of the one-year SKE course is to enhance subject knowledge in physics or chemistry with a fully

examined and accredited Graduate Diploma for Intending Teachers in one of those subjects.

The application of a Judgemental Model of Assessment appears to be more appropriate to the simulation or practice and demonstration of competence stages. These stages can be equated with a trainee's experience on teaching placement and the NQT's reassessment during their induction year. However, it should be noted that during this study the assessment data for subject knowledge in chemistry and physics for comparing PGCE trainees on SKE and non-SKE pathways arises entirely from a Judgemental Model of Assessment. As such, the assessment data are the results of mentor and university tutor judgements against criteria for competencies based on evidence from lesson observations, planning materials and resources prepared by the trainee.

This approach is aimed at evaluating subject knowledge as it is applied in the classroom during teaching, learning and assessment. It cannot claim to assess chemistry or physics subject knowledge in the same way that an examination would. It does demonstrate the levels of confidence in trainee subject knowledge expressed by practising science teachers working with LJMU science PGCE trainees during the 2011–2012 academic year. The accuracy, reliability and validity of such assessment data can always be questioned, as these characteristics depend on the procedures in place for quality assurance and moderation of assessment. Given the issues associated with the assessment of competencies and standards, and of subject knowledge in particular, care is needed when attempting to interpret the data collected during this study. However, it should be remembered that these Judgemental Model style assessments remain and are the evidence on which all recommendations are made to award QTS or not.

Data and findings

The SKE student exit survey 2011–12
This group started their PGCE in September 2012, and 35 of 37 students (95%) returned their anonymous questionnaires concerning the SKE course in the previous academic year.

Students were asked for their level of agreement or disagreement with eight positive statements (Table 1) covering the key areas for which tutors were seeking feedback. Qualitative

Table 1 SKE student exit survey responses, May 2012; the two students (5%) who did not return their questionnaires are included in the "Did not answer" column percentage for each statement

Statement	% of cohort ($n = 37$)				
	Strongly agree	Agree	Disagree	Strongly disagree	Did not answer
I feel highly motivated to participate in my PGCE next year due to my involvement in my SKE course	81	11	0	0	8
The SKE course has developed my SKU throughout the year	78	16	0	0	5
The teaching on the SKE course has supported my learning well	70	22	0	0	8
I feel well prepared to embark on my PGCE course next year due to my engagement on the SKE course this year	62	30	0	0	8
I have enjoyed the content in the course this year	62	27	0	0	11
The feedback in the sessions has supported my development and understanding	51	43	0	0	5
The tutor(s) have given me sufficient support throughout the year	51	35	5	0	8
The feedback I have received regarding my assignments has supported my development and understanding	46	49	0	0	5

SKU = subject knowledge and understanding

statements were also collected but only for internal diagnostic use.

The response on exit from the course was over-whelmingly positive. All the respondents strongly agreed or agreed with the eight statements, apart from two students who disagreed with the statement that there had been sufficient support from their tutor. So, at the end of their course, most LJMU 2011–12 SKE chemistry and physics students expressed positive perceptions about the course content, its teaching, its impact on their learning and its role in preparing them for the PGCE year. The less positive responses centred on feedback and support but even in these areas the majority of students held positive viewpoints.

Assessment data from the science PGCE cohort 2011–12

Just over half (29) of the 2011–12 PGCE science cohort (53) entered the course directly from university or employment and the remainder (24) had just successfully completed a one-year LJMU SKE Graduate Diploma in Chemistry or Physics.

Both populations were routinely assessed using a Judgemental Model of Assessment (Hager and Butler, 1996) against the previously used Government teaching standard for subject knowledge (Q14) requiring that subject

knowledge and understanding and relevant pedagogy is secure for the age and ability range to be taught. Under the new *Teachers' Standards* (Department for Education, 2011), subject knowledge requirements are located mainly in standard S3, although aspects of subject knowledge arise in descriptors within standards S2 and S4.

Subject knowledge was assessed formatively throughout school placements by examining a variety of sources:

- lesson observations;
- planning forms;
- feedback on pupils' work;
- teaching resources prepared or modified by the trainee.

Summative judgements were recorded at review points during the course and finally at the end of the second school placement. It should be noted that the assessment of science subject knowledge was not divided into separate science subjects and the authors have no way of knowing exactly how many biology, chemistry or physics lessons actually contributed to any individual's assessment. However, those trainees designated as chemistry or physics PGCEs (SKE and non-SKE route) were placed with school science

Table 2 Formative assessment with written feedback given to science PGCE trainees during their last school placement in 2012 from a voluntary sample of 34 lesson observation forms

Teaching competency	% of observations assessed as			Mean % of assessments supported by written comments
	Working beyond or achieved	Working towards	Not evidenced or not assessed	
Subject knowledge and understanding	79	15	6	8
Classroom management and organisation (including behaviour management)	53	41	6	24
Suitability of resources	53	38	9	13
Teacher exposition	53	38	9	13
Lesson structure and focus	50	41	9	24
Pupil experience, interest and challenge	50	44	6	40
Planning and preparation	47	44	9	15
Personalised learning and differentiation	47	47	6	28
Homework/out of classroom	41	21	38	9
Monitoring, assessment and giving feedback	29	50	21	31

departments and mentors intending to provide a timetable rich in those subjects for the trainee.

At a final triangulation meeting, school mentors and university liaison tutors compared evidence from a variety of sources with the trainees' own portfolio of evidence. Competence was judged against Government teaching standards, including subject knowledge, in preparation for completing each trainee's final review form. Based on this assessment profile, the award of QTS and an overall teaching grade were recommended for successful trainees.

Information from lesson observation forms

During 2011–2012, schools, mentors and trainees were invited to take part in a project looking at assessments made during lesson observations and at the written feedback given on LJMU lesson observation forms. Participation was voluntary and anonymous. Thirty-four lesson observation forms were received during the final school placement. The LJMU partnership agreement requests that trainees are observed formally at least twice a week. Therefore, this represented a very small proportion of the potential number of forms that could have been returned. As such, the sample was too small to allow a comparison of SKE route and non-SKE route trainees or to be considered representative of the science mentors working with trainees in LJMU partnership schools. They do indicate the formative assessment and feedback

practices of some of the teachers working with 2011–12 PGCE science trainees during their last phase of training and the teachers' attitude towards subject knowledge. The assessment categories used on the lesson observation forms were:

- working beyond;
- achieved;
- working towards;
- not achieved;
- not evidenced.

Teachers, school-based tutors and professional mentors were required to judge trainees' performance against teaching standards or clusters of related standards (Table 2 shows the teaching competencies assessed) in the context of the phase of training and the trainees' experience at the time of the observation. This continues to be the Government's expectation:

Providers of initial teacher training (ITT) will assess trainees against the standards in a way that is consistent with what could reasonably be expected of a trainee teacher prior to the award of QTS. (Department for Education, 2011: 3)

There was no expectation that a judgement be recorded against every standard or cluster for each observation. Free response boxes provided the opportunity to support judgements with commentary and analysis that could be used as evidence.

PGCE training at LJMU is currently divided between time spent at university and two 12-week school experience placements. The school-based training is divided into three phases:

1 Orientation;
2 Beginning/developing teaching and classroom management;
3 Qualifying to teach.

The first placement emphasises trainee performance during phases 1 and 2. However, during the final placement in phase 3, the emphasis changes to developing strategies for maximising pupil performance and constructing a portfolio of evidence for the award of QTS. Judgements about pupil experience, interest and challenge, monitoring, assessment and giving feedback, and personalised learning and differentiation were most likely to be supported by written feedback. This reflected the planned course progression described above, as these clusters of standards were identified in trainee and mentor handbooks as focus areas for the second school placement. More fundamental teacher competencies developed in the first placement – classroom management and organisation (including behaviour management), and lesson structure and focus – formed the second tier of written comment.

In the lessons observed, subject knowledge and understanding (SKU) was most likely, by a considerable margin, to be perceived positively by teachers and assessed as achieved or working beyond and least likely to be assessed as working towards. It also shared the least rank with three other clusters of standards for no evidence or no assessment made. However, SKU was also the least likely standard or cluster to have any written comment or evidence accompanying the judgements made. A property shared with most of the other categories was that the more negative the assessment the more likely it was to be supported by written feedback (Table 2).

Although the LJMU PGCE course requires that judgements should be evidence based, one possible interpretation for the lack of written feedback on SKU in this small-scale survey of lesson observation forms is that in the second semester placement both the teachers and their trainees had reached consensus over what constituted acceptable and good SKU in science lessons and how to assess it. The teachers routinely and positively reinforced trainees' SKU more than other skills and competencies. Mistakes were still noted and discussed but these seem to have been the exception not the rule. As the participants did not have prior knowledge that SKU was a particular focus for the study, this could reflect their perceptions of the fundamental importance of SKU to aspiring teachers. It also suggests high levels of teacher confidence in its assessment.

Information from final review forms and results summary

In order to complete a trainee's final review form, the school-based mentor and a visiting university tutor applied QTS standards and Ofsted descriptors for final year students to lesson observations and portfolio evidence using a four-point scale:

1 outstanding;
2 good;
3 satisfactory;
4 fail.

Each teaching standard was graded separately and the trainee's profile of grades used to arrive at an overall teaching grade. Trainees must provide evidence for all standards to at least a satisfactory level to be awarded QTS and are routinely allowed to submit evidence gathered during the PGCE course from sources other than their final placement to strengthen the evidence base.

Final phase 3 review forms and the final results summary for the 2011–2012 cohort of PGCE trainees were used to compare the frequencies with which different final overall teaching grades and subject knowledge grades were awarded to those from the SKE and non-SKE routes. The cohort's results summary spreadsheet was used as the source for the overall teaching grade, as the grade recommended on the final review form can occasionally be amended later in the light of assessment of evidence in portfolios. Statistical correlations between final subject knowledge and overall teaching grades were also investigated. Non-parametric statistical tests of significance (Pearson's chi-squared and Spearman's rank correlation) were used to avoid any issues associated with small sample size and non-normal distributions in the data.

No statistically significant difference was found between the observed and expected

frequencies with which final grades 1 and 2 (or below) for subject knowledge or overall teaching were awarded to PGCE science trainees who had followed the SKE route or those who been accepted directly (Tables 3 and 4 and Box 1).

There was a strong positive correlation between the overall teaching grades and the subject knowledge grades within the whole PGCE group. The correlation was highly significant for non-SKE route (direct entry) trainees but among SKE route trainees the correlation was weaker (Table 5). In fact, in seven of the nine occasions when the SKE cohort final grades did not match, it was the subject knowledge grade that exceeded the overall teaching grade. Used in this context, the calculation of correlation coefficients does not seek to establish a causal link but to test the strength of a relationship which should already exist. The relationship between subject knowledge grades and overall teaching grades is prescribed by the assessment procedures followed at the end of the PGCE course described earlier. Subject knowledge is one of the standards judged separately by mentors before looking at the trainees' profile of strengths and areas for development and deciding a grade for overall

Table 3 Contingency table: final subject knowledge grades awarded to 2011–2012 PGCE science trainees

Grade	Observed (expected) counts		Total
	SKE route	**Non-SKE route**	
1	14 (15)	19 (18)	**33**
2 or below	6 (5)	6 (7)	**12**
Total	**20**	**25**	**45**

Chi-squared value 0.47; 1 degree of freedom but Yates's correction not necessary; critical value 3.84 (5% level); null hypothesis accepted

Table 4 Contingency table: final overall teaching grades awarded to 2011–2012 PGCE science trainees

Grade	Observed (expected) counts		Total
	SKE route	**Non-SKE route**	
1	8 (9)	13 (12)	**21**
2 or below	12 (11)	12(13)	**24**
Total	**20**	**25**	**45**

Chi-squared value 0.36; 1 degree of freedom but Yates's correction not necessary; critical value 3.84 (5% level); null hypothesis accepted

---

BOX 1 Using the chi-squared test and contingency tables to compare grades

For those not familiar with the chi-squared test and contingency tables, the observed count totals can be used to calculate the expected numbers of SKE route PGCE science trainees gaining a grade 1 in each contingency table. For Tables 3 and 4, the calculation from first principles would be:

probability of being an SKE route trainee (p) = SKE route total/all PGCE total

probability of gaining grade 1 (q) = grade 1 total/all PGCE total

expected number of SKE route trainees awarded grade 1= all PGCE total $\times p \times q$

The other expected values are obtained by subtracting this calculated figure from the appropriate row and column totals. For example, in Table 4, if 9 out of a total 20 SKE route PGCE trainees were expected to gain a grade 1 for overall teaching then 11 could be expected to get grade 2 or below.

The chi-squared value is obtained by summing the results of the calculation (observed − expected)^2/expected for each pair of observed and expected counts in the contingency table. The number of degrees of freedom for a contingency table is defined as the (number of rows − 1) × (number of columns − 1). So for

Tables 3 and 4 it is 1. If there is only 1 degree of freedom when the chi-squared value is calculated, Yates's correction (subtract 0.5 from the difference between the observed and expected count regardless of sign before squaring) may be applied particularly if there are expected counts of less than 5 in any part of the table. This was not necessary and it could be argued that the calculated expected values were so close to the observed counts in Tables 3 and 4 that a statistical test was redundant. However, the point of applying a statistical test of significance is to remove any subjectivity in drawing conclusions about the raw numerical data.

Table 5 Spearman's rank correlation two-tailed test on final overall teaching grades and subject knowledge grades for 2011–2012 LJMU PGCE science trainees

Trainees	Correlation between final overall teaching grade and subject knowledge grade			
	Correlation coefficient	n	Significant	Probability of error
All science PGCE	0.639	45	Highly	<1% (critical value 0.382)
All non-SKE route (direct entry)	0.726	25	Highly	<1% (critical value 0.511)
All SKE route	0.536	20	Yes	5% (critical value 0.447)

teaching. The weaker correlation among SKE route trainees could suggest that their subject knowledge grade contributed less to the decision about their overall teaching grade than it did for non-SKE route trainees. It could also indicate a distinction in the minds of assessors between knowledge of a subject and knowledge of how best to teach it.

LJMU 2011–12 PGCE science trainees' first destinations

Information on the destinations of NQTs after their course updates continually but the exit survey indicated that around half the SKE route PGCE science trainees had been successful in obtaining a first teaching position. At first glance, the data suggest that their reported success rate was better than that of the direct entrants to the science PGCE (Table 6). This might be thought a reasonable outcome because SKE applicants have their first degree specialism as well as their SKE subject Graduate Diploma to offer their prospective employer together with a year's extra training and experience to illuminate their responses at interview. However, percentages can be misleading. Chi-squared contingency table analysis of the differences between the observed and expected frequencies of SKE route and non-SKE route PGCE graduates notifying LJMU of successful job applications indicated that this difference was not statistically significant (Table 7).

Although it is probable that those who did not respond to the survey had no teaching post to report, this may be a false assumption. A cautious conclusion would be that there were no indications of a difference between the early reports of success in gaining employment in schools between the two PGCE routes.

Discussion

The nature of subject knowledge for teachers is still a matter of debate but a commonly cited

Table 6 Early notifications of destinations for 2012 PGCE science graduates

Graduates	Destinations (summer 2012)		
	Teaching job	Did not respond	Left teaching
Non-SKE route (direct) entrants			
Applied science	2	4	0
Biology	2	4	0
Chemistry	7	4	0
Physics	3	2	1
Total	14 (48%)	14 (48%)	1 (3%)
SKE route entrants			
Chemistry	10	6	0
Physics	4	4	0
Total	14 (58%)	10 (42%)	0
All PGCE science			
Total	28 (53%)	24 (45%)	1 (2%)

Table 7 Contingency table: early notifications of destinations for 2012 PGCE science graduates

First destination	Observed (expected) counts		Total
	SKE route	Non-SKE foute	
Teaching	14(13)	14(15)	28
Other	10(11)	15 (14)	25
Total	24	29	53

Chi-squared value 0.31; 1 degree of freedom but Yates's correction not necessary; critical value 3.84 (5% level); null hypothesis accepted

model is that proposed by Shulman (1986), who made the distinction between subject matter content knowledge (SMCK), pedagogical content knowledge (PCK) and curricular knowledge (CK). This approach is reflected in the relevant standards set down (Department for Education, 2011). SMCK and CK appear in standard S3 and PCK appears in standards S2 and S4.

Since their introduction, the use of competencies and standards in ITT/E has never been seriously challenged by practitioners in schools or higher education. However, some researchers have expressed concerns about relying solely on the use of teaching standards criteria to judge and accredit new teachers:

Although the standards can be useful as criteria for judging the abilities and attainment of beginning teachers, this article contends that the model of teaching this list presents is impoverished. (Turner-Bisset, 2006: 40)

To illustrate her thesis, Turner-Bisset (2006) presented a detailed and comprehensive discussion of the PCK demonstrated by a history teacher during a lesson. She then sought to provide a model to supplement the ideas of subject knowledge competency implicit in the QTS standards.

The ability to reliably assess subject knowledge and other teacher competencies with consistency across and within PGCE providers begins with the provision of clear descriptors or criteria. The new *Teachers' Standards* (Department for Education, 2011) seek to provide sufficient clarity about what is required of teachers but continue to give little guidance on how standards should be interpreted or assessed by those involved in teacher training or appraisal.

Leshem and Bar-Hama (2008) debated the issues surrounding the use of teaching competencies and criteria compared with holistic assessment during teacher training in Israel. Their study found that students initially perceived lessons analytically but saw quality as the sum of the parts. The students needed clear criteria and disliked assessment based on unknowns. Students felt criterion-based assessments were valid and shared their tutors' view that impressionist marking was subjective and unreliable. The analytical use of criteria was considered valuable for all students during feedback sessions to focus and aid discussion. However, students still expressed a preference for holistic assessment when summative judgements were made during observations. PGCE trainees in England must provide evidence that they have reached a satisfactory level of competence in all the *Teachers' Standards* (Department for Education, 2011) in order to be recommended for QTS. This implies that an analytical approach

to their assessment should be taken. However, in the absence of clear assessment guidelines to accompany statements of standards, this may not be the case for individual standards or judgements of overall teaching ability. This is an issue that reduces confidence in trainee assessment outcomes.

It was beyond the scope of this study to investigate the consistency of approach and assessment methods used to assess trainees on school placement across LJMU–school partnerships. The study gave no indication of the way university tutors, trainees, teachers and mentors viewed the various aspects of science subject knowledge for aspiring teachers or how to assess it. Nor did it investigate the priority assigned by assessors to knowing how to teach the subject (PCK) compared with knowing the subject discipline in the first place (SMCK). However, while accepting the limitations of a quantitative investigation of course documentation as a research methodology, it is still possible to make tentative conclusions and recommendations useful in the context of ITT/E at LJMU.

If the data collected from the PGCE science trainee lesson observation forms (Table 2) are typical across the LJMU partnerships, it would suggest that trainee subject knowledge was assessed during most lesson observations and usually positively reinforced. Written feedback tended to be given when subject knowledge was not adequate. It was not clear whether assessors separated chemistry, physics and biology subject knowledge or assessed science as a single entity. It is important to acknowledge that teachers, mentors, tutors and trainees will differ in the way they define and then assess subject knowledge. A qualitative study of teacher feedback would be required to illuminate these issues and enable the evidence used when making judgements to be investigated.

There is a suggestion from the correlation coefficient calculations (Table 5) that mentors may have distinguished between SMCK and PCK and reflected this in judgements of SKE route trainees' overall teaching grade. The correlation between final overall teaching and subject knowledge grades was weaker for SKE trainees, and where different the subject knowledge grade usually exceeded the overall teaching grade. Using Shulman's (1986) model, the purpose of SKE courses is to teach SMCK not PCK or CK. It is difficult to separate these completely and,

in response to student feedback in the first years of the SKE course at LJMU, delivery of PCK and CK has been increased. There may be a case for an even more integrated approach to SKE to ensure the other aspects of subject knowledge for teachers, apart from SMCK, are developed to a similar level in the first year.

In summary, SKE route trainees were very positive about their course and their level of preparation for PGCE and it appears from assessment data that this confidence is justified in terms of PGCE course outcomes. Teacher assessments of final overall teaching and subject knowledge grades for the 2011–12 PGCE cohort indicated that SKE route overall teaching and subject knowledge was perceived to be of a

similar standard to that of direct entry trainees. The weaker correlation between final overall teaching and subject knowledge grades for SKE route trainees suggests, at least, that other competencies or PCK may have had a greater impact on this assessment outcome. SKE route PGCE graduates were not found to be at any disadvantage on early returns when seeking a job for their induction year. These findings support the view that the LJMU SKE chemistry and physics Graduate Diploma courses are capable of succeeding in their aim to equip more science graduates with the subject knowledge that enables them to find employment and teach these shortage subjects.

References

Barton, L., Barrett, E. Whitty, G., Miles, S. and Furlong, J. (1994) Teacher education and teacher professionalism in England: some emerging issues. *British Journal of Sociology of Education*, **15**(4), 529–543.

Breuer, S. W. (2002) Does chemistry have a future? *University Chemistry Education*, **6**(1), 13–16.

Department for Education (2011) *Teachers' Standards: Statutory Guidance for School Leaders, School Staff and Governing Bodies*. Department for Education. Available at: www.gov.uk/government/publications/teachers-standards.

Gibson, S., O'Toole, G., Dennison, M. and Oliver, L. (2013) *Evaluation of Subject Knowledge Enhancement Courses: Annual Report – 2011-12. Research Report*. Department for Education/CooperGibson Research. Available at: www.gov.uk/government/publications/evaluation-of-subject-knowledge-enhancement-courses-annual-report-2011-to-2012.

Hager, P. and Butler, J. (1996) Two models of educational assessment. *Assessment and Evaluation in Higher Education*, **21**(4), 367–378.

HM Government (2002) *Education Act 2002*. Available at: www.legislation.gov.uk/ukpga/2002/32/contents.

Institute of Physics (2012) *The Future of Higher Education in England: Institute of Physics response to an IPPR*

call for evidence. London: IOP. www.iop.org/policy/consultations/higher_education/file_58559.pdf.

Kerry, T. and Shelton Mayes, A. ed. (1995) *Issues in Mentoring*. London: Open University/Routledge.

Leshem, S. and Bar-Hama, R. (2008) Evaluating teaching practice. *ELT Journal*, **62**(3), 257–265.

Martin, S. and Cloke, C. (2000) Standards for the Award of Qualified Teacher Status: reflections on assessment implications. *Assessment and Evaluation in Higher Education*, **25**(2), 183–190.

Shulman, L. S. (1986) Those who understand: knowledge growth in teaching. *Educational Researcher*, **15**(2), 4–14.

Stevens, D. (2010) A Freirean critique of the competence model of teacher education, focusing on the standards for qualified teacher status in England, *Journal of Education for Teaching: International Research and Pedagogy*, **36**(2), 187–196.

Teaching Agency (2012) *A Guide to School Direct 2013–14*. Teaching Agency. Available at: http://media.education.gov.uk/assets/files/pdf/a/a%20guide%20to%20school%20direct%202013%202014%20v041012.pdf.

Turner-Bisset, R. (2006) The knowledge bases of the expert teacher. *British Educational Research Journal*, **25**(1), 39–55.

Rick Tynan is a Senior Lecturer in Science Education (Biology) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: R.J.Tynan@ljmu.ac.uk

Andrea Mallaburn is a Senior Lecturer in Science Education (Chemistry) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: A.Mallaburn@ljmu.ac.uk

Robert Bryn Jones is a Senior Lecturer in Science Education (Chemistry) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: R.B.Jones@ljmu.ac.uk

Ken Clays is a Senior Lecturer in Science Education (Physics) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: K.Clays@ljmu.ac.uk

Working towards evidence-based practice in science teaching and learning

Rick Tynan, Robert Bryn Jones, Andrea Mallaburn and Ken Clays

ABSTRACT High-performing international education systems integrate evidence-based practice into their initial teacher education programmes. It is the authors' experience that the usefulness of education research to education practitioners is not always easy to judge and this leads to a justifiably cautious approach to evidence-based practice among trainee science teachers and their mentors in schools. An example of informal practitioner research is described and discussed. This involved using a science in society or socio-scientific approach to deliver a science subject knowledge module to two different cohorts of intending science teachers. The module was taught separately to 22 undergraduate students in their final year of a Primary/Secondary Education Honours degree with Qualified Teacher Status, and to 50 students following Graduate Diploma Subject Knowledge Enhancement courses in chemistry and physics who were preparing to take up places on science Postgraduate Certificate in Education courses. The aim was to demonstrate strategies for facilitating the development of critical thinking and scientific literacy in school science lessons. The use of anonymous voting devices during sessions indicated a polarisation of opinions among participants, rather than a more considered or critical response to the scientific questions. This discussion seeks to illustrate the value and drawbacks of informal practitioner research and how this evidence-based approach might be beneficial to teaching and learning in science.

Introduction

The interim report (BERA-RSA, 2014) of the joint enquiry by the British Educational Research Association (BERA) and the Royal Society for the Encouragement of the Arts, Manufactures and Commerce (RSA) seeks to establish a clear link between school improvement and research-based activity during initial teacher training and continuing professional development (Mincu, 2013). Burn and Mutton (2013) have surveyed the range of relationships between research and initial teacher education in selected international education systems, seeking links between pupil performance and evidence-based clinical practice in education. The report identifies some common attributes shared by high-performing international education systems. These include the provision of high-quality teacher education that progressively develops research skills and the ability to engage critically with evidence. However, the published requirements for qualified teacher status across the UK indicate a difference in attitude towards the role of research in teacher education (BERA-RSA, 2014). Whereas in Scotland and Northern Ireland the importance of research and evidence-based teaching is explicitly emphasised, in Wales and England this is only implied (Department for Education, 2011a; General Teaching Council for Northern Ireland, 2011; General Teaching Council for Scotland, 2012; Welsh Government, 2011).

Ben Goldacre, who is well known for the *Bad Science* column in the *Guardian* and his blog at www.badscience.net, responded to a request from the Department for Education to produce a discussion document (Goldacre, 2013) on the need for teachers to understand the importance of evidence-based teaching and learning and the strengths and limitations of quantitative and qualitative research methods in education. With some success, Goldacre has advocated (Department for Education, 2013) the increased use of randomised controlled trials in education in order to generate high-quality quantitative data to answer questions about what works in schools and what does not. Among Goldacre's considerations is that teachers should be encouraged to participate in large-scale quantitative research, while using their own

small-scale qualitative research to help identify the ideas that need examining. Goldacre argues that this is necessary to counter the current state of affairs in which education policy and practice is vulnerable to the influence of senior, sometimes charismatic, people who claim to have answers to challenges in schools, even when these are not based on significant evidence. Also, much of the small-scale qualitative research referred to by Goldacre (2013; Department for Education, 2013) is currently undertaken as part of professional development programmes of one sort or another. There is no infrastructure for following up research aimed at qualifications with larger scale studies that have more scope to inform and influence practice. As this article is concerned with a small-scale and informal piece of practitioner research, it illustrates many of the arguments raised in Goldacre's paper.

The learning, teaching and assessment strategies investigated

Science in Society, a subject knowledge module, was taught to 22 undergraduate students in their final year of a Primary/Secondary Education Honours degree with Qualified Teacher Status. Later during the same academic year, this module was also taught to a cohort of 50 postgraduates as part of Graduate Diploma Subject Knowledge Enhancement courses for intending teachers of chemistry and physics. The module was designed to enhance science subject knowledge and understanding and also place it in a technological and social context. The informal research was carried out during the topic, Genetically Modified Organisms (GMO). Learning, teaching and assessment (LTA) strategies that students were likely to need while on school experience placements were modelled in the teaching. This acknowledged the role of pedagogical content knowledge (Shulman, 1986) as a component of subject knowledge for teachers.

The format for each topic in the module was similar in that a lecturer introduced the scientific concepts involved and then set group tasks that required further research in preparation for the next session. For the GMO topic, groups were allocated and then assigned to one of a range of stakeholder roles. Each group then had a week to research GMO from that particular perspective and make the case for or against the genetic modification of organisms for agricultural or

medical use during a mock public enquiry debate in the second session. This strategy was intended to facilitate skills involved in the explanation, feedback and dissemination of group research outcomes to the whole cohort and to allow the student teachers to experience strategies that could improve scientific literacy and develop critical approaches to evidence in secondary school learners.

Some specific benefits of using role play and debate as a group activity with pupils were proposed by Simonneaux (2001), whose work suggested that developing critical thinking and scientific literacy are reasonable learning objectives for this strategy:

- understanding the complexity of decisions involving social issues;
- understanding relevant scientific principles;
- expressing, defending and/or criticising viewpoints;
- distinguishing between statements based upon evidence and those based upon values;
- evaluating evidence.

In his meta-study of 54 articles, Cavagnetto (2010) identified three main approaches to teaching pupils scientific argumentation in order to improve their scientific literacy. The structural approach included activities that emphasise the structure and practice of scientific argument and debate in comparison with other sorts of argument, such as political or legal. Immersion approaches included all activities where the emphasis was on scientific method and investigation, and learners were encouraged throughout to talk about the scientific process and evaluation of its outcomes. The use of group work and role-play debate, described above for the GMO topic, fell easily within the scope of the third, science in society or socio-scientific, approach, with activities that set scientific arguments in moral, ethical and political contexts.

Hand-held voting devices known as 'clickers' were used in conjunction with *TurningPoint 2008* software (Turning Technologies) to survey attitudes and display the results at various points during the topic. Compared with a show of hands or an assessment for learning (AfL) strategy such as 'traffic lights' or 'wipe boards', mobile voting devices have the perceived benefits of engaging students and being anonymous. Anonymous voting can encourage participation and honest

attempts to answer questions. Another pedagogic advantage for the teacher is that the results of the survey can be displayed instantly as a chart for discussion and can be saved for future reference. It is the authors' opinion that this addresses one issue associated with many less technology-dependent strategies: how to record formative assessment outcomes for diagnostic use and so inform future planning.

Voting occurred at the start and end of the topic for the undergraduate cohort but, for the postgraduate group, opinions were also sought just before the group work began. The same Likert-type items were used on every occasion. Using a 5-point scale, voters were asked to submit anonymously their level of agreement or disagreement with three statements. These were about the safety of GMO technology and its use in the fields of agriculture and health. The authors' intuition was that increased scientific literacy and critical thinking skills might be reflected in the students' anonymous voting behaviour through fewer voters expressing strong agreement or disagreement and more neutral votes.

The possible implications of the survey results for refining science pedagogy were discussed with the student participants and colleagues. A limited literature search was also conducted within the time constraints allowed. This is consistent with a systematic approach to critical reflection using Brookfield's Four Lenses (Brookfield, 1998, 2002).

Concerning methodology

The various guises that practitioner research can take are set out and discussed by Burton and Bartlett (2005), for whom a working definition of educational research would be research motivated by the need for improvements in LTA rather than to make advances within a subject discipline. The informal investigation described briefly above demonstrates elements of action research and case study, but satisfies neither approach completely.

Action research in the classroom consists of repeated cycles of planned interventions based upon previous observations. Each intervention is evaluated for its impact to inform the planning for the next (Baumfield, Hall and Wall, 2008). In the GMO topic, the intervention can be considered to be the use of role-play debate. Analysis of the data gathered when using the clickers provided an indication of the impact of the LTA strategies

adopted on student engagement with the topic and their attitudes to the use of GMO. This also gave clues to their use of critical thinking and demonstration of scientific literacy skills. This allowed an evaluation of the effectiveness and fitness for purpose of the role-play debate approach and informed future refinements to the LTA strategy. An extra opinion survey was conducted with the second cohort before the start of the role-play debate, so that, if the change in voting pattern was repeated, it might be pinpointed more closely to the first or second session activities.

Using Bassey's (1999) reconstructed approach to educational case studies, the authors believe that the investigation meets the criteria for an educational case study in the following ways:

- The research can be described as educational because it investigated a specific LTA strategy and evaluated its fitness for purpose;
- It was empirical and natural because it used anonymous cohort data collected in contact sessions during the planned LTA activities associated with the topic;
- It was concerned with a singularity involving a set of instances clearly limited by time and locality;
- The area had relevance for practitioners and students and was of interest to them;
- The study was informative and generated cautious conclusions and recommendations for future improvements in the LTA strategy.

However, also using Bassey's (1999) guidelines, it should be noted that, as a case study, the investigation could be questioned in a number of areas. Was there sufficient data to have confidence that all the significant features of the case were identified? Plausible explanations could be constructed based upon the data but, without further qualitative data-gathering, they can be considered as neither fully trustworthy nor fully convincing.

Results

The undergraduate group vote at the start of the topic demonstrated a range of opinion skewed towards caution with respect to the potential safety of GMO (Figure 1), but also in favour of making GMO foods available to the public (Figure 2). The use of GMO for medical purposes was strongly supported (Figure 3). After the topic
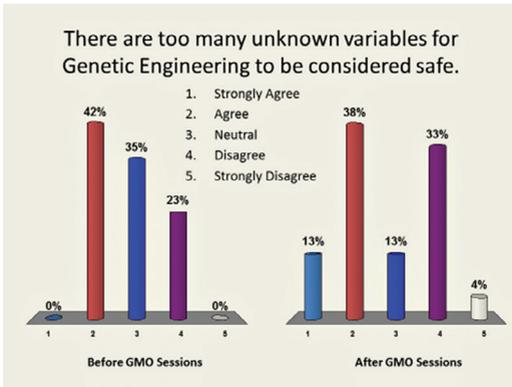
Figure 1 Confidence in the safe use of GMOs: voting results slide for the undergraduate cohort (*n* = 22) before and after the topic was taught
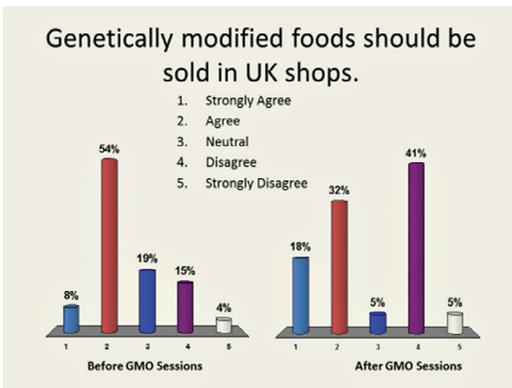
Figure 2 Use of GMOs in the human food chain: voting results slide for the undergraduate cohort (*n* = 22) before and after the topic was taught
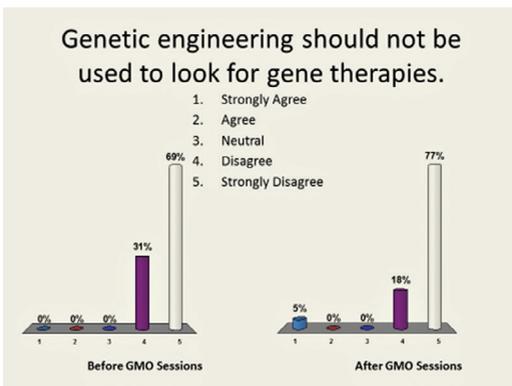
Figure 3 Use of GMOs for medical purposes: voting results slide for the undergraduate cohort (*n* = 22) before and after the topic was taught

was completed, the distribution was bimodal for the first two questions, apparently due to more 'neutrals' choosing a side (Figures 1 and 2). Attitudes to the medical use of GMO remained positive, apart from a small number now expressing strongly opposing views (Figure 3). Experiencing the activities associated with this topic had apparently polarised the opinions expressed by the group.

The teaching of this module and topic to a second cohort provided an opportunity to amend the delivery slightly, by adding another opinion survey point midway through the topic before group work began. Opinions did change during the first, information-based session, but the number expressing neutrality was much reduced after the group work component and debate and even reversed the trend demonstrated by the first two surveys (Figures 5 and 6). In short, the same polarisation of viewpoints after the group work was observed at different times with both cohorts (Figures 4, 5 and 6).

The voting results were discussed by the students during plenary sessions and the effect of rational and irrational influences on decision-making was an issue raised. One irrational influence noted by the students during group work was conformity to group aims and values. However, it was not clear how this might affect anonymous voting. These are the strategies that students observed during debates and additional ploys suggested later in discussion: ignoring contrary evidence and only reporting supporting evidence, distorting evidence, misrepresenting the opponents' cases, ignoring established causal relationships or reporting fictitious ones, and deliberately mixing up cause and effect.

Discussion

Most teachers routinely conduct informal research. In order to improve their pupils' learning, they observe the impact of what they do on their learners' progress, consult their pupils, question trusted peers and read up on areas of interest. In this way, they are applying the principles of critically reflective practice (Brookfield, 1998, 2002) and conducting informal research (Burton and Bartlett, 2005). They have little time to allocate to more formal research involving formal publication or to contributing to large-scale research projects.
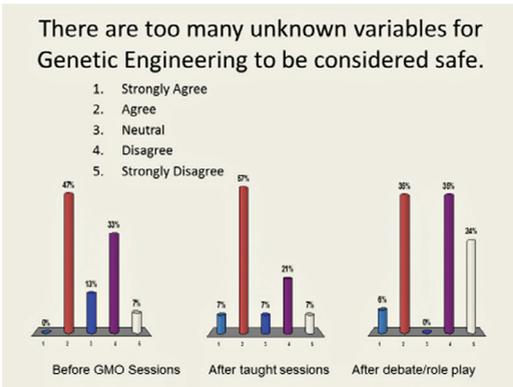
Figure 4 Confidence in the safe use of GMOs: voting results slide for the postgraduate cohort (*n* = 50) at the start of the topic, before group work and after the topic
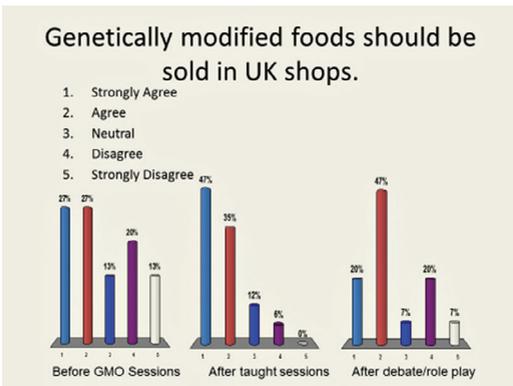
Figure 5 Use of GMOs in the human food chain: voting results slide for the postgraduate cohort (*n* = 50) at the start of the topic, before group work and after the topic
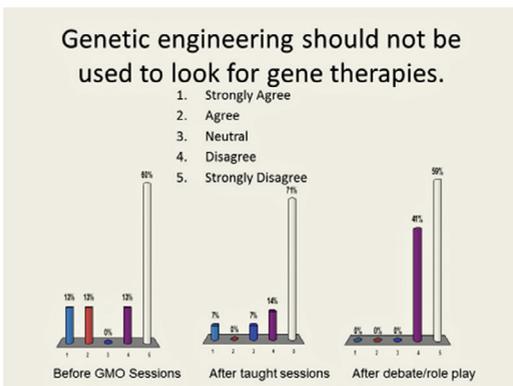
Figure 6 Use of GMOs for medical purposes: voting results slide for the postgraduate cohort (*n* = 50) at the start of the topic, before group work and after the topic

Concerning Likert scales

This research used a 5-point Likert scale to indicate the direction and intensity of respondents' opinions. The issues surrounding the use of such scales have been much researched and discussed. Arguments that Likert items cannot be used as interval scales and should be considered unsuitable for use with parametric statistics were robustly countered by Norman (2010). Leung (2011) recommended that social workers and teachers use 11-point scales mainly to increase sensitivity. Cummins and Gullone's (2000) earlier influential work explored the use of different Likert scales and favoured a 10-point scale. However, Leung's study (2011) found little difference in the psychometric properties and statistical behaviour of different-sized Likert scales. Leung (2011) also considered briefly the argument over odd- or even-numbered Likert scales. Even-numbered Likert scales have no neutral category. This may reduce certain sorts of bias, but may also distort the data in other ways. For instance, forcing neutral voters to choose a directional response can hide true opinions about complex or sensitive issues.

This study made use of the visual presentation of voting behaviours based upon a smaller, odd-numbered Likert scale with a neutral point. Future data gathering could be planned to include statistical analysis and more interval points, but an odd-numbered scale is probably still recommended for surveying opinion on complex topics such as GMO. However, as the same questions were used on each occasion, the possible questionnaire effects caused by the Likert scale or item wordings selected were common to each survey. It is, therefore, a reasonable step to question the assumptions made about the participants' voting behaviour and/or those concerning the expected learning outcomes for the LTA strategy.

Concerning the LTA strategy

In Cavagnetto's (2010) view, all approaches to teaching science that develop science argumentation skills will lead to improvements in pupils' communication skills, metacognition and critical thinking. However, Cavagnetto's study suggested that, while a socio-scientific approach can provide authentic contexts for science learning, adopting approaches with activities emphasising immersion in the process of science may be the most effective way to develop all the aspects of scientific literacy. Cavagnetto (2010)

also highlighted the unique competitive but collaborative nature of scientific argumentation. In role play, participants can adopt more adversarial styles of argument in order to win the debate. The polarised voting viewpoints observed at the end of the second session might indicate responses to individual and group debating performance rather than a more critical evaluation of the evidence provided. The students observed that the competitive nature of debates had led some participants to employ unscientific tactics and strategies in order to secure 'a win'. This increased the levels of engagement for participants, but the implications for the teacher running the activity are to be aware of these debating strategies in advance and how to manage them. Depending on the learners involved, unscientific styles of argumentation and debating ploys could either be disallowed during the debate, or allowed and discussed fully during subsequent debriefing. Trying out these alternatives would be future planned interventions in the action research cycle.

Concerning methodology

One cycle of action research was completed with each cohort and, to date, there have been no further opportunities to test the future interventions suggested above to improve the use of socio-scientific LTA strategies and evaluate them.

Taking a socio-economic approach to complex scientific issues such as GMO requires learners to consider multiple perspectives using background knowledge and understanding of a range of scientific concepts. For both cohorts of aspiring teachers, it arose independently that the group work and debate were associated with a marked polarisation of views during anonymous voting. This was made more apparent by the instant visual display of the voting results within the *PowerPoint* presentation and the ability to compare responses over time using previously saved results. The results of the voting provided a stimulus for lively discussion about the use of role-play debate as an LTA strategy, together with a consideration of the assumptions that practitioners made about the purpose and outcomes expected from this style of group work. This reminded us that accepted pedagogical practice is often based upon assumptions that may or may not appear to agree with practitioners' observations.

Our observations were gathered during teaching and learning activities with intending

teachers and were used in the discussion of science pedagogy and its fitness for its intended purpose. The voting response data were examined informally during student feedback discussions and our explanations were later found to be plausible in the context of the peer-reviewed studies consulted. To some extent, this is triangulation of evidence and at least consistent with the aim of developing critically reflective evidence-based practice envisaged by Brookfield (1998, 2002). As such, the findings had considerable utility at a local level. However, in terms of more formal research considerations, further planned investigation would be needed in order to evaluate the reliability (repeatability) of the data gathered, their validity and whether any generalisations could or should be made from them.

The survey results raised many questions, opening up several avenues for possible further work. It would have been interesting and informative to investigate links between short-term voting behaviours and long-term learning and the individuals' reasons for maintaining or changing their opinions. Whether voting behaviours in the same context would be similar or different in other groups of adult or school-age learners could only be answered by systematic observations on a much larger scale beyond the resources of practitioner researchers.

Implications for action research in schools

In applying Brookfield's Four Lenses (Brookfield, 1998) with two groups of aspiring science teachers, this informal research may or may not have indicated something worthy of further investigation. However, given the time constraints placed upon educators, it is unlikely that this or similar informal research in schools will be taken forward in a more formal context. Formal and informal small-scale research can be powerful improvement tools in schools for responding to issues and opportunities quickly and effectively, but does not provide the large-scale quantitative data required by centralised policy-generating bodies. Goldacre would argue (2013; Department for Education, 2013) that this means that greater emphasis needs to be placed on schools participating in large-scale quantitative studies, preferably in the form of randomised controlled trials.

In the drive for school improvement, Masters-level study and schools' engagement with evidence- and research-based learning and teaching

continue to be highlighted and encouraged as important components of initial and continuing teacher education (Department for Education, 2011b). The small-scale practitioner research associated with this has great potential to inform and guide large-scale research. The international evidence appears to indicate that both approaches are needed to inform practice with research-based evidence:

That research – be this delivered or stimulated by external interventions or through on-site collaborative inquiry processes – is a vital component of a school's capacity for self-improvement, and that such research is likely to play a vital role in ensuring that effective teaching and learning processes are in place. (Mincu, 2013: 2)

References

Bassey, M. (1999) *Case Study Research in Educational Settings*. Milton Keynes: Open University Press.

Baumfield, V., Hall, E. and Wall, K. (2008) *Action Research in the Classroom*. London: SAGE Publications.

BERA-RSA (2014) *The Role of Research in Teacher Education: Reviewing the Evidence*. Interim Report of the BERA-RSA Inquiry. British Educational Research Association (BERA).

Brookfield, S. (1998) Critically reflective practice. *Journal of Continuing Education in the Health Professions*, **18**(4), 197–205.

Brookfield, S. D. (2002) Using the lenses of critically reflective teaching in the community college classroom. *New Directions for Community Colleges*, **2002**(118), 31–38.

Burn, K. and Mutton, T. (2013) *Review of 'Research-Informed Clinical Practice' in Initial Teacher Education*. The BERA-RSA Inquiry. British Educational Research Association (BERA).

Burton, D. and Bartlett, S. (2005) *Practitioner Research for Teachers*. London: SAGE Publications.

Cavagnetto, A. R. (2010) Argument to foster scientific literacy: a review of argument interventions in K-12 science contexts. *Review of Educational Research*, **80**(3), 336–371.

Cummins, R. A. and Gullone, E. (2000) Why we should not use 5-point likert scales: the case for subjective quality of life measurement. *Proceedings of the Second International Conference on Quality of Life in Cities*, pp. 74–93. Singapore: National University of Singapore.

Department for Education (2011a) *Teachers' Standards: Guidance for School Leaders, School Staff and Governing Bodies*. Available at: www.gov.uk/government/publications/teachers-standards.

Department for Education (2011b) *Training Our Next Generation of Outstanding Teachers: An Improvement Strategy for Discussion*. Available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/175363/DFE-00054-2011.pdf.

Department for Education (2013) *Randomised Controlled Trials (RCTs) Will Raise the Bar in Education and Children's Services by Increasing the Use of Quantitative Evidence*. Available at: www.gov.uk/government/news/new-randomised-controlled-trials-will-drive-forward-evidence-based-research.

General Teaching Council for Northern Ireland (2011) *Teaching: The Reflective Profession*. Available at: epublishbyus.com/ebook/ebook?id=10020354.

General Teaching Council for Scotland (2012) *The Standards for Registration: Mandatory Requirements for Registration with the General Teaching Council for Scotland*. Available at: www.gtcs.org.uk/nmsruntime/saveasdialog.aspx?lID=3029&sID=7408.

Goldacre, B. (2013) *Building Evidence into Education*. Department for Education. Available at: media.education.gov.uk/assets/files/pdf/b/ben%20goldacre%20paper.pdf.

Leung, S. (2011) A Comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research*, **37**(4), 412–421.

Mincu, M. (2013) *Teacher Quality and School Improvement: What Is the Role of Research?* Research and teacher education: The BERA-RSA Inquiry. British Educational Research Association (BERA).

Norman, G. (2010) Likert scales, levels of measurement and the 'laws' of statistics. *Advances in Health Science Education*, **15**(5), 625–632.

Shulman, L. S. (1986) Those who understand: knowledge growth in teaching. *Educational Researcher*, **15**(2), 4–14.

Simonneaux, L. (2001) Role-play or debate to promote students' argumentation and justification on an issue in animal transgenesis. *International Journal of Science Education*, **23**(9), 903–927.

Turning Technologies. Available at: www.turningtechnologies.com/polling-solutions/turningpoint.

Welsh Government (2011) *Revised Professional Standards for Education Practitioners in Wales*. Available at: learning.gov.wales/docs/learningwales/publications/140630-revised-professional-standards-en.pdf.

Rick Tynan is a Senior Lecturer in Science Education (Biology) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: R.J.Tynan@ljmu.ac.uk

Robert Bryn Jones is a Senior Lecturer in Science Education (Chemistry) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: R.B.Jones@ljmu.ac.uk

Andrea Mallaburn is a Senior Lecturer in Science Education (Chemistry) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: A.Mallaburn@ljmu.ac.uk

Ken Clays is a Senior Lecturer in Science Education (Physics) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: K.Clays@ljmu.ac.uk

Subject knowledge enhancement courses for creating new chemistry and physics teachers: the students' perceptions

Richard Tynan, Robert Bryn Jones, Andrea Mallaburn and Ken Clays

ABSTRACT Subject knowledge enhancement (SKE) courses are one option open in England to graduates with a science background whose first degree content is judged to be insufficient to train to become chemistry or physics teachers. Previous articles in *School Science Review* have discussed the structure of one type of extended SKE course offered at Liverpool John Moores University, and its impact on outcomes for students. This article focuses on the qualitative responses collected from the anonymous exit questionnaire surveys returned by chemistry and physics SKE students about to continue on to their science Postgraduate Certificate in Education course. The main positive course characteristic cited was student support. The main positive outcomes perceived were improved subject knowledge and skills and increased confidence.

Aspiring teachers in England can currently apply to a number of different types of programmes that lead to Qualified Teacher Status accredited by a variety of higher education and school providers. One response to severe teacher shortages in key subjects such as the physical sciences has been to use a bursary system to encourage graduates in related science subjects to enhance their subject knowledge to enable them to teach chemistry or physics. Such training has taken a variety of forms, from online distance learning to traditional face-to-face offerings, over varying time frames from a few weeks to whole-year courses (Gibson *et al.,* 2013a, 2013b).

Funding changes led to the suspension of subject knowledge enhancement (SKE) courses during the academic year 2013–14. However, the difficulty in recruiting to shortage subjects remained and funding for SKE courses was later reinstated for candidates who held a conditional offer of a place on a teacher-training course. Employers of science teachers will continue to receive applications from candidates who may not have a first degree in that science subject but in a related science subject with an additional relevant SKE course accreditation.

However, recent experience of the application process at Liverpool John Moores University (LJMU) has indicated that the need to attract applicants to teach science shortage subjects and also accelerate science teacher recruitment to schools has affected the balance of SKE course provision. These constraints have encouraged the use of shorter SKE courses, which range from those requiring full attendance to distance-learning packages that take place fully online and whose effectiveness has yet to be evaluated.

This is the final article of three published in *School Science Review* investigating the impact of one-year, full-time, higher education certificate chemistry and physics SKE courses offered at LJMU between 2008 and 2012. In the first article, the course structure and aims were discussed (Inglis *et al.*, 2013). The second (Tynan *et al.*, 2014) examined quantitative data from course documentation in order to evaluate the outcomes for SKE students. These outcomes were similar to those of the Postgraduate Certificate in Education (PGCE) trainees who joined the course with honours degrees in chemistry and physics. In this article, the emphasis is on the qualitative data found within open-response boxes linked to eight Likert-type items (Clason and Dormody, 1994) and two free-response questions from the anonymous exit survey for the 2012–2013 SKE chemistry and physics cohort. Taken together, the three articles offer an insight into one type of extended science SKE course with a significant 'face-to-face' component that will be of interest to those who seek to recruit teachers in science shortage subjects.

Administering the survey

At the conclusion of the 2012–13 chemistry and physics SKE courses, all students were given the opportunity to complete an anonymous exit feedback questionnaire within a timetabled session. At this time, the SKE courses were expected to recruit and run during the next academic year and the emphasis was on seeking student feedback for the purpose of course improvement. Students were informed of the intention to analyse the anonymous data for publication and that ethical approval had been granted by the appropriate university committee. As the students were about to start their PGCE science course at LJMU after the summer break, it was necessary to take care to ensure that all students knew that completion of the questionnaire was completely voluntary, with no possible repercussions for those not wishing to participate. The return of a completed or partially completed questionnaire implied informed consent.

Findings

The total possible number of respondents for each question was 26, with similar numbers of chemistry and physics SKE course participants (Table 1). There were more female chemists and more male physicists but the total numbers of male and female participants across both courses were almost equal (Table 1).

All students answered all the Likert-type items except for one who did not respond to Question 6. Numerically, the feedback was very positive. Most respondents chose the most positive category for the majority of questions (Table 2). For three questions there was a more even split between the two positive categories and one respondent used a negative category for one of the questions (Table 2). This indicated a less positive endorsement of the feedback received and enjoyment of the course content.

Additional comments for each question

1 **The SKE course has developed my subject knowledge and understanding throughout the year.** Most free responses reinforced the respondents' perception that their subject knowledge and understanding had improved through the course and that self-confidence in their subject knowledge had increased. More chemistry SKE than physics SKE students commented on their improved confidence levels. Other strong themes were that the

Table 1 SKE course cohort characteristics 2012–13

Subject	Male	Female	Totals
Chemistry	5	10	15
Physics	7	4	11
Totals	12	14	26

Table 2 Numerical responses to questionnaire items

Question	Strongly agree	Agree	Disagree	Strongly disagree	Total
1. The SKE course has developed my subject knowledge and understanding throughout the year.	18	8	0	0	26
2. The teaching on the SKE course has supported my learning well throughout the year.	20	6	0	0	26
4. The feedback in the sessions has supported my development and understanding.	15	11	0	0	26
5. The feedback I have received regarding my assignments has supported my development and understanding.	15	11	0	0	26
6. I have enjoyed the content in the course this year.	14	10	1	0	25
7. I feel highly motivated to participate in my PGCE next year due to my involvement in my SKE course.	21	5	0	0	26
8. I feel well prepared to embark on my PGCE next year due to my involvement in my SKE course this year.	20	6	0	0	26
	Strong impact	Good impact	Slight impact	No impact	Total
3. The tutor(s) has/have given me sufficient support throughout the year.	21	5	0	0	26

course had been good revision of previously known material and that it had been valued as good preparation for teaching. A group of respondents clearly saw their course as a framework for their own personal programme of study. Several others cited the course content and delivery as supporting their learning. However, there were four individuals who had personal preferences for content or modes of delivery other than those on offer during the course.

2 **The teaching on the SKE course has supported my learning well throughout the year.** The open responses cited a variety of reasons for a positive perception. The most common themes were the ease of access to tutors and the high quality of support. This was followed by the high quality of the teaching. A variety of other aspects were also mentioned by one or two respondents. For example, the course format was considered to be effective in giving opportunities for clarifying subject content and misconceptions. Also, the inclusion of skills needed in preparation for PGCE had increased confidence.

3 **The tutor(s) has/have given me sufficient support throughout the year.** The participants clearly perceived this as an area of strength for this type of SKE course format. There were no negative or qualified open response comments for this survey item.

4 **The feedback in the sessions has supported my development and understanding**. No strong common themes emerged from the open responses. Respondents made a variety of comments on the quality and nature of the feedback received in sessions and the ways it had been personally useful to them.

5 **The feedback I have received regarding my assignments has supported my development and understanding.** The respondents wrote about several themes. They commented positively on the nature and sources of the feedback received and gave examples of where they had individually found it useful. Comments mentioned attributes of feedback received on assignments that respondents valued. In descending order of frequency they were: constructive, clear, focused, necessary, informative and prompt. Comments demonstrated a keen awareness of the utility of feedback in relation to identifying strengths and areas for development, and improving general performance or achieving grade criteria.

6 **I have enjoyed the content in the course this year**. The descriptors linked to enjoyment of the content of the course cited by one or more respondents, in descending order of frequency, were: fun, interesting, the subject content, engagement, the tutors, personal development, pedagogy and links between separate sciences. Some sessions that challenged individual respondents resulted in a range of negative comments: some sessions were not enjoyable, reflective log was a 'necessary evil', uneven distribution of workload, negative references to a specific module, coverage in sessions, lack of challenge, more depth wanted and specified content that had been omitted and should be included in future.

7 **I feel highly motivated to participate in my PGCE next year due to my involvement in my SKE course.** Common themes among respondents were based upon confidence or increased confidence to start their PGCE course. Some expressed this as feeling well prepared and looking forward to starting PGCE after SKE. Although not a primary function of the course, pedagogical issues were a key and integral part and respondents welcomed opportunities to microteach and present in order to develop teaching skills in preparation for the PGCE course.

8 **I feel well prepared to embark on my PGCE next year due to my involvement in my SKE course this year.** Again, increased confidence and feeling well or better prepared for the PGCE year were the explanations most often linked to the positive response to this item. Most respondents saw this in terms of having acquired new subject knowledge and understanding, being more up to date or recently revising the topics needed for teaching. A few cited modelling of pedagogy, microteaching, gaining experience or developing presentation skills as positive experiences. The feeling that support would continue into the next course was valued. However, several respondents were more guarded and demonstrated awareness that even more subject knowledge development would be needed during the PGCE year.

Free response questions

- **Write a few sentences to summarise what the SKE course/tutors do well**. A wide variety of comments were made that fell into two main categories: comments based upon individual preferences related to aspects of the course and those related to attributes of tutors; there were no strong common themes.
- **Write a few sentences to summarise what we could do better.** Comments about a particular module and its mode of delivery demonstrated some respondents' preference for the adoption of a didactic approach throughout the module and for all tasks within it to contribute to assessment outcomes.

Summary of main findings

Positive outcomes

- Respondents ended the course with high levels of motivation towards their PGCE course and feeling prepared for its challenges.
- Respondents felt their subject knowledge and understanding in chemistry or physics had improved over the course.
- Respondents felt more confident to teach chemistry or physics.
- Most participants enjoyed the course.
- Tutor support was highly valued and seen as a strong positive aspect of the course.

Areas for development

- Session and assignment feedback was found to be useful for a variety of reasons, but some qualified statements sought formal feedback more often, more quickly after assignments and in more detail.
- Structural course improvements were suggested by some respondents. The most common was to increase the amount of contact time. On occasion this suggestion was linked to a desire to reduce the overall course length.

Discussion

A wide-ranging and detailed evaluation of SKE courses was undertaken by CooperGibson Research for the UK government Department for Education (Gibson *et al*., 2013a, 2013b). This entailed a large-scale national survey of students at the start and end of SKE courses, on PGCE courses and during the newly qualified teacher (NQT) year. All SKE subjects and courses were included across the range of course lengths and models of delivery on offer. This research indicated that SKE courses were very well perceived by students and seen as excellent preparation for PGCE courses (Gibson *et al*., 2013a). Course satisfaction ratings were mostly high and students expressed perceptions of improved subject knowledge and greater confidence in their SKE subject and their ability to teach it (Gibson *et al*., 2013b). Although the national survey respondents constituted a self-selecting sample across all SKE subjects, the survey data from the start and end of SKE courses originated mostly from respondents preparing to teach chemistry, mathematics and physics (Gibson *et al*., 2013b). This was useful for comparison with the local data collected for this article. The data from LJMU students is indicative of the findings of the national survey but more focused in referring only to chemistry and physics SKE students and one type of extended SKE course.

Respondents to the LJMU chemistry and physics SKE student feedback survey gave very positive feedback about its content and delivery. Philpott (2014) considered a wide range of models for professional learning that are used by teacher trainers and educators. Initially, it might appear that models that emphasise the individual's cognitive and psychological aspects of learning, such as Kolb's (1983) experiential learning cycle, would be most relevant to the design of SKE courses. Models that focus on the social aspects of learning, such as Wenger's (1998) communities of practice, might appear to have more relevance to the process of initial teacher training/education (ITT/E) in school. When making positive comments, the LJMU SKE students appear to have considered their own progress from a social learning perspective. When thinking of areas for course improvement, a more personal and psychological learning model was adopted.

The students agreed on the value of regular contact sessions that enabled them to receive support from tutors and each other over an extended period of time. However, the reservations expressed often reflected the differences in their personal attitudes to individual learning and in their expectations of what a SKE course should cover. Some favoured a shorter, more intense course with a didactic approach throughout in order to include more subject knowledge covered efficiently in more depth. While most welcomed the opportunities

provided to practise skills and develop chemistry and physics pedagogy, this was not to be at the expense of the subject knowledge content covered.

Both the national data (Gibson *et al*., 2013b) and the local findings reported in this article suggest that SKE chemistry and physics students are likely to feel confident and in possession of the necessary subject knowledge to teach. However, student self-evaluation means little unless tested and confirmed by third-party evaluations or assessments. For LJMU SKE chemists and physicists completing their PGCE courses in 2012, Tynan *et al*. (2014) found no difference in the grade distributions on final review forms for subject knowledge and overall teaching grades compared with those who had commenced the PGCE with a chemistry or physics first degree. Early employment data for both groups were also similar (Tynan *et al*., 2014). Furthermore, to pass an accredited higher education certificate course, participants must meet all its assessment requirements. This suggests that the LJMU SKE student survey expressions of confidence and perceptions of improved subject knowledge in chemistry and physics are linked to measurable outcomes. The national survey did not look at outcomes for SKE students (Gibson *et al*., 2013b).

A model for development that helps make sense of the differing approaches to assessment encountered while learning to do a job or profession was proposed by Hager and Butler (1996). This was applied to the assessment of teacher competencies by Martin and Cloke (2000), who described a model for professional

development congruent with the pattern of initial teacher training at LJMU at the time of this study (Figure 1). SKE courses offered prior to the commencement of teacher training can be seen as an extension of the initial phase of professional development involving the acquisition of fundamental subject-specific knowledge, attitudes and skills (Figure 1). However, this model of professional development suggests that the methods of assessing new subject knowledge will not only vary depending upon the model of delivery and length of the SKE course but also with the phase of professional development.

Once participants embark upon their teacher training in school, the assessment emphasis shifts away from testing and a scientific measurement model towards a qualitative judgemental one based upon observations of learning and teaching (Martin and Cloke, 2000). This reflects different purposes of subject knowledge for teachers during the practice and demonstrating personal competency phases of development (Hager and Butler, 1996; Martin and Cloke, 2000). Here, the emphasis is to develop and practise teaching and assessment skills to enhance another's learning of subject knowledge and skills. Competence in the teachers' standards (Department for Education, 2011) is judged by the impact on pupils' learning over time so there must be a change in emphasis from assessing a teacher's personal knowledge and understanding of a subject discipline to judging the impact that their subject knowledge has upon their learners' progress.

This client-led approach has resulted in the current resurgent popularity of clinical practice ITT/E models among education policy makers

Developmental Stage	Acquisition of knowledge, attitudes and skills		Simulated performance in practice domains		Demonstration of personal competence in real situations		
Training Phase	Subject qualification		Phases 1 and 2a		Phase 2b	Phase 3	NQT Year
Assessment Model	Summative scientific measurement (eg examination)		Formative judgemental approach (eg observation)		Formative judgemental approach (eg observation)	Summative judgemental approach (eg paired observation, portfolio, triangulation)	

Figure 1 A model for ITT/E professional development; after Hager and Butler (1996) and Martin and Cloke (2000)

(Philpott, 2014). The perceptions of improved chemistry and physics subject knowledge expressed by the LJMU SKE participants in their feedback and their increased levels of confidence would be important in helping them perform as credible teachers of physical sciences in the eyes of learners and teaching colleagues. This could also help explain the parity of assessment and then early employment outcomes of LJMU PGCE students with SKE qualifications compared with those with first degrees in physical sciences reported by Tynan *et al.* (2014).

SKE courses in chemistry or physics do not claim to provide the same depth and range of subject knowledge and understanding as a first degree in those subjects. Furthermore, the way subject knowledge is acquired and assessed can differ greatly between undergraduate physical science, SKE and PGCE courses. Together, these considerations should lead potential employers of chemistry and physics teachers to take care when considering the subject knowledge of an applicant with a SKE qualification. On the other hand, there is a pressing need to accelerate the process of training new teachers of chemistry and physics. The evidence gathered so far supports the idea that SKE courses in chemistry and physics can help recruitment to these shortage science subjects and provide confident and capable physical science teachers.

References

Clason, D. L. and Dormody, T. J. (1994) Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, **35**(4), 31–35.

Department for Education (2011) *Teachers' Standards: Guidance for School Leaders, School Staff and Governing BNodies*. London: Department for Education.

Gibson, S., O'Toole, G., Dennison, M. and Oliver, L. (2013a) *Evaluation of Subject Knowledge Enhancement Courses: Annual Report – 2011–12. Research Report*. Department for Education/CooperGibson Research.

Gibson, S., O'Toole, G., Dennison, M. and Oliver, L. (2013b) *Evaluation of Subject Knowledge Enhancement Courses: Technical Report, Analysis of Survey Data 2011–12. Research Report*. Department for Education/ CooperGibson Research.

Hager, P. and Butler, J. (1996) Two models of educational assessment. *Assessment and Evaluation in Higher Education*, **21**(4), 367–378.

Inglis, M., Mallaburn, A., Tynan, R., Clays, K. and Jones, R. B. (2013) Insights from a subject knowledge enhancement course for preparing new chemistry and physics teachers. *School Science Review*, **94**(349), 101–107.

Kolb, D. A. (1983) *Experiential Learning: Experience as the Source of Learning and Development*. London: Prentice Hall.

Martin, S. and Cloke, C. (2000) Standards for the award of qualified teacher status: reflections on assessment implications. *Assessment and Evaluation in Higher Education*, **25**(2), 183–190.

Philpott, C. (2014) *Theories of Professional Learning: A Critical Guide for Teacher Educators*. Northwich: Critical Publishing.

Tynan, R., Mallaburn, A., Jones, R. B. and Clays, K. (2014) Subject knowledge enhancement (SKE) courses for creating new chemistry and physics teachers: do they work? *School Science Review*, **95**(353), 85–94.

Wenger, E. (1998), *Communities of Practice: Learning, Meaning and Identity*. Cambridge University Press.

Rick Tynan is a Senior Lecturer in Science Education (Biology) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: R.J.Tynan@ljmu.ac.uk

Robert Bryn Jones is a Senior Lecturer in Science Education (Chemistry) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: R.B.Jones@ljmu.ac.uk

Andrea Mallaburn is a Senior Lecturer in Science Education (Chemistry) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: A.Mallaburn@ljmu.ac.uk

Ken Clays is a Senior Lecturer in Science Education (Physics) in the Faculty of Education, Health and Community at Liverpool John Moores University. Email: K.Clays@ljmu.ac.uk

Consistency counts – or does it?

Rick Tynan, Andrea Mallaburn
Liverpool John Moores University

Abstract
All stake holders in competency based teacher training systems share an interest in the consistency of assessment outcomes and practice. Assessment data from more than 200 trainees participating in Initial Teacher Training/Education (ITT/E) programmes and partnerships at a Higher Education (HE) provider in the Northwest of England were analysed during the academic year 2014-15.

At four formal review points the overall teaching grades received by trainees were compared across five ITT/E programmes leading to Qualified Teacher Status (QTS). Several statistical approaches were employed and compared. All the methods indicated consistency of outcomes across the programmes for the final summative assessment.

Two statistical methods were used to investigate the strength of correlations between grades awarded for individual teaching standards and the trainees' overall teaching grades. Both demonstrated that all individual standards were positively correlated with overall teaching grades.

The second and qualitative phase of the study is ongoing and uses Q-Analysis to illuminate these initial findings by seeking to identify clusters of subjectivity amongst mentors and tutors when prioritising statements about assessment. It is too early to report any results from this phase.

Key Words
assessment; placements; mentors; tutors; standards; competencies; ITT/E; consistency; primary; secondary.

Context
Achieving and demonstrating consistency in both assessment outcomes and practice is of interest to both HE (Higher Education) and School based accreditors of QTS (Qualified Teacher Status). Ofsted (2015) (Office for Standards in Education) uses consistency across partnerships as a performance indicator in the inspection of ITT/E (Initial Teacher Training/Education) provision. We set up this study to test the assumption that statistical analysis of assessment outcomes supported by qualitative evidence of assessment procedures can be used to demonstrate consistency in these areas. The aim was to use quick and reliable analysis tools and apply them diagnostically throughout the year in order to redress any inconsistencies detected between programmes or assessment points.

There are reasons other than inspection for seeking to improve consistency. Our experiences across school/university partnerships indicate that the quality of mentoring and coaching relationships can be adversely affected when trainees perceive their assessments to be inaccurate or unfair. On the other hand, there is relatively little central guidance to help

Citation
Tynan, R., Mallaburn, A. (2017) 'Consistency counts – or does it?', *TEAN Journal*, 9(1), pp 90-99.

assessors make objective and reliable judgements about their trainees' competencies with respect to the Teachers' Standards (Department for Education, 2011) in England.

Recommendation for QTS is currently based upon the assessment of teacher competencies described by eight teaching standards split into a number of sub-divisions together with a set of professional expectations (Department for Education, 2011). Individual standards and overall teaching are graded using a four-point scale: 1 (Outstanding), 2 (Good), 3 (Requires improvement) and 4 (Inadequate). Assessors are instructed to take into account trainee experience and stage of training and to adopt a holistic approach to sub-section criteria when reaching a judgement about the grade for an individual standard (Department for Education, 2011). The standard descriptors set out minimum expectations for performance but provide no indication of what is required for the award of grades 2 and 1.

A wide range of schools and HE providers involved in ITT/E in the Northwest of England have collaborated over time to apply general Ofsted descriptors for the assessment of final year trainees to the Teachers' Standards criteria for minimum performance (Department for Education, 2011). The result has been an individual trainee standards tracking document containing a set of performance descriptors for all sub-sections of the standards at every point on the four-point grading scale. To facilitate consistency across the HE provider's partnerships all assessors were expected to use the tracking document to reference their grading decisions.

Across the programmes involved in the study the subject mentors supervising the trainee teachers in school assessed them at three formative and one final summative assessment point. These corresponded to the completion of each phase of training (Figures 2 and 3). They awarded grades for each of the Teachers' Standards and collated these judgements to arrive at an overall teaching grade. Professional mentors moderated assessments made by different teachers within their school and school liaison tutors from the HE provider visited schools to conduct training and quality assure the mentoring and assessment processes. However, from our experience across the partnerships, despite this high level of professional, organisational and individual effort, assessment and grading continues to challenge new and experienced mentors and tutors. In turn, achieving and gathering evidence of consistency in assessment within and across multiple partnerships and programmes is a challenge for those with quality assurance roles. An obvious place to look for evidence was the assessments data and partnership documentation generated by trainees.

This paper reports the quantitative results from the first year of a mixed method, practitioner research investigation into the consistency of assessment outcomes across one North West of England HE provider's ITT/E programmes and partnerships. We also report on progress with data gathering for the second qualitative phase of the project. The project is on-going but early indications are that these approaches are worth pursuing.

Methodology and methods
This investigation is a practitioner led staff project. It links to local perceptions of issues and opportunities around consistency of assessment outcomes and practice for schools working in ITT/E partnership with a Northwest of England HE provider. The study evaluates the impact of interventions intended to improve consistency within and across partnerships and has the potential to become cyclic. As such it fits well with an action research model of investigation (Burton & Bartlett, 2009: 9).

The planned interventions were to:

- increase attendance and participation in Mentor Training by including a training element in all liaison visits and supplementing the HE provider programme of training meetings by on-site training in partner schools
- maximise consistency of assessment outcomes and practice by referencing all assessments to the minimum performance descriptors set down in the Teachers' Standards (Department for Education, 2011) and the criteria set down in the individual trainee standards tracking document
- increase the rigour of the final assessment process through longer, more structured triangulation meetings chaired by HE Tutors

The assessment data analysed statistically were drawn from five programmes at four formal review points in the academic year 2014-15. The programmes involved were Primary PGCE, Secondary PGCE (Postgraduate Certificate in Education), Secondary Salaried School Direct, Primary Education Honours degree with QTS (3 Year) and Primary Education Honours degree with QTS (4 Year). Non-Salaried School Direct trainees were grouped with core PGCE trainees. The number of trainees following each programme varied according to the quotas allowed and final uptake by applicants. The trial included three ways of analysing overall teaching grades at different review points across programmes and two ways of comparing overall grades to the grades for individual standards.

The statistical analyses used to compare overall teaching grades across programmes and assessment points were:

- The visual presentation of mean grades and their 5% confidence limits
- Single factor analysis of variance (ANOVA) on counts for grades across programmes
- Chi squared analysis on counts for grades across programmes

- The statistical analyses used to compare individual standard grades with overall teaching grades were:
- Pearson's correlation coefficient
- Spearman's Rank correlation coefficient

For the qualitative phase of the investigation, a concourse of around 40 statements concerning the assessment of trainees on school placement has been constructed from policy and course documentation. Mentors and tutors from across programmes and partnerships have been invited to participate in an anonymous on-line activity to place these in personal priority order for Q analysis (Brown, 1980, van Exel et al., 2005). This will identify clusters of subjectivity with respect to the concourse of statements amongst respondents. Participation will be anonymous and voluntary. Informed consent will be implied by completing and submitting the on-line activity.

The full project will collate the findings from both the quantitative and qualitative data analysis and, as such, constitutes a mixed methods study.

Interpreting the statistical conclusions
The methods trialled were a mixture of parametric and non-parametric statistical analysis. Parametric tests assume that data are distributed in a particular way whereas non-parametric

tests do not. If you know the distribution of your data, you can usually place more confidence in a conclusion reached using an appropriate parametric method of analysis. The parametric methods trialled were all designed for use with normally distributed data. This distribution is often seen when observations from large populations are presented graphically. Plotting the frequencies with which values occur results in a characteristic bell shaped curve that is symmetrical either side of the average value. Norman (2010) argued strongly that there is evidence that conclusions reached using parametric methods can be robust and valid even when the assumptions underlying their use are not in place. However, when data do not (or are not known to) conform to any distribution then a non-parametric approach may give more reliable and accurate conclusions.

Statistical analysis starts with the Null Hypothesis (H^0). This is always a neutral or cautious hypothesis e.g. there is no difference between the mean grades or numbers of grades awarded by the five programmes at a particular assessment point. This is accepted or rejected and the four possible results from a statistical test are shown in Figure 1. Statisticians prefer to reject H^0 but only because the probability of a wrong conclusion is quantified when this happens. The maximum risk of error allowed in statistical investigations is typically a probability of 0.05 or 5%.

Null Hypothesis (H^0)

		H^0 Valid	H^0 Invalid
Conclusion	Reject H^0	**Type 1 Error (5% chance of error)**	**Correct**
	Accept H^0	**Correct**	**Type 2 Error (error unknown)**

Figure 1. Statistical errors.

Using statistical tests ensured that objective conclusions were reached about consistency or inconsistency in the assessment data. All the statistical methods used to compare overall teaching grades awarded across the five ITT/E programmes were interpreted in the same way. Accepting H^0 indicated consistency in assessment outcomes and rejecting it demonstrated inconsistency. For the investigation of correlations H^0 was that there was no correlation between grades for individual teachers' standards and the overall teaching grades awarded. Rejecting H^0 with a positive correlation suggested an association between the standard and overall teaching performance in the minds of the assessor. However, a negative correlation or no correlation indicated grading decisions about individual standards and overall teaching that were inconsistent with the guidance and training the assessors had received.

Results

The use of means and 5% confidence limits to demonstrate assessment data pictorially (Figure 2.) was visual and easily understood. It demonstrated the progression in overall teaching grades throughout the year awarded across all the programmes. The number of assessments made within programmes varied across the assessment points as some students deferred their studies, returned to study or permanently left their course. The 5% confidence limits of the mean overall grades awarded by programmes at the end of each phase of training overlapped

except for one pair of programmes at the third formative assessment point (Figure 2.). This suggested that there was a high degree of consistency between programmes.

Single factor ANOVA indicated no differences in the distribution of grades across the five ITT/E programmes at any assessment point (Table 1) suggesting consistency between programmes.

Chi squared analysis demonstrated specific differences between the observed and calculated expected frequencies of overall teaching grades between programmes for the first three assessment points but not the last (Table 2).

Norman (2010) discussed and defended the parametric analysis of data derived from number scales similar to those used for our trainee teacher assessments. However, for our data there were differences in the conclusions reached by different methods and, without going into detailed mathematical and statistical arguments, the method that we had most confidence in was the Chi Squared analysis. This is a non-parametric method comparing observed numbers of grades awarded with expected numbers calculated using a contingency table. It indicated differences between individual programmes and the rest of the partnerships at the formative assessment points but consistency between all of them at the final summative assessment.

The correlation coefficient analysis compared the grades awarded for each individual teaching standard with the overall teaching grade at each assessment point for each programme. The correlation study did not indicate any 'rogue' standards (not positively associated with overall teaching grade) in any programme at any assessment point. All correlations were positive with a 5% or less chance of this conclusion being in error. This indicated consistency of outcomes across programmes at all assessment points. The results were statistically interesting because both the parametric and non-parametric methods gave similar results with identical conclusions in all cases.

Figure 2. A pictorial representation of mean overall teaching grades and their 5% confidence limits across five programmes at each review point.

Table 1. ANOVA summary table.

Assessment	F-value (the result of the ANOVA test)	Probability (P) of error if H^0 is rejected	Conclusion (reject H^0 if P is 0.05 or less)
Summative	1.02	0.43	H^0 accepted
3rd Formative	0.64	0.64	H^0 accepted
2nd Formative	0.94	0.47	H^0 accepted
1st Formative	1.02	0.43	H^0 accepted

Table 2. Chi-squared summary table.

Assessment	Degrees of freedom	Chi^2 value	Critical values Probability (P) of error if H^0 is rejected: 0.05 0.01	Conclusion (reject H^0 if the Chi^2 value is larger than either critical value)
Summative	8	9.51	15.51 20.09	H^0 accepted
3rd Formative	4	15.55	9.49 13.28	H^0 rejected (P= 0.01)
2nd Formative	4	14.73	9.49 13.28	H^0 rejected (P= 0.01)
1st Formative	4	14.24	9.49 13.28	H^0 rejected (P= 0.01)

Discussion

Our study arose from our perception that the nature of professional learning and of criteria referenced assessment of competencies contribute strongly to the challenges associated with achieving and maintaining consistency in assessment across ITT/E programmes and partnerships. Philpott (2014) provided a critical summary of professional learning models and their relationship to the current school led model of ITT/E provision. He divided these broadly into those that focus on the psychology of individual learning and those that emphasise group or social aspects of learning. Kolb's experiential learning model and the clinical practice approach are examples of models that start with the individual's cognitive development and the increase in knowledge and skills based upon the evidence of practical experience (Philpott, 2014). Communities of practice and apprenticeship models emphasise social aspects of learning and the need for trainees to demonstrate independent competence to gain acceptance as a practitioner (Philpott, 2014).

Models emphasising the trainee's individual development of professional knowledge and skills appear to give more opportunity for assessors to be objective in their judgements. However, criterion referenced assessment of students in HE is liable to a variety of subjective influences even when assessments have been designed to reduce this (Donovan, Price and Rust, 2001). Further, the Teachers' Standards (Department for Education, 2011) describe the minimum criteria for competence but give no guidance on acceptable evidence to use in judging when a standard has been achieved or at which grade. Ofsted (2015) perceives successful teaching in terms of the pupils' learning outcomes. The argument is that a standard is met when its impact on learning is at least satisfactory over time. Learner rather than teacher performance then

becomes the evidence for standards and the bigger and more consistent over time the impact on learning the higher the grade awarded. However, our experience indicates that the problem of establishing how much a trainee's competency in a particular standard contributes directly to the overall impact on pupil learning appears to remain essentially subjective.

Learning to become a teacher is not a straightforward process that can be tracked using simple assessment tools. Hager and Butler (1996) proposed a model for professional development that also considered assessment. Martin & Cloke (2000) applied this model to teaching and the assessment of teacher competencies. Without troubling too much about the individual or social processes involved in professional learning their model highlights differences in activity and expectations as professional development progresses, and the assessment models associated with each stage. Figure 3 maps their model to the phases of teacher training currently in common use in ITT/E partnerships in the Northwest of England. We have found the judgemental assessment model described by Martin and Cloke, (2000) useful when considering factors that may affect consistency when assessing trainee teachers on school placement.

Developmental Stage	Acquisition of knowledge, attitudes and skills		Simulated performance in practice domains		Demonstration of personal competence in real situations	
Assessment Model	Summative scientific measurement (eg examination)	Formative judgemental approach (eg observation)	Formative judgemental approach (eg observation)		Summative judgemental approach (eg paired observation, portfolio, triangulation)	
Training Phase	Subject qualification	Induction, Orientation & Beginning Learning, Teaching & Assessment (LTA)	Consolidating LTA	Extending LTA & Qualifying to Teach	NQT Year	

Figure 3. A model for the professional development of trainee teachers (developed from Hager & Butler (1996) and Martin & Cloke (2000).

The basic pre-requisite for aspiring teachers is adequate subject knowledge and trainee teachers in England must possess an honours degree in an appropriate subject or an equivalent qualification. Arriving at the start of teacher training, success in a subject at undergraduate level will have been judged using mainly scientific measurement assessment tools such as written assignments, portfolios and examinations (Hager and Butler, 1996). The current compulsory skill tests in English and Mathematics are further examples of filters applied to applicants for ITT/E programmes using a scientific measurement assessment tool.

Trainees extend their knowledge and skills beyond their own subject discipline as they progress through the various stages of their professional development (Figure 3). Shulman (1986) referred to subject knowledge for teachers in terms of three areas: subject matter content knowledge, pedagogical content knowledge and curriculum knowledge. Banks, Moon and Leach (2005) considered professional teacher knowledge to be a personal construct of subject and pedagogical knowledge together with school knowledge. The Teachers' Standards (Department for Education, 2011) include all these elements within the various standard descriptors that are used to assess trainee competencies. However, the assessment of trainees' knowledge on school placement is likely to involve qualitative judgements rather than formal testing (Martin & Cloke, 2000). As trainees practice and develop their teaching skills and then demonstrate

competency (Figure 3) they take on increasingly independent responsibility for their classes. Successful application of professional knowledge in the classroom becomes the assessment focus. Assessors then rely entirely upon a judgemental approach based upon qualitative evidence that is often considered less reliable and more subjective than scientific measurement (Martin & Cloke, 2000).

Leshem & Bar-hama (2008) investigated and discussed the issues that arose when criteria based assessment of teaching competencies was adopted by their ITT/E programme. Tutors used the criteria analytically or to confirm their overall holistic assessment decisions about teaching and learning. Their students perceived a role for clear assessment criteria and criterion based assessments during feedback but preferred holistic approaches to summative assessment.

The mentors and tutors in our study were expected to make evidence based judgements based upon qualitative evidence. However, holistic, analytical or combined approaches to assessment were all consistent with the framework and guidelines agreed with partner schools. Our interventions constituted new or amended organisational steps intended to reduce the potential for inconsistency between assessors due to subjective interpretation of assessment criteria and personal differences when applying the guidelines.

The assessment data for 2014-15 yielded evidence of a high degree of consistency across the five ITT/E programmes and their partnerships especially for the final summative assessment of the overall teaching grade. No firm conclusions can be reached at this stage about the reasons for this. It is, however, reasonable to speculate on the list of interventions and identify those which are associated solely with the final summative assessment. The interventions designed to counter inconsistency through assessor subjectivity can be summarised as:

- increased emphasis on mentor training,
- the application of rigorous, common assessment procedures based upon the Teachers' Standards (Department for Education, 2011) and grade descriptors developed and set down in the individual trainee standards tracking document, and
- the formalisation of final triangulation meeting procedures for quality assuring the summative assessment of trainees.

The evidence is circumstantial and causal links have yet to be established but the nature of the revised final assessment triangulation meeting with the presence of an external quality assurer are possible influences on the high degree of consistency of final assessment outcomes.

With this in mind the concourse of statements about the assessment of trainees on placement required for Q analysis (Brown, 1980, van Exel & de Graaf, 2005) was constructed. Just over forty statements based upon the Teachers' Standards (Department for Education, 2011), partnership documentation and training materials were selected with reference to the results of the quantitative phase. Mentors and tutors have been invited to carry out an on-line exercise to place the statements into a personal priority order. Q analysis of the results will identify clusters of subjectivity due to groups of respondents with differing assessment priorities. It is possible that this may provide a link to one of the interventions put in place to encourage consistency in outcomes and practice or may identify a different influence.

Conclusions and recommendations:

All the statistical methods trialled indicated consistency of overall teaching grade assessment outcomes across all programmes for the final summative assessment just prior to the recommendation of QTS. This constitutes strong evidence of consistency in assessment outcomes across all programmes. Future comparisons should be routinely performed at each formal review point as part of quality assurance procedures. Chi squared analysis based upon assessment data and contingency table calculations is the method recommended from those trialled in this study.

The comparison of the grades awarded for individual teachers' standards and the overall teaching grade using correlation coefficients demonstrated only strong positive correlations. This gives some indication that assessors are keeping to the guidelines provided. As there were no differences in the conclusions reached using parametric and non-parametric methods, applying the quicker of the two methods, Pearson's Correlation Coefficient is recommended after each review point.

The rigorous and formal nature of the triangulation meeting that confirms final assessment judgements should be retained until there is evidence to the contrary that this has contributed to the consistency of assessment outcomes at this point.

References

Brown, S. R. (1980), *Political Subjectivity: Applications of Q Methodology in Political Science. Yale:* Yale University.

Banks, F., Leach, J. & Moon, B. (2005) 'Extract from new understandings of teachers' pedagogic knowledge', *The Curriculum Journal*, 16(3), pp.331-340.

Burton, D. & Bartlett, S. (2009) *Key Issues for Education Researchers*. London: Sage.

Department for Education (2011) *Teachers' Standards: Guidance for School Leaders, School Staff and Governing Bodies*, Crown copyright 2013 Available at: www.gov.uk/government/publications/teachers-standards (Accessed 23 November 2016).

Van Exel, N. Job A. & de Graaf, G. (2005). Q methodology: A sneak preview. Available at: www.jobvanexel.nl] or visit http://www.qmethodology.net (Accessed 23 November 2016).

Hager, P. and Butler, J. (1996) 'Two models of educational assessment', *Assessment and Evaluation in Higher Education*, 21(4), pp. 367–378.

Leshem, S. & Bar-Hama, R. (2008) 'Evaluating teaching practice', ELTJournal, 62(3), pp.257-265.

Martin, S. and Cloke, C. (2000)' Standards for the Award of Qualified Teacher Status: reflections on assessment implications', *Assessment and Evaluation in Higher Education*, 25(2), pp. 183–190.

Norman, G. (2010) 'Likert scales, levels of measurement and the ''laws''of statistics', *Advances in Health Science Education*, 15, pp.625–632.

O'Donovan, B., Price, M. and Rust, C. (2001) 'The Student Experience of Criterion-Referenced Assessment (Through the Introduction of a Common Criteria Assessment Grid)', *Innovations in Education and Teaching International*, 38(1), pp. 74-85.

Ofsted (2015) *Initial teacher education inspection handbook*. London: Crown.

Philpott, C. (2014), T*heories of Professional Learning: A critical guide for teacher educators.* Critical Publishing.

Shulman, L. S. (1986) 'Those who understand: knowledge growth in teaching', *Educational Researcher*, 15(2), pp.4-14.

Assessing trainee secondary teachers on
school placement: Subject knowledge and
overall teaching grades

Rick Tynan and Robert Bryn Jones
Liverpool John Moores University

Abstract
Schools and Initial Teacher Training/Education providers have joint responsibility for developing trainee subject knowledge. Due to the current curriculum and training emphasis placed upon the importance of subject knowledge, the relationship between it and overall teaching grades is of interest when monitoring trainee assessment data collated from school mentors in placement schools.

This paper reports a statistical analysis of numerical grades awarded on progress review forms completed by mentors using the teaching competencies described in Teachers' Standards in England. It includes the assessment data gathered from two consecutive cohorts of secondary Post Graduate Certificate/Diploma in Education trainees whilst on school placement experience. All the schools were in partnership with a single Higher Education provider in the North West of England. The focus for the analysis was the distribution of grades assigned to trainees in English National Curriculum core subjects for overall teaching and two standards with descriptors covering aspects of teacher subject knowledge.

Of twenty-four comparisons, only six indicated significant differences. In these instances, more high grades than expected were assigned for the standard describing teacher subject content and curriculum knowledge compared to the standard describing pedagogy and/or overall teaching.

Key Words
ITT/E; partnership; school placement; subject knowledge; assessment; teachers' standards; evidence; statistical analysis; consistency.

Context
Currently there are a variety of routes into teaching in England. School Centred Initial Teacher Training (SCITT) and School Consortia offer Qualified Teacher Status (QTS) either working alone or in conjunction with a Higher Education (HE) organisation that can accredit this qualification. Organisations that can accredit QTS are termed Initial Teacher Training (ITT) providers. HE providers working in partnership with schools can also offer ITT courses leading to QTS but usually offer alongside Masters level Initial Teacher Education (ITE) courses. The HE provider in this study offered mainly Post Graduate Diploma in Education (PGDE) courses leading to QTS with two thirds of the credits needed for a full Masters degree. It also worked with school consortia that required a Post Graduate Certificate in Education (PGCE) leading to QTS with one third of the credits required for a full Master's Degree.

In England, eight Teachers' Standards, each split into several descriptors, together with a set of professional expectations describe minimum levels of performance in competencies that trainees must demonstrate before they are recommended for QTS (Department for Education, 2011). The assessors in this study followed one Initial Teacher Training/Education (ITT/E) practice in England by grading individual standards and overall teaching at formal review points using a four-point scale: 1 (Outstanding), 2 (Good), 3 (Requires improvement) and 4 (Inadequate). This was in line with the number grading system in use at the time by the Office for Standards in Education (OFSTED) for assessing all teachers.[1] Assessors in this study used a locally produced expansion of the descriptors to guide judgements about trainees performing at a level above the minimum set down in the Teachers' Standards (Department for Education, 2011). The profile of grades for individual standards was then used to arrive at an overall grade for teaching. Guidance in the Teachers' Standards (Department for Education, 2011) states that in reaching judgements assessors should adopt a holistic approach to descriptors contributing to each standard and take into account level of experience or stage of training. ITT/E programmes in England take place largely or entirely in schools. Subject to moderation and quality assurance by school or HE ITT/E providers, school mentors have the first responsibility for both training and assessment.

After the education White Paper of 2010 (Department for Education, 2010) the government acted in England to re-establish the importance of subjects in school curricula. At the same time, it set in motion alterations to ITT/E that, amongst other things, ensured that recruiters would strongly associate teacher quality with subject discipline and degree classification. Namely, withdrawal of funding for applicants with less than a 2:2 degree classification, an expansion of its Teach First scheme for attracting top graduates to challenging schools and financial incentives for those with degrees in shortage subjects (Department for Education, 2010). Through these actions, policy makers demonstrated that they consider good subject knowledge a vital pre-requisite for successful teachers. The relationship between subject knowledge and overall teaching competency is, therefore, of interest when monitoring the trainee assessment data collected from school mentors in placement schools.

These government measures drew qualified approval from a range of sources with differing political perspectives who appeared, nonetheless, to be in broad agreement with the government's curriculum and assessment initiatives (Beck, 2012). Young (2011), for example, whilst seeing little benefit in a curriculum composed of fixed and unchanging traditional subjects, detailed the educational advantages of a curriculum organised by subject compared to one aimed at developing generic skills. Specifications in England (e.g. AQA, 2016) for first teaching in 2015 and 2016 list end-test only - GCSE and GCE AS/A Level subjects. Some subjects considered more peripheral or difficult to examine are no longer offered. This reflects the government's emphasis on traditional mainstream subjects and methods of assessment. For this reason, we were interested in investigating the assessment of ITT/E trainees in the traditional core subjects in addition to the pooled data for all the secondary subjects offered across the ITT/E provider's partnerships.

Analyses of what constitutes subject knowledge for teachers can be complex (Turner-Bisset, 1999). However, the components of the more straightforward models on offer (Shulman 1986, Banks, Leach and Moon, 2005) usually include subject content and skills, subject specific pedagogy and the curriculum requirements for the subject's learning, teaching and assessment (LTA).

---

[1] https://www.gov.uk/guidance/being-inspected-as-a-further-education-and-skills-provider

Shulman's influential model (1986) categorises these as Subject Matter Content Knowledge (SMCK), Pedagogical Content Knowledge (PCK) and Curriculum Knowledge (CK). Banks, Leach and Moon (2005) described professional knowledge for teachers in terms of subject knowledge, pedagogy and school knowledge. It is often difficult to separate such categories in practice or avoid the use of sub-categories (Lucero, Petrosini and Delgado, 2017). However, it is clear that Shulman's (1986) broad categories have general utility for analysing LTA (Lehane and Bertram, 2016; Kleickmann et al., 2015; Diezmann and Watters, 2015). For this reason, wherever possible, we have adopted Shulman's (1986) model in order to refer to aspects of subject knowledge for teachers. Elements of it are discernible within The Teachers' Standards (Department for Education, 2011) that intend to describe minimum performance for teaching competencies that trainees must demonstrate in order to achieve QTS in England. SMCK and CK appear partly in the Teachers' Standards descriptors for Teachers' Standard 3 and PK is contained partly within descriptors for Teachers' Standard 4 (Department for Education, 2011).

Assessment decisions in the NW ITT/E provider's partner schools were largely justified by mentor observations and the interpretation of documentary evidence. The evidence for final summative judgements was triangulated during a rigorous meeting involving trainee, mentor and HE tutor. Tynan and Mallaburn (2017) mapped Martin and Cloke's (2000) application of Hager and Butler's (1996) model for professional learning and its assessment to ITT/E programmes at the HE provider in the same North West of England that provided the data for this current study (Figure 1.). Applying this model, the assessment system and procedures described above are an example of the qualitative, judgemental assessment model suggested by Martin and Cloke (2000).

Martin and Cloke's (2000) model suggests that trainees arrive with a level of SMCK described by their previous qualifications and then develop their CK and PCK whilst training. Diezmann and Watters' (2015) case study of a professional microbiologist's transfer to the teaching profession sought to identify the effect of specialist discipline knowledge on the transition to teaching. They used a domain map of knowledge for Science, Technology, Engineering and Mathematics (STEM) teaching adapted from Hill *et al* (2008) where specialised discipline content knowledge is a component of SMCK. Interviews and classroom observations indicated that the career changer employed different subject matter whilst teaching microbiology compared to that possessed as a scientist and also identified a need to develop PCK. This finding challenged the assumption that specialised discipline subject knowledge was readily transferable to teaching.

Tynan and Mallaburn (2017) explored the possibility of using several different statistical tests to monitor the consistency or comparability of grades assigned across and within an HE provider's ITT/E programmes. Their study looked at partnerships across five ITT/E programmes at a HE provider in the North West of England. It demonstrated only positive, strong positive correlations between separate standard and overall teaching grades at each review point. Our investigation of grades assigned by school mentors for specific standards and overall teaching develops this quantitative approach. It focusses on assessment data from the Secondary Post Graduate Certificate/Diploma in Education (PGC/DE) Core and Non-Salaried School Direct Programme collected from two consecutive cohorts of trainees during the period September 2014 to July 2016. For each cohort, the analysis collates data from all possible review points within the programme to investigate in more depth possible relationships between the grades for overall teaching and the standards that refer to aspects of subject knowledge for teachers (Department for Education, 2011). The data were analysed for trainees in all subjects combined and, separately, for mathematics, English and science.
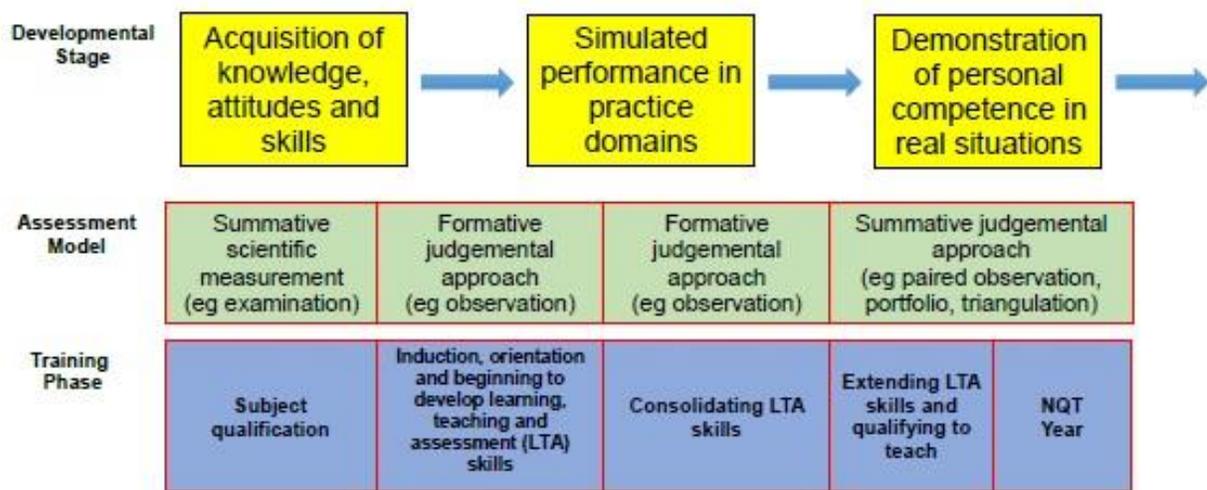
Figure 1. A model for the professional development of trainee teachers (developed from Hager & Butler (1996) and Martin & Cloke (2000).

Whilst policy makers have moved to strengthen teachers' specialist subject knowledge and skills (Department for Education, 2010), teacher trainers and educators may find it difficult to separate this from the additional elements of subject knowledge for teachers defined by Shulman (1986) and others. The aim of this study was to investigate any differences in the distribution of grades awarded by teachers acting as trainees' subject mentors in school for different aspects of subject knowledge compared to the grades assigned for teaching overall.

Methodology and methods

Practitioner research is often associated with local, small scale, qualitative research that draws criticism because the researcher is too close to the investigation and may be less than objective in seeking changes to the system under investigation (Anderson and Herr, 1999; Ebbutt, Worrall and Robson, 2000; Open University, 2005). However, even though it is quantitative and uses statistical analysis, this study fits well within a practitioner research model of investigation. It links to local perceptions of issues and opportunities around assessment outcomes and practice for secondary schools working in ITT/E partnership with a Northwest of England HE provider. It has the potential to identify issues and recommendations for future interventions and, as such, it corresponds to the early data-gathering phase of an action research cycle (Burton and Bartlett, 2009).

The statistical method used was the Chi Squared calculation and test of significance based upon observed and expected counts. Researchers in fields as varied as medicine, biology, social sciences and education have used the Chi Squared calculation and distribution in a variety of ways. Most introductory statistics texts (Hinton, 2014; Upton and Cook,1996) explain how Chi Squared calculations, used to test goodness of fit, compares observed counts with expected results predicted by theoretical models or known distributions e.g. allele frequencies predicted by Mendelian laws of genetics or normal distributions.

However, it is possible to derive expected results empirically from observed counts placed in contingency tables. Statistical texts (Hinton, 2014; Upton and Cook, 1996) explain Chi Squared used as a test of independence to compare two or more patterns of counts or frequencies. For instance, medical researchers use it to compare recovery rates from different treatments with the recovery rate for the trial as a whole. In a different but related use, plant ecologists have a long history of using Chi squared as a test for association to investigate if two plant species are found together, alone or absent more often than random distribution would predict (Dice, 1945).

Chi squared is a non-parametric statistical method that does not assume data has any particular distribution and does not require large populations or samples to give reliable conclusions. A statistical test used in this manner does not seek to identify predictive generalisations about any larger population of trainee teachers. It simply removes any subjectivity when deciding if differences between observed and expected counts for the local cohorts investigated were sufficiently large to stimulate further investigation into the possible reasons behind them. As such, the cohorts' assessment data do not constitute a sample but include all the data from the target populations.

For each subject or group of subjects, Chi Squared was used to test the hypothesis that there was no difference between the observed and expected numbers of grades assigned at different levels for a specific standard compared to those assigned for another standard, or for overall teaching. Specifically, we compared grades for Teachers' Standard 3 with Teachers' Standard 4 and both separately with overall teaching. Introductory statistical texts provide clear instructions for this approach (Langley, 1968). To illustrate the use of Chi Squared with the assessment grade data, Tables 1, 2, and their supporting text show one example, from the full data analysis, of a comparison that demonstrated the independence of two frequency distributions of grades assigned by school mentors.

Table 1. A contingency table showing observed counts and expected counts in brackets. (Expected counts for any box = row total X column total/ grand total).

Grade	All subjects Teachers' Standard 3	Overall teaching	Totals
1	96 (77)	59 (78)	155
2	189(190)	192(191)	381
3	66 (81)	96 (81)	162
4	3 (6)	10 (7)	13
Totals	354	357	711

Simple comparison of the observed and expected counts in Table 1 would suggest that the two sets of data are varying independently of each other. More Grade 1s than expected were awarded for Teachers' Standard 3 and fewer than expected for the overall teaching grade. Conversely, mentors awarded fewer Grade 3 and 4 than expected for Teachers' Standard 3 and more than

expected for the overall teaching grade. This approach compares both data sets with the overall frequencies calculated using the totals boxes as described in Table 1.

The statistical null hypothesis (H^0) for a Chi Squared calculation and test is there is no difference between observed and expected counts. Rejecting H^0 indicates independent distributions. The bigger the Chi Squared value calculated in Table 2 the smaller the probability of this set of results occurring if the H^0 is correct. The scientific standard for rejecting H^0 is a probability of 0.05 or a 5% chance of error. The test of significance described below indicated that the differences in Table 1 were large enough to represent a rare result if the two sets of data constituted similar distributions.

Statistical tables (Lindley and Miller, 1953) provided the probability of obtaining the Chi Squared value of 18.17, calculated in Table 2, occurring if H^0 was correct. The tables take into account the number of calculations summed to make this table or degrees of freedom for the data. For a contingency table the degrees of freedom are calculated as the (number of rows -1) x (number of columns -1). In this example, the table has four rows and two columns giving 3 degrees of freedom. The critical Chi squared value for rejecting H^0 with three degrees of freedom is 7.82 with 5% chance of error. As 18.17 is larger than this, H^0 was rejected. However, the tables give critical values for rejecting H^0 with lower chances of error, and 18.17 exceeds 16.27, which is the critical value for rejecting H^0 with only a 0.1% chance of error. Returning to Table 1 we can conclude that in 2014-15, across all the secondary PGDE subjects at this North West of England HE provider, assessors assigned more Grade 1s and fewer Grade 3s and 4s for Teachers' Standard 3 than for overall teaching performance. The chance of this being a false conclusion is less than one in a thousand.

Table 2. Chi squared calculation for Table 1.

Counts				
observed	expected	difference	difference^2	difference^2/expected
96	77	19	354.46	4.59
189	190	-1	0.48	0.00
66	81	-15	214.86	2.66
3	6	-3	12.06	1.86
59	78	-19	354.46	4.55
192	191	1	0.48	0.00
96	81	15	214.86	2.64
10	7	3	12.06	1.85
	Total	0	Total	18.17

We have similarly analysed data collected from the Secondary PGC/DE Core and Non-Salaried School Direct ITT/E programme during the academic years 2014-15 and 2015-16. Each year trainees received three formative and one summative assessment in order to complete formal progress review forms at the end of discrete training phases (Figure 1.). In the second year of the study, the programme decided not to award numerical grades for the second formative assessment. This was because it took place during a short school experience placement at a

different school to the other assessments. The statistical analysis compared all the graded assessment data across all review points for each year.

For each academic year, Teachers' Standard 3 and Teachers' Standard 4 counts were compared separately to the counts for the overall teaching grade and then with each other for four sets of trainees: all subjects combined, English, mathematics and science. H^0 in each comparison was: there was no difference between the observed and expected counts for the grades awarded. This was rejected if there was a 5% or less probability this conclusion being in error. The chance of error was read from standard statistical probability tables (Lindley and Miller, 1953).

Some data sets had different degrees of freedom to the example cited above. This was because in certain circumstances Chi squared calculations give too large a value increasing the chance of rejecting H^0 in error (Langley, 1968). To avoid this, where the expected count was five or fewer, counts for adjacent grade category rows were pooled. When pooling data resulted in a two by two contingency table with one degree of freedom then Yates' Correction was applied. This follows principles described in any basic text on statistical analysis (Langley, 1968).

The rationale for the approach adopted and described above makes several assumptions about the data and the assessment processes involved. Firstly, that grades for Teachers' Standard 3 recorded on the trainees' formal progress review forms are, at least to some extent, an assessment of SMCK and CK for teachers.

> 3. Demonstrate good subject and curriculum knowledge
> - have a secure knowledge of the relevant subject(s) and curriculum areas, foster and maintain pupils' interest in the subject, and address misunderstandings
> - demonstrate a critical understanding of developments in the subject and curriculum areas, and promote the value of scholarship
>
> (Department for Education, 2011:11).

Similarly, it assumes that Teachers' Standard 4 grades are, at least to some extent, an assessment of PCK.

> 4. Plan and teach well structured lessons
> - impart knowledge and develop understanding through effective use of lesson time
> - promote a love of learning and children's intellectual curiosity
> - set homework and plan other out-of-class activities to consolidate and extend the knowledge and understanding pupils have acquired
> - reflect systematically on the effectiveness of lessons and approaches to teaching
> - contribute to the design and provision of an engaging curriculum within the relevant subject area(s)
>
> (Department for Education, 2011:11).

This does not preclude SMCK, CK and PCK contributing to other standard descriptors nor claim that these standards only address subject knowledge areas.

The HE provider's guidelines for completing formal progress reviews instructed assessors to grade the individual standards and take into account the profile of grades before arriving at an overall teaching grade. If assessors followed these, then both Teachers' Standard 3 and Teachers'

Standard 4 grades contribute to the assessment of overall teaching and lead to large positive correlations between the distributions of grades assigned. Tynan and Mallaburn (2017) found this broadly to be the case in the academic year 2014-15 across the HE provider's full range of ITT/E programmes. Therefore, there is some basis for assuming that mentors followed the assessment guidelines and that any differences demonstrated give an indication of the relative importance attributed to Teachers' Standard 3 and Teachers' Standard 4 when determining overall teaching grades.

Results

For the statistically minded, the results of the Chi squared tests where significant differences were demonstrated are collated in Table 3. Figures 2 and 3 are visual presentations of the conclusions in Table 3. The main findings for the secondary PGDE cohort at the NW of England HE provider can be summarised as follows:

- For 18 out of 24 comparisons, there were no significant differences between the observed and expected counts.
- For 6 out of 24 comparisons, for Teachers' Standard 3 the numbers of top grades awarded were higher and the number of lower grades fewer than expected than for overall teaching or Teachers' Standard 4.
- All subjects taken together, there were more top grades and fewer lower grades awarded for Teachers' Standard 3 than overall teaching in both the academic years 2014-15 and 2015-16. In the first year, this was similar for Teachers' Standard 3 compared to Teachers' Standard 4.
- For mathematics, there were more top grades and fewer low grades awarded for Teachers' Standard 3 compared to Teachers' Standard 4 in 2014-15 and compared to overall teaching in 2015-16.
- In science, more top grades and fewer low grades were awarded for Teachers' Standard 3 than for overall teaching in 2014-15.
- Science and mathematics were different to the other core subject English where the number of grades awarded for overall teaching, Teachers' Standard 3 and Teachers' Standard 4 were comparable to that expected during both the years studied.

Key:

OTG   Overall teaching
S3   Teachers' Standard 3 grades counts
S4   Teachers' Standard 4 grades counts
⬅️   More high grades and less low grades than expected
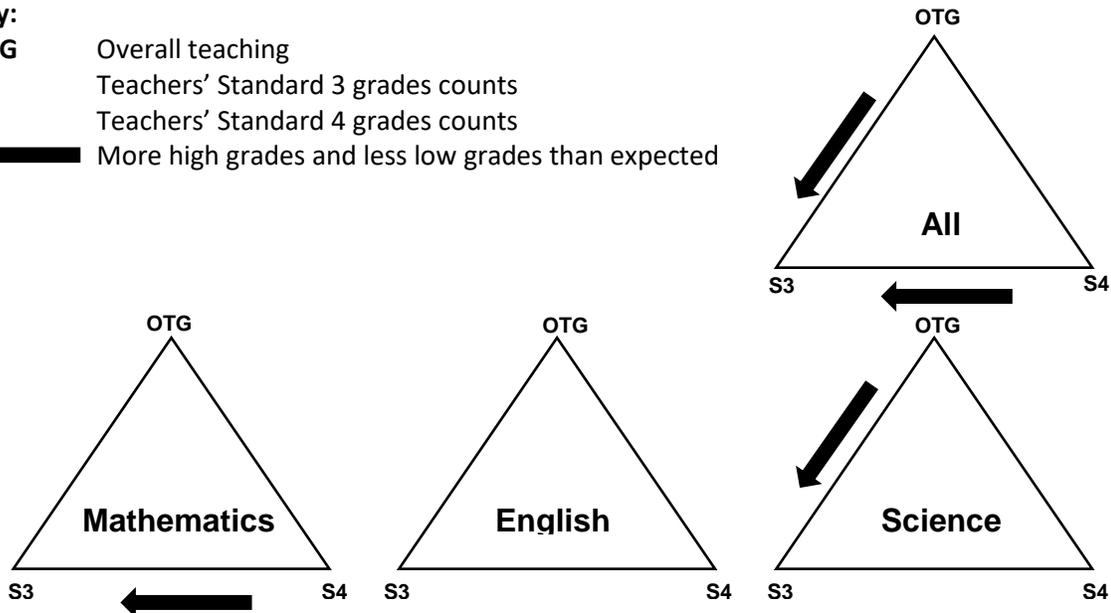
Figure 2. Visual presentation of Chi Squared analysis for the Academic Year 2014-15.

Key:

OTG   Overall teaching

S3   Teachers' Standard 3 grades counts

S4   Teachers' Standard 4 grades counts

⬅ More high grades and less low grades than expected
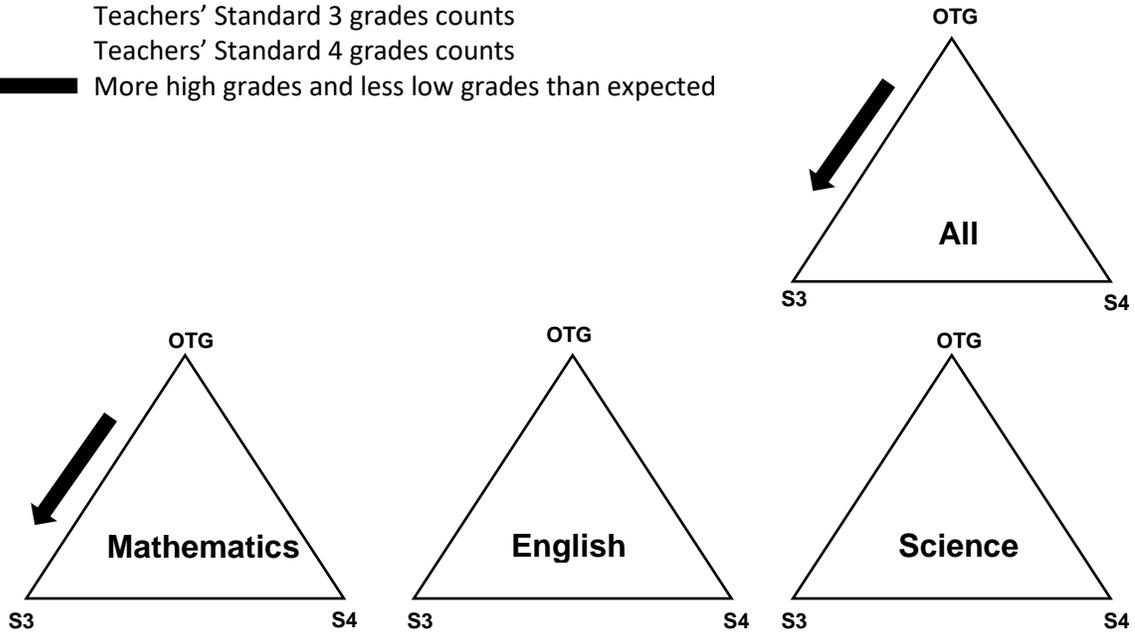
Figure 3. Visual presentation of Chi Squared analysis for the Academic Year 2015-15.

Table 3. Expanded conclusions for comparisons demonstrating significant differences between observed and expected frequencies of grades awarded.

Year	Subject	Assessments compared	H^0	Probability of Error	Number of Assessments	Conclusion
2014-15	**All**	Teachers' Standard 3 and OTG	Rejected	0.001	711	More Grades 1 and fewer Grades 2, 3 & 4 than expected for Teachers' Standard 3 compared to overall teaching grades
	All	Teachers' Standard 3 and Teachers' Standard 4	Rejected	0.01	707	More Grades 1 & 2 and fewer Grades 3 & 4 than expected for Teachers' Standard 3 compared to Teachers' Standard 4 grades

	Mathematics	Teachers' Standard 3 and Teachers' Standard 4	Rejected	0.05	150	More Grades 1 & 2 and fewer Grades 3 & 4 than expected for Teachers' Standard 3 compared to Teachers' Standard 4 grades
	Sciences	Teachers' Standard 3 and OTG	Rejected	0.01	142	More Grades 1 and fewer Grades 2, 3 & 4 than expected for Teachers' Standard 3 compared to overall teaching grades
2015-16	**All**	Teachers' Standard 3 and OTG	Rejected	0.01	767	More Grades 1 & 2 and fewer Grades 3 than expected for Teachers' Standard 3 compared to overall teaching grades
	Mathematics	Teachers' Standard 3 and OTG	Rejected	0.01	111	More Grades 1 & 2 and fewer Grades 3 & 4 than expected for Teachers' Standard 3 compared to overall teaching grades

Discussion

The tests of significance were applied according to their accepted principles and protocols, and offer valid conclusions with respect to the secondary PGDE/CE cohorts described. One drawback of making multiple single statistical comparisons is that the laws of probability necessitate the summation of errors. During the study, as a whole, there is a 9% probability that H^0 was rejected in error on one occasion. In addition, tests of statistical significance do not quantify the possible error when H^0 is accepted. However, applying vote counting, one of the simpler principles involved in meta-analysis (Cooper, 2017), to all the statistical conclusions, including the non-significant ones not presented in Table 3. suggests that the overall pattern of statistical conclusions would represent a rare event if caused only by random variation in the assessment grades.

For the two secondary PGDE cohorts studied, the results and main findings suggest differences between ITT/E mentors in core subjects during grading of Teachers' Standard 3, Teachers' Standard 4 and overall teaching. Overall, there is an acceptable level of confidence that there was a high degree of agreement between the numbers of observed and expected grades awarded for the standards relating to subject knowledge for teachers and for overall teaching. Where there were significant differences, the findings suggest that assessors in mathematics and science (and perhaps other subjects subsumed in the combined results for all subjects) linked the grade for Teachers' Standard 4 to their grade for overall teaching more often than Teachers' Standard 3. No significant differences occurred in English. In the second cohort studied, the number of significant differences reduced and Science became like English. In mathematics, the grades for Teachers' Standard 3 and Teachers' Standard 4 became congruent whilst the difference between observed and expected grades in Teachers' Standard 3 and overall teaching became significant. Such differences between grading in different core subjects suggest subject specific differences in subject knowledge for teachers or mentors' interaction with the assessment procedures described here.

There is some evidence that the categories suggested by Shulman (1986) and developed by many others are measurable and constant. This appears so, even in the subject knowledge of teachers from teaching systems in different cultural and national contexts (Kleickmann *et al*, 2015). There is certainly overlap between categories and difficulties measuring them separately (Kleickmann et al., 2015) but, even so, there is evidence that the categories are useful research tools. For instance, Lehane and Bertram (2016) examined one widely adopted measure of PCK and surveyed its many uses in education research.

There has been interest in the relationship between SMCK and PCK for some time. At around the same time that Shulman (1986) proposed his categories for teacher subject knowledge, Hashweh (1987) worked with a small group of teachers specialising in biology and physics to investigate the link between their subject content knowledge and their planning to teach a biology and physics topic from a text book. Subject knowledge was tested in three ways (including concept mapping) and their planning evaluated through a thinking aloud activity followed by eight questions about teaching each topic. Working within their own subject specialisms, the teachers demonstrated more content knowledge and knowledge of subject specific higher order principles and concepts. They also demonstrated the ability to link this to other areas within the subject and from the wider curriculum. This translated as greater independence and willingness to move away from the textbook treatment of the subject.

Also investigating this link, Lucero, Petrosini and Delgado (2017) focused upon knowledge of student conceptions (KOSC) as an indicator of PCK. Teachers at a large American high school answered SMCK questions and predicted their students' most likely alternate conceptions using a concept inventory. Their students answered the same inventory allowing the authors to compare the teachers' success in predicting the most likely alternate concepts with their SMCK scores. The authors viewed PCK as a multidimensional construct overlapping with SMCK but suggested there was a minimum level SMCK necessary to predict KOSC. Otherwise, they found no correlation between these two aspects of teacher subject knowledge at the school.

Earlier some of the perspectives on subject knowledge for teachers were described and linked to their location in the standards (Department for Education, 2011) in order to justify the use of

Teachers' Standard 3 and Teachers' Standard 4 as assessment measures of trainee subject knowledge. The work of Turner-Bisset (1999) on the professional knowledge demonstrated by History teachers has suggested that standards will never be able to provide more than a restricted description of teacher competency in this area. A quantitative study cannot reveal which, if any, model for subject knowledge teachers adopted during the grading process or their interpretation of the Teachers' Standards (Department for Education, 2011) descriptors with respect to this teacher competency. Further, the work of Hager and Butler (1996) and Martin and Cloke (2000) (illustrated in Figure 1) suggests the qualitative and potentially subjective nature of assessment during teacher training. For an insight into ITT/E assessors' differing approaches (holistic/analytical) when adopting competency based assessments see Leshem and Bar-Hama (2008). As assessors' experiences are likely to be different, it would be reasonable to assume that their views on how to assess their trainees might also differ, perhaps leading to more variation in assessment outcomes than demonstrated by the data. However, mentors tasked pragmatically with grading their trainees' teaching competencies will have, to some extent, met our assumptions about subject knowledge for teachers and Teachers' Standard 3 and 4 descriptors.

During the two years covered by this study, the provider aimed to reduce assessor subjectivity by increasing consistency of practice across its range of programmes and partnerships through a series of interventions agreed by school partners and supported by routine quality assurance procedures. Mentor training participation greatly increased leading to improved dissemination and implementation of the guidelines for evidence-based assessment using a set of agreed criteria. These agreed criteria extended the Teachers' Standards minimum performance descriptors (Department for Education, 2011) to guide the award of higher grades. The pivotal intervention most likely to have affected the grades assigned in the direction of the increasing consistency observed was referencing all assessments at all review points to the performance expected of trainees at the point of recommendation for QTS. This varies from written advice to assessors given in the Teachers' Standards (Department for Education, 2011) to take experience and stage of training into account but is, arguably, less subjective in the absence of any central or locally agreed performance criteria describing trainee teachers' performance at different stages of training. The interventions did not give specific guidelines on the evidence that might demonstrate different levels of subject knowledge for teachers nor how the grade descriptors might be interpreted.

Writing about the academic assessment of trainee teachers, Tummons (2010) suggested that quality assurance and managerial requirements might override complex assessment processes to influence the outcomes. If so, we might suspect something similar of the assessment of teaching competencies in school. In turn, this might, amongst other plausible possibilities, explain the high levels of consistency in the numbers of grades assigned for individual standards compared to overall teaching. However, using Shulman's model (1986), a speculation supported by the data and findings is that, assessors in certain subjects graded their trainees' SMCK and CK higher than their ability to teach their subject (PCK) and their overall teaching effectiveness. In the core subjects, mathematics and science demonstrated this but not English. This should not be particularly surprising as assessors and appraisers are currently guided to assess trainees and teachers by their perceived and measured impact on pupil learning. However, this effect reduced in the second year data suggesting the increased impact of interventions aimed at increasing consistency of assessment practice and outcomes across the provider's partnerships.

Conclusions and recommendations for further work

With respect to grades assigned for Teachers' Standards 3 and 4, and overall teaching, the data suggest, over two academic years, increasing consistency of grading outcomes between assessors on this PGDE programme. Given the qualitative nature of assessment and grading of trainees on school experience placement this is surprising and worth further investigation.

However, there is also evidence that Teachers' Standards related to different aspects of subject knowledge for teachers contributed differently to mentors' decisions about the overall teaching grade in English, Mathematics and Science.

Further, mentors consistently linked grades for Teachers' Standard 4 with overall teaching grades whereas grades for Teachers' Standard 3 were sometimes higher. This suggests that the mentors did always perceive subject matter content knowledge as an indicator of good pedagogy or overall teaching skill.

Statistical analysis only indicates the probability of the patterns observed arising by random variation. Qualitative research would be required to investigate what gave rise to increasingly high levels of consistency in the data and the reasons for the differences observed between trainees' grade distributions in English, mathematics and science. Explanations could lie with the nature of the assessment of competencies against standards using grading categories and/or differences in the nature of teacher subject knowledge in the core subjects.

Possible refinements in mentor training at the HE provider could in the future include materials exploring the application of models of subject knowledge for teachers in different subjects. These could also explore the possible impact of mentors' own models of subject knowledge for teachers on the assessment of trainees.

The government's current position is that teacher subject knowledge should be used as a strong indicator of teaching ability. This links to curriculum and examination changes encouraging the return to traditional subjects and single end-test methods of assessment. Against this backdrop, the relationship between the assessment of teacher subject matter content knowledge, pedagogical content knowledge and overall teaching skill will continue to be of interest to teacher trainers and educators.

References

Anderson, G. and Herr, K. (1999) 'The New Paradigm Wars: Is There Room for Rigorous Practitioner Knowledge in Schools and Universities?', *Educational Researcher*, 28:5, 12-21.

AQA (2016) Background to new GCSEs, Available at: http://www.aqa.org.uk/about-us/what-we-do/policy/gcse-and-a-level-changes/background-to-new-gcses (Accessed: 21/12/2016).

Banks, F., Leach, J. and Moon, B. (2005) 'Extract from new understandings of teachers' pedagogic knowledge', *The Curriculum Journal*, 16(3), pp.331-340.

Beck, J. (2012) 'Reinstating knowledge: diagnoses and prescriptions for England's curriculum ills', *International Studies in Sociology of Education*, 22(1), pp.1-18.

Burton, D. and Bartlett, S. (2009) *Key Issues for Education Researchers*. London: Sage.

Cooper, H. (2017) *Research Synthesis and Meta-analysis: A step-by-step approach: 5th Edition*. London: Sage.

Department for Education (2010) *The Importance of Teaching: The Schools White Paper 2010*.
    London: The Stationary Office.

Department for Education (2011) *Teachers' Standards: Guidance for School Leaders, School Staff
    and Governing Bodies*. Available at: www.gov.uk/government/publications/teachers-
    standards (Accessed: 22.12.17).

Dice, L.R. (1945) Measures of the amount of ecological association between Species, *Ecology*,
    26(3), pp. 297-302.

Diezmann, C. D. and Watters J. J. (2015) 'The knowledge base of subject matter experts in
    teaching: A case study of a professional scientist as a beginning teacher', *International
    Journal of Science and Mathematics Education*, 13, pp. 1517-1537.

Ebbutt, D., Worrall, N. and Robson, R. (2000) 'Educational research partnership: differences and
    tensions at the interface between the professional cultures of practitioners in schools and
    researchers in higher education', *Teacher Development*, 4(3), pp. 319-338.

Hager, P. and Butler, J. (1996) 'Two models of educational assessment', *Assessment and
    Evaluation in Higher Education*, 21(4), pp.367–378.

Hashweh, M. Z. (1987) 'Effects of subject-matter knowledge in the teaching of biology and
    physics', *Teaching & Teacher Education*, 3 (2) pp. 109-120.

Hill, H., Ball, D. L. and Schilling, S. (2008). 'Unpacking "pedagogical content knowledge":
    Conceptualizing and measuring teachers' topic-specific knowledge of students', *Journal for
    Research in Mathematics Education*, 39(4), pp. 372–400.

Hinton, P. R. (2014) *Statistics Explained, Third Edition*, London: Routledge.

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., Cheo, M. and Baumert, J.
    (2015) 'Content knowledge and pedagogical content knowledge in Taiwanese and German
    mathematics teachers', *Teaching and Teacher Education*, 46, pp. 115-126.

Langley, R. (1968) *Practical Statistics: for non-mathematical people*. Lichfield: Pan Piper.

Lehane, L and Bertram, A. (2016) 'Getting to the CoRe of it: A review of a specific PCK conceptual
    lens in science education research', *Educacion Quimica*, 27, pp. 52-58.

Leshem, S. and Bar-Hama, R. (2008) 'Evaluating teaching practice', *ELT Journal*, 62(3), pp. 257–
    265.

Lindley, D. V. and Miller, J. C. P. (1953) *Cambridge Elementary Statistical Tables*. Cambridge:
    Cambridge University Press.

Lucero, M. M., Petrosini, A.J. and Delgado, C. (2017) 'Exploring the relationship between
    secondary science teachers' subject matter knowledge and knowledge of student
    conceptions while teaching evolution and natural selection', *Journal of Research in Science
    Teaching*, 54 (2), pp 219-246.

Martin, S. and Cloke, C. (2000) 'Standards for the Award of Qualified Teacher Status: reflections
    on assessment implications', *Assessment and Evaluation in Higher Education*, 25(2), pp.
    183–190.

Shulman, L. S. (1986) 'Those who understand: knowledge growth in teaching', *Educational
    Researcher*, 15(2), pp. 4-14.

Tummons, J. (2010) 'Institutional ethnography and actor network theory: a framework for
    researching the assessment of trainee teachers', *Ethnography and Education*, 5 (3), pp. 345-
    357.

Turner-Bisset, R. (1999) 'The knowledge bases of the expert teacher', *British Educational Research
    Journal*, 25(1), 39-55.

Tynan, R. and Mallaburn, A. (2017) 'Consistency Counts- or does it?' *Teacher Education
    Advancement Network Journal*, 9(1), pp. 90-99.

Upton, G. and Cook, I. (1996) *Understanding Statistics*, Oxford: Oxford University Press.

Young, M. (2011) 'The return to subjects: a sociological perspective on the UK Coalition government's approach to the 14-19 curriculum', *The Curriculum Journal*, 22(2), pp265-278.

Can effect sizes give any clue to the way mentors ascribe numerical grades when assessing trainee teachers against the teachers' standards in England?

Rick Tynan and Robert Bryn Jones
Liverpool John Moores University

Abstract

Some teacher educators use numerical grades when assessing teaching competencies. In this situation, statistical analysis can be used to monitor consistency and look for correlations between assessment outcomes across teacher training partnerships and at different stages in training. Another approach is to calculate effect size metrics. These do not claim statistical significance but do seek to explain the practical impact of patterns in quantitative data. This study looks at number grade assessment data from a large secondary initial teacher education programme across schools working in partnership with a higher education provider in the Northwest of England. The proportion of variance between numerical grades for individual Teachers' Standards and overall teaching was calculated at each formal review point over three consecutive years. Despite the complex process involved in assessing teaching competencies against performance criteria and the potential for subjective variation between individual assessors, the data consistently demonstrated underlying patterns. These suggested that quality assurance and management of assessment issues could have been a major influence on the assessors.

Key words

Context and Review of Literature

Currently, Initial Teacher Education (ITE) programmes in England take place largely or entirely in schools, academies and colleges in partnership with providers who can accredit Qualified Teacher Status (QTS). This study is located in partnerships between secondary schools in the Northwest of England and a single Higher Education (HE) provider. School based mentors had first responsibility for both training and assessing trainee teachers subject to moderation and quality assurance by the provider. In England, HE and other providers are responsible for accrediting recommendations for the award of QTS. Such recommendations are based upon trainees demonstrating the minimum performance criteria described in the eight areas of teacher competency and section on professional expectations that are set out in the Teachers' Standards (Department for Education, 2011). The purpose of this investigation was look for clues in quantitative grading data to the priorities assessors gave to individual standards when considering overall teaching grades.

ITE partnerships in England are inspected and monitored by a government agency, the Office for Standards in Education (Ofsted). Ofsted judges ITE providers according to trainee outcomes (retention, grades and employment rates), the consistency of their training experience acrosspartnerships and the accuracy of mentors' assessments (Ofsted, 2018). The Northwest of

Citation

England HE provider in this study used Ofsted number grades for both formative and summative reviews of trainees' teaching skills in order to monitor and demonstrate their progress. Assessors all followed one possible assessment practice in England by numerically grading the eight individual standards and overall teaching at several formal review points during training. They used a four-point scale: 1 (Outstanding), 2 (Good), 3 (Requires improvement) and 4 (Inadequate). Although not all ITE providers use number grades to formatively assess their trainee teachers' performance, in England they must provide Ofsted inspectors with summative assessments of their trainees' teaching performance. In turn they are, themselves, judged on their ability to produce Good (2) and Outstanding (1) teachers (Ofsted, 2018).

In an effort to improve the consistency of assessment practice across a large number of partnerships, the HE provider in this study adopted some changes in 2011. These intended to improve the quality assurance of assessment practices. The provider also sought ways of demonstrating and monitoring consistency in the assessment grading outcomes of trainees. The steps taken were:

- Increased participation in mentor training by delivering this in partnership schools in addition to centrally, at the provider.
- Insisting on the central role of an agreed set of performance criteria, contained in an individual trainee standards tracking document, when making grading judgements against the Teachers' Standards (Department for Education, 2011).
- Adopting a rigorous and structured format for triangulation meetings between the trainee and mentor. These were chaired by a tutor from the provider and considered the evidence for the trainees' final indicative grades.
- Training tutors from the provider to emphasise their quality assurance and mentor training roles when visiting partnership schools.
- Agreeing clear documentation and guidelines through partnership steering groups.
- Emphasising preparation for inspection during mentor and visiting tutor training using feedback from external examiners and an Ofsted consultant.
- Using statistical tests to monitor consistency in grading outcomes and using this to inform training.

Practitioner researchers at the provider have found quantitative evidence of consistency that masks subjectivity in graded assessment outcomes across the partnerships monitored. Tynan and Mallaburn (2017) assumed that consistency in assessment practice would be reflected by consistency in assessment outcomes. They explored the use of statistical tests of significance to monitor consistency in numerical grades awarded by school based assessors. They demonstrated significant positive correlations between grades awarded for individual standards and the grades awarded for overall teaching. They also found consistency in final summative grades awarded for overall teaching across five ITE programmes (Tynan and Mallaburn, 2017). Whilst accepting there could be other explanations, Tynan and Mallaburn (2017) attributed their findings, at least in part, to the package of the interventions described above that were introduced to improve consistency of practice between assessors.

Tynan and Jones (2018) were interested in the assessment of trainees' subject knowledge on a secondary ITE programme and chose to focus on grades awarded for standards that include different aspects of subject knowledge for teachers. They used the most sensitive test of statistical significance

indicated by Tynan and Mallaburn's (2017) study to look at the relationship between grades awarded for Teachers' Standards S3 and S4 (Department for Education, 2011) and overall teaching grades in English, mathematics and science. Again, there was much consistency in the core subjects but in science and mathematics there were   occasions when significantly more high grades were assigned for the standard associated with subject content and curriculum knowledge compared to overall teaching or the standard more associated with pedagogy (Tynan and Jones, 2018). This hinted at some subjectivity between assessors in different core subject areas not readily apparent in Tynan and Mallaburn's (2017) wider survey and trial of statistical analyses.

However, there are a number of issues associated with the use of Ofsted number grades when assessing trainee teachers' performance that make achieving consistency and accuracy across a large number of partnerships problematic. The Teachers' Standards (Department for Education, 2011) give information on the minimum performance criteria necessary for the recommendation of QTS in England. However, the Teachers' Standards (Department for Education, 2011) contain neither guidance on appropriate assessment tools nor acceptable evidence to be used for assessments. Further, they do not contain any information on criteria for judging Good (2) or Outstanding (1) performance. ITE providers must use locally agreed criteria to evaluate performance above the minimum required. At the provider in this study, these were formulated initially by a consortium of local bodies involved in ITE. Over time, the provider has developed these with partnership schools using extrapolations from the Standards descriptors, Ofsted descriptions of the characteristics of undergraduate final year trainees and, more recently, clues from Ofsted ITE partnership inspections. Inherent in this approach is the opportunity for subjective differences between regions and local partnerships in the choice of assessment tools, construction and interpretation of criteria and choice of evidence when assessing trainee teachers.

In addition to the potential sources of variability inherent in basing assessments on Ofsted categories coded as numbers, there are some theoretical issues that predict that more variability in assessment grades might be expected than the practitioner investigations cited above actually demonstrated. The way in which assessors perceive professional learning could have important implications for their approach to assessment and pose another potential source of subjectivity and variation in grades. Philpott (2014) provided a summary and critical review of a range of models for professional learning and their implications for teacher educators. These constituted a continuum with individual cognitive and psychological approaches at the opposite end to those that consider learning to be a social construct. For instance, Kolb's (1983) model, which focuses on the acquisition of knowledge and skills through experiential learning, seems to invite the assessor to concentrate on the aspiring teacher's performance. On the other hand, Wenger's (1998) model, which emphasises acceptance into a community of practice, suggests judgements based upon norms, expectations and aspiring practitioners' perceived impact on learners as clients (Philpott, 2014).

Hager and Butler (1996) considered two models of assessment to be necessary during professional learning and Martin and Cloke (2000) applied these to the assessment of teaching competencies. They contended that, whatever model of professional learning is assumed, the evaluation of trainee teachers becomes more judgemental and qualitative and less measurable scientifically as professional learning proceeds. Tynan *et al* (2014) compared the assessment outcomes of Post Graduate Certificate in Education (PGCE) science trainees preparing to teach chemistry and physics. Some arrived at the HE QTS provider with a first degree in their specialist teaching subject whilst others arrived with a one year Subject Knowledge Enhancement (SKE) qualification accredited by an HE Certificate. Tynan *et al*

(2014) found that school based assessors did not distinguish between these two groups of science trainees and that the grades awarded for subject knowledge and overall teaching ability were similar no matter the level of qualification in chemistry or physics. As it would seem impossible to cover as much science subject content in one year compared to a three year undergraduate course, these findings would seem to suggest that assessors were assessing subject knowledge in a different manner to that used at the end of a first degree. This would seem to validate the application of Martin and Cloke's (2000) model in that context. Tynan and Mallaburn (2017) mapped this model to the delivery and assessment of ITE programmes at one HE provider and noted the implications for increased variability in grades if the model was assumed to be valid.

Criteria based assessment of competencies can be viewed as an approach aimed at reducing the subjectivity inherent in a qualitative judgemental assessment model. Leshem and Bar-Hama (2008) discussed issues around the introduction of criteria based assessment of teaching competencies to the Israeli ITE system. They explored the attitudes, perceptions and preferences of tutors and students in comparing this more analytical approach to the previous practice of using professional judgement more holistically. Tutors in their study reported different approaches to using criteria for assessing teaching competencies. Some started with a holistic judgement and used the criteria as a check, whilst others started more analytically with the criteria and then compared the resulting assessment outcome against their professional judgement. No matter their approach, tutors noted difficulties in reconciling their holistic judgments with criteria based assessment. In the context of this study, assessors following the guidelines agreed by partnership schools and the HE QTS provider graded individual standards first and then used these to arrive at an overall grade for teaching. However, as Leshem and Bar-Hama (2008) reported, some assessors may find this approach difficult and could have started with a holistic assessment and grading of overall teaching ability and then graded the individual standards accordingly afterwards.

Tummons (2010a, 2010b and 2011) has also considered in depth assessment across PGCE programmes provided by a northern university. Tummons (2010b) investigated the validity and reliability of assessments of trainee lesson plans and also the issues associated with making valid assessments of trainees' reflective practice (Tummons, 2011). However, when considering possible reasons for consistency in number grade assessments, Tummon's (2010a) application of institutional ethnography (IE) and actor network theory (ANT) to the assessment of post graduate trainees appear useful. This approach perceives assessment as closely governed by IE and ANT. IE can be described as the way an institution documents its courses and assessment activities which, in turn, becomes inseparable from the way these documents are sponsored by tutors and teachers (ANT). The application of this approach to student teacher assessment led Tummons (2010a) to suggest that complex assessment activities had been subsumed in practice by quality assurance and managerial issues. These ideas might help explain the high levels of consistency in grading assessment data reported by Tynan and Mallaburn (2017) and Tynan and Jones (2018) for competency based assessment against the Teachers' Standards (Department for Education, 2011).

Methodology and methods
This study constitutes local, small scale, practitioner research involving one secondary ITE programme at a single HE QTS provider in the Northwest of England. It is a quantitative survey and analysis of numerical grades for individual teaching standards and overall teaching. These were collated from formal progress review forms during the period September 2014 to July 2017. The programme selected was the largest of those available for study at the QTS provider, which earlier work (Tynan

and Mallaburn, 2017) suggested was representative of the other programmes. School based mentors routinely assessed trainees preparing to teach a range of subjects during their Post Graduate Diploma in Education/Certificate in Education (PGDE/CE) courses. All the trainees were in secondary schools. Trainees were assessed against the Teachers' Standards (Department for Education, 2011) using the descriptors in the standards and a locally produced trainee progress tracking document used across all the partnerships within the programme. Numerical grading data on formal review forms were collected from three consecutive cohorts.

Quantitative studies based upon tests of statistical significance can be criticised if they omit to attempt an explanation of the practical significance, or impact in everyday terms, of their statistical findings (Ellis, 2010). Proportion of variance (POV) is one approach to addressing this using an r-family effect size metric that looks at the practical impact of correlations (Ellis, 2010). A further advantage of using an effect size metric is that the statistic is scale free. This allows comparison of data from different studies (Ellis, 2010) and, in this study, the data collected from three different years, despite differences in cohort sizes.

Previous work by Tynan and Mallaburn (2017) on numerical grades established that the use of either Pearson's or Spearman's Rank correlation coefficients (r) led to identical statistical conclusions (for a starter statistical text see Hinton, 2014). No matter which test was used, positive correlations were demonstrated between the grades awarded for individual standards and overall teaching, with very low probabilities of these being due to random patterns in the data (Tynan and Mallaburn, 2017). In light of this finding, the most straightforward calculation was adopted. Pearson's correlation coefficient (r) was calculated using the standard function formula available in standard spreadsheet software. Grades for each individual standard were compared to grades for overall teaching at every formal review point where number grades were ascribed. POV (r^2) was calculated by squaring r (Ellis, 2010) using a standard spreadsheet formula. This effect size metric can be reported as the proportion of variance or as expressed as the percentage of the variation shared by two sets of data, simply by multiplying r^2 by one hundred (Ellis, 2010). Percentages may be more intuitively understood than proportions written as decimals.

Considering POV does not seek to prove statistical significance but to establish a practical indication of the size of an effect, no matter the cause. The aim of this study was to use POV to establish if the grades assigned for some individual standards might be linked more closely than others to overall teaching grades. In turn, this might give clues to the priorities ascribed to different standards by assessors when deciding on an overall grade for teaching.

Several qualifications should be noted before considering the findings below. POV metrics are calculated from correlation coefficients. However, demonstrating correlations between sets of grades does not describe causal reasons for them, if any exist. Further, even if the correlations underlying the calculations of POV are significant, any discussion of the differences between effect sizes must include the distinct possibility that these might be the result of chance variation in the data. Lastly, the cohorts were treated as full populations not samples, so there is a high degree of validity in the data for the programme investigated but no claim that the findings from this programme should be extrapolated to a larger population of trainee teachers.

Findings
Table 1 provides a quick reference to the Teachers' Standards headings (Department for Education, 2011). In the interests of reducing the amount of statistical data presented, only the results for the

final summative assessment point are presented in full (Table 2 and Figure 1 below). However, in Table 3 (below), for every review point, the pairs of standards whose grades shared the highest and lowest percentage variation overlap with the grades for overall teaching are presented. The difference in the percentage overlap from first to last ranked standard is also reported as an indication of the range of the effect sizes.

Table 1: Key to Part 1 Teachers' Standards headings (Department for Education, 2011)

1. Set high expectations which inspire, motivate and challenge pupils.
2. Promote good progress and outcomes by pupils
3. Demonstrate good subject and curriculum knowledge
4. Plan and teach well-structured lessons
5. Adapt teaching to respond to the strengths and needs of all pupils
6. Make accurate and productive use of assessment
7. Manage behaviour effectively to ensure a good and safe learning environment
8. Fulfil wider professional responsibilities

Main Findings
There were differences each year between the POV for summative grades ascribed for individual standards and overall teaching (Table 2 and Figure 1)

Table 2: The percentage proportion of variance (POV) for the final summative grades of individual standards compared to grades awarded for teaching overall and their yearly rankings.

Teachers' Standard	Proportion of variance (%)			Ranking		
	2014-15	2015-16	2016-17	2014-15	2015-16	2016-17
1	65	56	64	3	4	3
2	70	60	68	1	2	2
3	27	24	47	8	8	6
4	66	64	69	2	1	1
5	52	57	57	5	3	4
6	48	52	51	6	5	5
7	54	50	43	4	6	7
8	41	35	37	7	7	8
n	85	126	238			
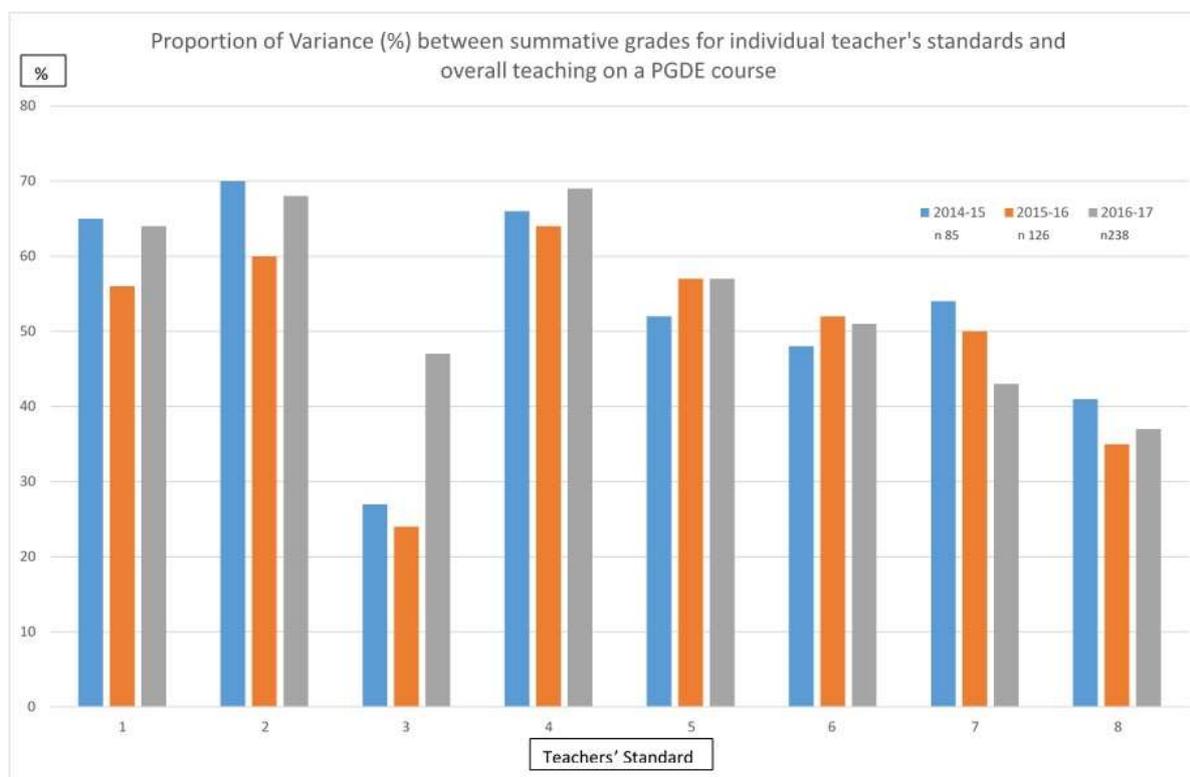
Figure 1. The percentage of shared variation between summative numerical grades ascribed for individual Teachers' Standards (Department for Education, 2011) and overall teaching over three consecutive years.

At all review points where number grades were ascribed, there were patterns in the POV for grades for individual standards and overall teaching that were similar over a three year period (Table 3).

Table 3: Pairs of Teachers' Standards (Department for Education, 20011) with the highest and lowest proportion of variance (POV) for a Secondary ITE Programme leading to QTS for all review periods assessed with number grades.

	2014-15		2015-16		2016-17
Highest Ranking Standards					
First formative review	1 & 4		1 & 5		2 & 4
Second formative review	2 & 4		-		-
Third formative review	2 & 4		1 &4		2 & 3
Final summative review	2 & 4		2 &4		2 & 4
Lowest ranking standards					
First formative review	6 & 8		6 & 8		6 & 8
Second formative review	3 & 8		-		-
Third formative review	3 & 8		3 & 6		7 & 8
Final summative review	3 & 8		3 & 8		7 & 8

For formative review points during 2014-2017, Standards 1, 2 and 4 were most likely to share the highest percentage of their variation in grades, with overall teaching and standards 3, 6 and 8 most likely to share the least (Table 3). For final summative grades during 2014-2017, Standards 2 and 4 always shared most variation with overall teaching and Standards 3 and 8 were most likely to share the least (Table 3 and Figure 1).

One possible interpretation of these findings is that the different effect sizes are caused by non-random differences in the way that number grades were ascribed. If this were the case then the differences in effect sizes could indicate differences between the ways assessors perceive different standards' contribution to overall teaching performance. For instance, for summative grades, one interpretation could then be that assessors associated the grades for Standard 2 and Standard 4 much more closely with the grade ascribed for overall teaching than the grades for Standard 3 and Standard 8 (Table 1). This approach does not help identify reasons why this might be the case, but the consistency of this pattern over three years encourages the idea that the pattern is non-random and a further investigation into the possible reasons for the pattern would be worthwhile.

Discussion

The interpretation of POVs as percentages (Ellis, 2010) allows an easily accessible interpretation of the practical significance of the correlation between the grades for a particular standard and overall teaching. For example, in the first line of Table 2 the POV for Standard 1 was found to be 64% in 2016-2017. This can be interpreted as 64% of the variation found in the grades for overall teaching and Standard 1 was common to both. Intuitively, this would seem to indicate that assessors were placing more importance on Standard 1 than Standard 8 which only shared 37% of its variation with overall teaching. Of course, this is a risky interpretation as neither correlations nor shared variation can be used alone to establish causal relationships. Further, the use of the effect size metric invites the interpretation that 64% of the variation in overall teaching grades was due to variation in the grades for Standard 1, whereas only 37% was due to variation in grades for Standard 8. This is one possible explanation but assumes that the differences in POV were due to non-random causes and that assessors had followed the agreed guidelines by using a profile of grades for individual standards to arrive at an overall teaching grade. Neither may be the case.

However, it is not necessary to make any conclusions about potential causal relationships for correlations for their effect size metrics to be useful tools. When tracking consistency in assessment outcomes and practices it is sufficient to know that there may be a trend that needs further monitoring whilst more qualitative evidence is gathered. Effect sizes are scale free and can be compared directly across different data sets collected during an investigation (Ellis, 2010) or the meta-analysis of data from different studies (Cooper, 2017). This allows the comparison of the effect sizes for different standards in the same year and also effect sizes for the same standard in different years. For example, in Table 2, in the academic year 2014-15, POV values suggest that assessors linked Standard 4 most often and Standard 8 least often with overall teaching grades. Using Table 1, this might suggest that assessors mentally associated a trainee's pedagogical knowledge more often with overall teaching ability than their contribution to wider school responsibilities. Also from Table 2, it can be seen that the POV values for Standard 1 differ in different academic years. This may indicate changing attitudes to the relative importance of this standard and other standards over time. However, such ideas have to be considered very cautiously as variations in effect size metrics could still represent chance fluctuations in data (Ellis, 2010).

The consideration of holistic and analytical competency based assessment practices by Leshem and Bar-Hama (2008) and Hager and Butler's (1996) discussion of the qualitative judgemental assessment model, applied to education by Martin and Cloke (2000), both appear to indicate that a degree of subjectivity amongst a large group of assessors should be expected. Similarly, differences in assessors' views on the psychological (Kolb, 1983) or social (Wenger, 1998) nature of professional learning could also lead to subjective differences in the way they ascribed numerical grades. A consideration of POVs has indicated more variation than was demonstrated by the use of statistical tests of significance by Tynan and Mallaburn (2017) and Tynan and Jones (2018). The variation in POV values between standards and the same standards over time may suggest the subtle effects of differences between assessors on grading outcomes. However, there were also consistent patterns in the POVs calculated and one possible explanation may lie in the interventions listed previously that were successfully applied during the period of this study. These encouraged a standards-first approach and sought to improve consistency of assessment practice across the school and HE provider partnerships involved. However, it is interesting to note that the steps implemented focused upon agreed procedures and consistency of outcomes. Assessors and liaison tutors were not formally called upon to consider or question the assessment process by which judgements and number grades were assigned.

A consideration of the contents of Table 3 suggests that this pragmatic approach may have affected the way assessors ascribed grades for individual standards and overall teaching. During the period 2014-2017 the partnerships were awaiting inspection by Ofsted and increasing emphasis was placed on the interventions described previously that were implemented to improve the quality of partnerships in line with published Ofsted ITE inspection criteria (Ofsted, 2018). Participation in training for mentors of trainees in school greatly increased and the reference point for awarding grades was the descriptors in the Teachers' Standards in England (Department for Education, 2011) and a locally agreed set of performance criteria. It was considered very important that summative grades were agreed at rigorous and structured triangulation meetings involving trainees, their mentors and the visiting tutor from the HE QTS provider in a quality assurance role. There can be little doubt that information considered important to establish consistency in preparation for inspection was cascaded repeatedly to mentors.

The three year patterns in POVs for summative grades are congruent with the information disseminated. Namely that it would be difficult to justify numerical grades for overall teaching that were widely different to the grades of some important standards. Standards 2 and 4 were considered the best indicators of the overall teaching grade. Standards 5 and 6 were also considered to be important predictors. This is not to suggest that this is or ever was actual Ofsted practice during inspections of ITE partnerships in England but merely to record the advice that was cascaded to all interested participants during the period of this study.

Tummons (2010a) found it useful to consider IE and ANT when discussing assessments in ITE. Tummons did not look specifically at competency based assessment of trainees on teaching experience placement by school mentors against the Teachers' Standards (Department for Education, 2011). However, the findings in Table 3 are congruent with the idea that the ITE programme's documentation together with its sponsors, the school liaison tutors and school mentors, were successful in sharing and implementing the programme's messages on consistency in grading in preparation for Ofsted. For formative review point grades, other standards might be included in the top pair and an inspection indicator standard might be included in the bottom pair. However, for every

year during the study, the summative grades for Standards 2 and 4 shared the highest POV with overall teaching and the summative grades for Standards 5 and 6 were in the middle group and never in the lowest pair.

It is difficult to imagine that chance fluctuations might give rise to the same pattern over three consecutive years. This might indicate that, whilst some assessor subjectivity was possible during formative grading, for summative assessments quality assurance and assessment managerial issues could mask the assessment process and its associated sources of variation. This could constitute a further example of IE and ANT in ITE assessment similar to that suggested by Tummons (2010a).

Conclusions

The use of an effect size metric for the quantitative investigation of grades assigned for individual standards and overall teaching has suggested some findings about the way assessors in schools ascribe grades and raised further questions. These are highly relevant to practitioners in the programme and institution studied and may be of general utility to all teacher educators involved in assessing trainee teachers against criteria describing teaching competencies.

In this study, the number grades assigned did not reflect the degree of variability predicted by several procedural and theoretical considerations. Further, consistency in numerical grading outcomes may have reflected consistency in assessment practices between assessors but there is a suggestion that compliance with quality assurance and management of assessment issues also contributed. In the minds of the authors, this begins to build a case questioning the use of numerical grades for individual standards and overall teaching during the formative and summative assessment of trainee teachers.

Professional learning deserves a valid and reliable assessment process with the aim of assessment to produce the most effective teachers possible. Given the issues discussed previously concerning the application and extrapolation of the Teachers' Standards (Department for Education, 2011), ascribing number grades or categorising teachers' teaching performance in these circumstances invites the subjective use of professional judgement whilst leading an external observer to believe that this is something that can be measured scientifically (Martin and Cloke, 2000). The Teachers' Standards (Department for Education, 2011) may provide a useful analytical approach for mentoring and coaching aspiring teachers. However, the use of number grades may not facilitate meaningful assessment and, particularly in the formative stages of an ITE programme, may reduce the validity and utility of feedback to trainee teachers.

At the HE provider studied, statistical analysis has been used to demonstrate high levels of consistency in numerical grading outcomes across partnerships, programmes and time (Tynan and Mallaburn, 2017, Tynan and Jones 2018). The findings of the current study give clues that this may be due to the effects of successful managerial and quality assurance interventions that ensure compliance with assessment guidelines rather than reducing subjective differences in assessment practice between assessors. In their present format the use of number grades and categories when assessing teachers may be masking actual assessment processes, which may be more valid and reliable than the current practice.

Next steps

Further qualitative projects are in progress or in the planning stages at the ITT/E HE QTS provider that seek to answer questions that cannot be addressed by further quantitative studies:

- How far can the standards descriptors be trusted to guide assessment of trainees when considering performance above the minimum required for QTS?
- What are the tensions perceived between assessment and its practice?
- Is potential subjectivity between assessors an issue?
- How do assessors perceive assessment using performance criteria?
- What evidence do assessors use to ascribe grades and how do they use it?
- Does a number grade approach lead to spurious perceptions of accuracy in, what is essentially, a qualitative assessment system?
- Can quality assurance and management considerations across many partnerships allow complex assessment processes to be implemented fairly and reliably?

References

Cooper, H. (2017) *Research Synthesis and Meta-analysis: A step-by-step approach: 5th Edition*. London: Sage.

Department for Education (2011) *Teachers' Standards: Guidance for School Leaders, School Staff and Governing Bodies*, Crown copyright 2013. Available at: www.gov.uk/government/publications/teachers-standards

Ellis, P. D. (2010) *The essential guide to effect sizes: Statistical Power, Meta-Analysis, and the interpretation of Research Results*, Cambridge University Press.

Hager, P. and Butler, J. (1996) Two models of educational assessment, *Assessment and Evaluation in Higher Education*, 21(4), 367–378.

Hinton, P. R. (2014) *Statistics Explained*, *Third Edition*, London: Routledge.

Kolb, D. A. (1983*) Experiential Learning: Experience as the Source of Learning and Development*, London: Prentice Hall.

Leshem, S. and Bar-Hama, R. (2008) 'Evaluating teaching practice', *ELT Journal*, 62(3), pp. 257–265.

Martin, S. and Cloke, C. (2000) 'Standards for the award of qualified teacher status: reflections on assessment implications', *Assessment and Evaluation in Higher Education*, 25(2), 183–190.

Ofsted (2018) *Initial teacher education inspection handbook*, Crown copyright

Philpott, C. (2014) *Theories of Professional Learning: A Critical Guide for Teacher Educators*, Northwich: Critical Publishing.

Tummons, J. (2010a) 'Institutional ethnography and actor network theory: a framework for researching the assessment of trainee teachers', *Ethnography and Education*, 5 (3), pp. 345-357.

Tummons, J. (2010b) 'The assessment of lesson plans in teacher education: a case study in assessment validity and reliability', *Assessment & Evaluation in Higher Education*, 35(7), pp. 847-857.

Tummons, J. (2011) 'It sort of feels uncomfortable': problematizing the assessment of reflective practice', *Studies in Higher Education*, 36(4), pp. 471-483.

Tynan, R. and Jones, R. B. (2018) 'Assessing trainee secondary teachers on school placement: subject knowledge and overall teaching grades', *TEAN Journal*, 10(1), pp. 20-34.

Tynan, R. and Mallaburn, A. (2017) 'Consistency Counts- or does it?' *TEAN Journal*, 9(1), pp. 90-99.

Tynan, R., Mallaburn, A., Jones, R. B. and Clays, K. (2014) 'Subject knowledge enhancement (SKE) courses for creating new chemistry and physics teachers: do they work?', *School Science Review*, 95(353), pp. 85–94.

Wenger, E. (1998), *Communities of Practice: Learning, Meaning and Identity*, Cambridge University Press.

Attitudes to assessing trainee teachers on school experience placement within a group involved in an initial teacher education partnership at an 11-16 school: a Q-methodology approach

Rick Tynan and Matt McLain
Liverpool John Moores University

Abstract
In England the recommendation and award of qualified teacher status (QTS) is currently linked to the assessment of trainee teachers' competencies against performance criteria descriptors. Q-methodology was used to look for subjective differences in attitudes to the assessment of trainee teachers in school. This is a quantitative approach to qualitative research combining the best of both. It statistically compares participants' perceptions of a wide range of ideas whilst demonstrating clearly the qualitative differences between any groups showing subjectivities in their responses.

A small group involved in initial teacher education at an 11-16 school took part. The school was in partnership with a higher education QTS provider. Each respondent independently completed an anonymous on-line sorting exercise using a concourse of 41 statements about the assessment of trainees. This was constructed from appropriate literature, national policy and the partnership's documentation and guidelines. The group included a visiting tutor, two school mentors and five trainees. Their responses were analysed using standard Q-methodology software. Participants demonstrated a clear consensus about prioritising statements concerning compliance with national requirements and local guidance. However, factor analysis identified one group, one pair and two individuals with subjective differences in their levels of agreement with statements about the assessment of trainees' teaching competencies.

A group of four, that included one mentor and his mentee, prioritised statements linked to fairness, validity and quality assurance processes within and external to the school. The other mentor and her mentee perceived assessment almost solely in terms defined by the Teachers' Standards in England. This suggested that the different perspectives on assessment were associated with influences other than the respondents' differing roles within ITE.

Key words
Q-methodology, subjectivity, assessment, Teachers Standards, competencies, criteria, descriptors, ITE, trainees, school placement

Context and Literature Review
Initial Teacher Education in England
For Robinson (2006) the history of initial teacher education (ITE) in England has been characterised by two important themes. There have been broad cyclical swings between school led models of provision and those based in college or university. These have been influenced by the link between ITE and the politics and provision of state education. Robinson (2006) was writing at the beginning of, what turned out to be, an extended and on-going period of central government control of teacher education with an emphasis upon school-led professional training and the multiplication of work-based routes into teaching. Over time, this approach to ITE in England has developed and extended the mentoring role

of school based teacher educators. The assessment of trainees' teaching has become a key part of the mentor's role (Department for Education, 2011).

Moore (2004) argued that the perceptions held by policy makers and the public concerning the attributes of good teachers have implications for ITE and the assessment of teachers. Moore (2004) identified and discussed three dominant discourses concerning good teachers. These described teachers in terms of the competent craftsperson who could learn the job from others, the reflective practitioner who could learn from their experience and the charismatic subject who was born rather than learned to teach. These discourses appear to persist and co-exist in current discussions of professional learning and school-led ITE (Philpott, 2014; Door, 2014) and the role of teacher educators (Czerniawski, 2018). However, it would be over simplification to identify the teacher as a competent craftsperson with school based training and the reflective practitioner with teacher education undertaken by HE providers (Jones and White, 2014). Moore (2004) predicted that the existence of such dominant discourses about teachers would make others more difficult to conceive and less likely to emerge.

Currently, higher education (HE) and school based ITE providers in England recommend trainee teachers for qualified teacher status (QTS) if they demonstrate competency in eight teaching standards (Department for Education, 2011). The Teachers' Standards (Department for Education, 2011) intend to describe learning, teaching and assessment using a finite list of competencies that set out criteria for minimum levels of performance. Training takes place mainly in schools involving partnerships between placement schools and HE or school-based QTS providers. The Office for Standards in Education (OfSTED) regularly inspect ITE partnerships. As one measure of trainee outcomes, inspection teams use a four-point numerical scale based upon the number of trainees judged to be performing at different levels (OfSTED, 2018). In turn, some providers in England use the OfSTED (2018) scale to assess teaching competence by ascribing number grades for each standard and overall teaching. The participants in this Q-method study were located in a partnership that used the number grade approach for formal formative progress reviews in addition to the summative assessment of trainees.

Cajkler and Wood (2016, p1) used the term 'reductive models' for ITE approaches such as this. They suggested that the need to gather evidence to meet standards must restrict training experiences and advocated a collaborative approach to planning, observing and evaluating lessons (lesson study) as a more effective way of developing classroom skills (Cajkler and Wood, 2016). However, the move in England towards describing teaching in terms of a list of competencies can be located alongside similar trends observable in the ITE provision of other countries. For example, in the USA Stiggins (1999) wrote as a member of the National Council on Measurement in Education (a professional association), and as President of a commercial publishing and training body, the Assessment Training Institute. Stiggins (1999) discussed self-evaluation by North American teacher education institutions to improve training provision in assessment of learners as a teacher skill. To support the need to describe the assessment of learning using key competencies, Stiggins (1999) reported a survey illustrating the issues faced by central policy makers in the USA at the time. This revealed a wide variation across states in requirements for aspiring teachers to demonstrate this key area of teaching skill. Fifteen states tested performance in some way whilst a further ten only required a piece of coursework on the assessment of learning. Twenty-five states did not require any formal demonstration of a teacher's ability to assess learners.

Assessing teachers and the potential for subjectivity
The Teachers' Standards (2011) give clear guidance that head teachers and their delegates have the first responsibility in schools for assessing teachers. However, there is no guidance on mechanisms or

the evidence for assessment. School based ITE mentors are free to choose their methods of assessment subject to internal and external quality assurance procedures. In the early stages of competency based assessment using standards in England, Jones (2001) compared the way 25 English and 25 German ITE mentors perceived their roles. Both groups agreed on their advisory role but English mentors placed more importance on their assessment role than their German counterparts. They were also less likely to perceive difficulties associated with the assessment of trainees or potential conflicts between their assessment and other mentoring roles. Anecdotal experience from our partnerships would suggest mentors perceive their assessment roles pragmatically in much the same way as the mentors from England in this early small-scale study.

Utilising extensive research and consultation, the Centre for the use of Research and Evidence in Education (CUREE) produced supporting materials setting out principles for mentoring and coaching teachers (Mentoring and Coaching CPD Capacity Building Project, 2005). The Welsh Government (2014) later updated and adopted CUREE's national framework. Both documents describe differences between mentoring, co-coaching and specialist coaching. They set out mentoring mainly in terms of identifying goals, formative assessment and giving feedback to guide trainee progress. Summative assessment was mentioned briefly as 'assessing, appraising or accrediting practice' (Mentoring and Coaching CPD Capacity Building Project, 2005, p5; Welsh Government, 2014, p6). The most recent guidance (Department for Education, 2016) sets out non-statutory national standards for school based ITT mentors in response to the Carter (2015) review of ITT provision. These re-visit some of the themes previously identified in the CUREE (2004) national guidelines. None of these documents intend to guide mentors on how to assess trainee progress against performance criteria. However, the mentor standards document (Department for Education, 2016) draws upon several case studies that indicate assessment was an ongoing issue for partnerships as they evaluated different strategies for improving accuracy, feedback, and moderation preparing for inspections.

Hager and Butler (1996) proposed a simple process model for professional development that linked professional learning with assessment models. Martin and Cloke (2000) later applied this model to teaching standards in England. Using this model, the final stages of ITE programmes require trainees to demonstrate their professional competence in real life situations. As trainees become independent and take on individual responsibility for their classes' learning, their mentors can increasingly concentrate on a summative assessor role. At this stage, the assessment model becomes judgemental and based upon qualitative evidence (Hager and Butler, 1996; Martin and Cloke, 2000). This model predicts many opportunities for subjective differences between assessors when using standard descriptors to arrive at grades. For example, assessors could differ in their interpretation of performance criteria, appropriate sources of evidence and the key characteristics of trainees at various levels of performance. This raises potential issues concerning validity, accuracy and reliability of assessment decisions that have been the subject of the earlier practitioner research conducted at our partnerships.

The participants in this Q-method study conducted two formal formative progress reviews during school experience placements before making a summative assessment at the end of the ITE programme. All required the numerical grading of individual standards and overall teaching. Another potential source of variation in assessment data would be individual differences between assessors in their use of grades for formative and summative assessment. The importance of making accurate summative assessments for reporting is laid out in the inspection frameworks (OfSTED, 2018). However, experienced mentors may choose to use formative review number grades differently to motivate as well as inform trainees. Matthews and Noyes (2016) discussed the balance between formative and summative assessment during the observation of further education trainee teachers and the issues associated with the use of grades for feedback. Whilst advocating increased use of

trainee self-assessment, they noted that trainees receiving developmental feedback as a grade were sometimes confused about what it actually meant.

Reolofs and Sanders (2007) provide an example of support for the use of performance criteria of the sort associated with competency based teacher assessment In England. In a thorough attempt to provide a framework for the assessment of teaching performance aimed primarily at Dutch teacher educators, Reolofs and Sanders (2007) maintained that applying a reductionist model was more likely than other approaches to result in valid inferences about teaching competence. They emphasised that this allows the assessor to focus on different areas of teaching separately when reaching decisions. However, the adoption of competency-based assessment of teacher performance in England stimulated some academic debate concerning its validity and reliability for assessing the performance of trainees on ITE programmes. Turner-Bisset (1999) raised early concerns over the use of reductionist standards and descriptors and considered these inadequate for providing a model of the subject knowledge demonstrated by teachers. Proposing an alternative model, Turner-Bisset (1999) emphasised that teacher self-knowledge was an important element missing from the descriptors at that time. Again anecdotally, it has been our experience as teacher educators that trainees and mentors have found the performance criteria encompassed in The Teachers' Standards (Department for Education, 2011) a useful focus for the formative analysis of teaching skills and for guiding mentoring and coaching feedback. Leshem and Bar-Hama (2008) reported that their students expressed similar sentiments during a study of the introduction of competency-based assessments to an ITE programme in Israel. However, their students also preferred assessors to use holistic judgements when making summative assessments.

Although not directly referring to English ITE provision, Korthagen (2017) has more recently proposed a model of professional teacher development, which gives equal emphasis to the teacher as a person as well as their practice and understanding of theory. In this model, excellent teachers express appropriate core beliefs through the application of competencies to make effective decisions about their behaviour that maximises outcomes for their learners. In England, the application of 'lesson study' as a strategy for encouraging professional development has led Cajkler and Wood (2016) to call for a more educationally literate use of the Teachers' Standards when considering teaching competency. Initially referring to 'communities of practice' (Wenger, 1998) as a model that gives place to teachers within their practitioner group, they advocate the use of 'lesson study', involving a system of collaborative planning, observation and evaluation by groups of teachers, to encourage pedagogic literacy in teachers. They perceived this as an improved measure of teacher worth compared to mastery of a list of stated performance criteria.

How this Q-method study informs our previous investigations
For teacher educators with responsibility for the quality assurance of ITE provision, a consideration of the validity, reliability and accuracy of assessment tools is important and must address the possibility of subjective differences between assessors. An initial study by Tynan and Mallaburn (2017) surveyed simple statistical tests of significance to identify a method of demonstrating and monitoring consistency in assessment outcomes between five ITE programmes at a large HE provider. They concluded that, for the 2014-15 cohorts, there was consistency in the summative overall teaching grades across all programmes not demonstrated at three other formative assessment points. Within the programmes, numerical grades for individual standards were significantly correlated to the overall teaching grades ascribed (Tynan and Mallaburn, 2017). This was the first part of a mixed methods investigation into assessment grading outcomes and practices within and across partnerships at a HE QTS provider in the North West of England. This Q-methodology study comprises the second part of the study.

Tynan and Mallaburn (2017) speculated that their findings were congruent with school-based assessors consistently applying agreed partnership assessment practices. The use of common partnership templates, documentation and guidelines promoted at mentor training was intended to encourage this. Further, they identified one aspect of the partnerships' quality assurance provision present only in the final summative assessment of trainees as potentially important in explaining the findings. The school's visiting tutor from the HE QTS provider always attended the final evidence triangulation meeting between the mentor and trainee. This was to quality assure the process of deciding final grades but could have influenced the grades awarded. The visiting tutor was not present when grades were ascribed for any formative assessments and these did not demonstrate consistency in overall teaching grades across partnerships in different programmes (Tynan and Mallaburn, 2017). The authors accepted that these speculations on the observed patterns in the assessment data needed to be informed by the second qualitative phase of the study in order to investigate the degree and nature of subjectivity about assessment and grading amongst assessors.

Tummons (2010) argued that complex assessment processes at an HE QTS provider in the North East of England might be influenced by issues associated with the management of assessment and its quality assurance. Tummons (2010), associated this with institutional ethnography (IE) and actor network theory (ANT). An educational institution's IE consists of the sum of its documentation including policies, written information, guidance and support materials, templates and forms. It is the task of the institution's representatives or actor network of trainers, quality assurers, leaders and communicators to translate these into practice when working with ITE participants. How far these might influence assessment outcomes would depend on the content of the documentation and the extent and influence of the actor network (Tummons, 2010).

This is congruent with the findings of Tynan and Mallaburn (2107). One possible explanation for the observed variation in assessment outcomes across ITE programmes could have been a more relaxed implementation (ANT) of partnership guidelines (IE) at formative review points. This might be explained by mentors responding to individual training needs using variety of subjective approaches to the award of number grades in order to motivate their mentees during the earlier stages of training. This would be additional to assessors' subjective interpretations of guidelines and standards. However, consistency in summative gradings might be explained by the presence of an external quality assurance observer from the provider at final grading meetings (ANT) ensuring adherence to agreed guidelines and compliance issues (IE).

A further statistical study at the provider (Tynan and Jones, 2018) was able to demonstrate some subjective differences in summative number grading at the HE provider. For two cohorts on a Secondary ITE programme, assessors in core subjects differed in the way numerical grades were ascribed for overall teaching and Standards 3 and 4. These standards have headings associated with aspects of teacher subject knowledge (Department for Education, 2011). There was still an overriding trend towards consistency in the summative grading data but mentors in science and mathematics sometimes associated the standard linked to pedagogical knowledge with overall teaching to ascribe similar grades. This was in preference to associating overall teaching with the standard linked with subject content and curriculum knowledge. Mentors in English appeared to place equal emphasis on both standards. Later work (Tynan and Jones, 2019) on assessment data from the same secondary programme used effect size metrics. This suggested that, on one ITE programme, differences in the associations between grades awarded for individual standards and overall teaching could have been influenced over a three-year period by information communicated to all assessors. This had suggested that OfSTED inspection teams might expect grades for key indicator standards to be more closely linked to grades for overall teaching than others. The patterns in the assessment data were consistent with the information disseminated over the period of the study (Tynan and Jones, 2019).

Cumulatively, the patterns in the data from the three previous quantitative studies suggest that partnerships at the HE provider have developed clear guidelines and documentation (IE) implemented by influential advocates (ANT) similar to those observed by Tummons (2016). This Q-method study aims to demonstrate any subjectivity about the assessment between ITE participants at one partnership and describe any similarities and differences. This may give clues to the validity, reliability and accuracy of assessments and whether IE and ANT could be important influences on assessment outcomes or not.

Methodology and methods
Overview
This paper is the second phase of a mixed methods research project which resulted in an initial report on quantitative data by Tynan and Mallaburn (2017). It utilises Q-methodology (Brown, 1980; van Exel and de Graaf, 2005) as a quantitative approach to qualitative research in order to answer the research question: Do participants in ITE at a partnership in the North West of England demonstrate subjectivity in their perspectives on the assessment of ITE students on school experience placement? If found, the second aim of the research was to describe the profiles of subjective differences between groups of participants indicated by the Q-Methodology factor analysis.

Q-Methodology (Watts and Stenner, 2012) investigates participants' subjective beliefs or "first person viewpoints" (p.4) "in pursuit of an explanation and new insight" (p. 39). In this study it was used to focus on the responses of a small group actively engaged in ITE on a secondary postgraduate ITE programme working in partnership with a HE provider in the North West of England. Participants placed a concourse of relevant statements about the assessment of ITE students on school placement in order according to their level of agreement and disagreement. Q-methodology groups participants with sufficiently similar patterns of responses, identifying clusters of subjectivity amongst respondents. Qualitatively, it is then possible to identify the profile of statements that characterised the clusters and construct a description of their attitudes at the time of the exercise (Brown, 1980; van Exel and de Graaf, 2005). A group of mentors interested in assessment of teaching competencies trialled the use of the statements and Q-sort software at a partnership conference session. The participants in the trial neither added nor removed any statements. However, following feedback, some statements were re-worded to make the participants' choice clearer and the online sorting process was also changed to make it easier to complete.

Participants for the study were invited from attendees at a mentor training session at a partner secondary school academy. Eight respondents volunteered and represented a range of ITE partnership roles. This captured the perspectives of trainees and subject mentors directly involved in the assessment and grading process and a HE school liaison tutor with a quality assurance role. This was a convenience sample (Etikan *et al*, 2016) of mentors and trainees available to the lead researcher from a wider population of ITE participants at the HE provider. This non-random sampling technique could result in over representation of more numerous categories of participant. However, this is appropriate for the study as inferences are made only about subjective differences in the responses of the actual participants with no reference to any larger population (Etikan *et al*, 2016) involved in ITE.

Participants sorted the statements by importance when assessing trainees using a forced-choice frequency distribution along a continuum from 'most agree' to 'most disagree'. Participants positioned each statement on one of seven levels of priority. Each rank was assigned a score ranging from 3 for 'most agree' to -3 for 'least agree'. The middle rank scored zero. All the statements chosen represented approaches to the assessment of trainee teachers in England in general use by participants in the study, so it was possible that a participant might not actually disagree with any of

the statements. If this occurred, participants were asked to substitute 'most disagree' with 'least agree' in their minds.

QSortWare (Pruneddu, 2014), an online Q-Sort survey tool, was used to record responses. The analysis of data was conducted using the PQ Method software (Schmolck, 2014). It is worth noting that Q-methodology uses factor analysis, which is more usually associated with R-methodology (Brown, 1980; van Exel and de Graaf, 2005). R-methodology looks for correlations between variables linked to participants- for example the sets of treatment conditions that correlated to quicker patient recovery times. Q-methodology applies factor analysis to find groups of participants who share the same profiles for a set of variables- for example clusters of people placing similar priorities on statements about a particular subject.

The concourse of statements
The concourse consisted of 41 statements relevant to the assessment of trainees' teaching competencies whilst on school experience placement. It was developed from relevant literature, the Teachers' Standards (Department for Education, 2011), local tracking document descriptors and agreed partnership assessment practices. These were encapsulated in the forms, guidance and institutional documents in use at the time. The statements were designed to represent a wide range of, often alternative, approaches to the assessment of ITE trainees intended to elicit different responses in participants (Brown, 1980; van Exel and de Graaf, 2005). Although presented in a randomised order to respondents, the concourse statements were categorised (Tables 1-8) to facilitate the interpretation of findings.

The statements in Category A (Table 1) were linked directly to the Teachers' Standards (Department for Education, 2011) titles in Parts 1 and 2 and easily recognised by all the respondents. This allowed participants to demonstrate differences in the importance ascribed to individual standards for assessment and grading.

Table 1. Statements in Category A: Individual Teachers' Standards.

Statement	A. Individual Teachers' Standards
1	The trainee sets high expectations which inspire, motivate and challenge pupils
23	The trainee demonstrates good subject and curriculum knowledge
24	The trainee promotes good progress and outcomes for pupils
18	When considering grades, the trainee's ability to respond positively to constructive advice is important.
29	The trainee manages behaviour effectively to ensure a good and safe learning environment
31	The trainee makes accurate and productive use of assessment
33	The trainee's personal and professional conduct
34	The trainee plans and teaches well-structured lessons
36	The trainee adapts teaching to respond to the strengths and needs of all pupils
39	The trainee fulfils wider professional responsibilities

Category B statements (Table 2) were taken from partnership documentation and allowed participants to prioritise statements about the impact of quality assurance measures by internal and external representatives of the partnership.

Table 2. Statements in Category B: Quality assurance processes.

Statement	B. Quality assurance processes
5	The structure of the Triangulation meeting leads to accurate grading decisions on the final Phase Review Form.
8	The Professional Mentor's role in Quality Assurance is important for the accuracy of grading.
10	The Liaison Tutor's role in Quality Assurance is important for the accuracy of grading.
20	The presence of a university Liaison Tutor in the Triangulation meeting leads to accurate grading decisions.
22	A Triangulation meeting of a minimum of 1 hour in length, leads to accurate grading decisions on the final Phase Review Form.
32	Paired observations with university Liaison Tutors help with grading decisions on Phase Review Forms.
37	Independent professionals' role in Quality Assurance is important for the accuracy of grading - such as external examiners or internal moderators.

Category C statements (Table 3) were taken from the national and local guidelines. They allowed differences in opinion to be expressed about analytical and holistic approaches to assessment, reliance on descriptors for assessing competencies and the role of grading individual standards in the process of assessing overall teaching grades.

Table 3. Statements in Category C. Compliance and following local guidelines.

Statement	C. Compliance and following local guidelines
13	When considering a trainee's overall grade, it is important to use professional judgement holistically.
15	Best fit assessments are more accurate than can-do lists.
17	Holistic assessments are more accurate than those arrived at through reference to descriptors.
16	When considering individual grades, judgements should be made against the Teachers' Standards as the baseline for the minimum performance.
19	When considering individual grades, judgements should be made against criteria, such as the North West Consortia Trainee Tracking document.
40	When reaching a judgement about a trainee's overall grade, it is important to assess individual standards first.

Category D statements (Table 4) allowed participants to express differences in their attitude towards the allowances that could be made for a trainee's stage in training during assessment.

Table 4. Statements in Category D. Differences in assessment priorities by training phase.

Statement	D. Differences in assessment priorities by training phase
2	When reaching a judgement about a trainee's overall grade, some standards are more important than others in the early to middle stages of training.
25	When reaching a judgement about a trainee's overall grade all standards are equally important in the early to middle stages of training.
28	When reaching a judgement about a trainee's overall grade, all standards are equally important in the final stages of their training.
35	When considering grades, allowance should be made for how far the trainee is into their training.
41	When reaching a judgement about a trainee's overall grade, some standards are more important than others in the final stages of their training.

Category E statements (Table 5) enabled respondents to prioritise the importance of some methods and sources of mentor support when formatively and summatively assessing their mentees.

Table 5. Statements in Category E. Support for mentors and its source.

Statement	E. Support for mentors and its source
3	Mentor training leads to accurate assessment and feedback for trainees.
4	Paired observations with other practitioners help with accurate assessment and feedback to trainees.
21	Paired observations with other practitioners help with grading decisions on Phase Review Forms.
27	Paired observations with university Liaison Tutors help with accurate assessment and feedback to trainees.
38	Mentor training leads to accurate grading decisions on Phase Review Forms.

Category F statements (Table 6) describe guidance from OfSTED sources concerning the assessment of teaching competencies. Category G statements (Table 7) gave the respondents an opportunity to prioritise partnership statements linking the final triangulation meeting to the celebration of mentoring and partnership. Category H statements (Table 8) enabled participants to agree or disagree with the importance of evidence from sources external to an assessor's own mentoring situation.

Table 6. Statements in Category F: OfSTED descriptors.

Statement	F. OfSTED descriptors
11	When considering grades, the trainee's ability to work independent to the mentor is important.
12	When considering grades, impact on learning in the lessons is important.
14	When considering grades, the trainee's ability to evaluate their own progress is important.
30	When considering grades, consistency over a period of time is important.

Table 7. Statements in Category G: Celebration.

Statement	G. Celebration
7	The Triangulation meeting is an important celebration of the success of the mentoring.
26	The Triangulation meeting is an important celebration of the success of the partnership between the school and university.

Table 8. Statements in Category H: Sources of evidence for grades.

Statement	H. Sources of evidence for grades
6	Evidence from the activities undertaken at the university is important when grading the trainee in the final stages of their placement.
9	Evidence from the Alternative Placement is important when grading the trainee in the final stages of their placement.

Findings

Factor analysis of the fixed choice forced distribution data from the respondents using the PQ Method software (Schmolck, 2014) identified two discrete profiles of responses from participants that fell into two statistically significant groupings (Table 9). Q-Methodology treats these as distinct clusters of subjectivity about respondents' level of agreement with the concourse of statements. These are called 'Factors' by the software. In Table 9, those participants marked with an 'X' in the factor columns shared similar profiles of responses that are not likely to be due to random variations in the data. The figures in those columns are a measure of correlation between participants' responses generated by the software. In Table 9, a 1 would indicate a complete match and -1 complete disagreement. A zero would indicate responses varying independently with respect to other respondents.

For the rest of this paper Factors will be referred to as Groups to facilitate reading. Group 1 participants 1, 2, 3 and 5 demonstrated a similar profile of responses and Group 2 participants 4 and 7 a different and distinct profile. The responses of two further participants neither matched either of the profiles identified nor formed a separate distinct profile of their own. This demonstrated clearly a range of subjective differences in participants' perceptions about the concourse of statements concerning the assessment of trainee teachers. Table 9 also records some characteristics of the respondents.

Table 9. Participant characteristics and factor analysis: clusters of participants sharing a profile of responses are marked with an X.

Participant	Role	Gender	Subject	Factor 1	Factor 2
1	Mentor	Male	History	0.4961X	-0.2648
2	Trainee	Male	History	0.6344X	-0.0312
3	Trainee	Female	Art and Design	0.8034X	-0.0082
4	Mentor	Female	Mathematics	0.4809	-0.07107X
5	Trainee	Male	Mathematics	0.5156X	-0.3651
6	Trainee	Female	Design and Technology	0.0395	0.1766
7	Trainee	Female	Mathematics	-0.0416	0.5115X
8	Liaison Tutor	Male	Science	0.2052	0.2268

Table 10 shows the Q sort and Z scores within each group identified by factor analysis in Table 9, for individual concourse statements. These are the average rank scores and standard deviation within the group for each statement. This identified the statements for which Group 1 and Group 2 demonstrated consensus and differences of opinion. The categorisation of statements as 'consensus' or 'distinguishing' for groups is a function of the PQ Method software (Schmolk, 2014) based upon the Q sort score and degree of shared variance between the factors for statements. For instance, a very strong difference in opinion between groups about the priority placed upon a statement would be demonstrated by a large difference in Q sort score (average ranking) and low Z scores (narrow range of ranking scores). Conversely, statements with little or no difference in Q sort score and large Z scores would be ascribed as consensus statements.

Table 10. Consensus and distinguishing statements.

Concourse statement	Factor 1		Factor 2		Relevance to factors	
	QSORT	Z-Score	QSORT	Z-Score	Concensus	Distinguishing
1	-3	-2.06	3	1.51		✓
2	0	0.25	-1	-0.64		✓
3	1	0.36	0	0.21	✓	
4	0	0.33	3	1.52		✓
5	1	0.69	-2	-1.08		✓
6	3	1.43	-3	-1.33		✓
7	1	0.55	-1	-0.23	✓	
8	-1	-0.43	-1	-0.23	✓	
9	1	0.61	-1	-0.23		✓
10	1	0.42	-3	-1.31		✓
11	-1	-0.32	2	1.11		✓
12	-3	-1.65	3	1.33		✓
13	1	0.48	2	1.31		✓
14	-1	-0.52	2	1.31		✓
15	-1	-0.07	-1	-0.44	✓	
16	-1	-0.25	0	-0.21	✓	
17	3	1.45	0	0.21		✓
18	0	-0.07	0	0.02	✓	
19	0	0.21	1	0.44	✓	
20	2	1.02	-2	-1.08		✓
21	2	0.86	1	0.66	✓	
22	3	1.39	-1	-0.46		✓
23	-3	-1.63	1	0.44		✓
24	-2	-1.47	3	1.97		✓
25	2	0.79	-1	-0.9		✓
26	0	0.02	-2	-1.1		✓
27	0	0.17	-3	-1.75		✓
28	-1	-0.39	-3	-1.97		✓
29	-2	-1.17	0	0.02		✓
30	-2	-0.75	-3	-1.31	✓	
31	-2	-1.17	2	1.11		✓
32	2	1.17	-2	-1.1		✓
33	-2	-1.18	0	0.02		✓
34	-3	-1.66	1	0.43		✓
35	1	0.72	1	0.23	✓	
36	-2	-1.46	2	1.1		✓
37	2	1.23	-2	-1.1		✓
38	0	0.3	-1	-0.44	✓	
39	-1	-0.73	1	0.43		✓
40	2	0.91	1	0.64	✓	
41	3	1.59	2	0.88	✓	

Interpretation of findings
Reference to Table 9 indicates that factor analysis identified four respondents in Group 1: a male history mentor and his male mentee, a female art and design trainee and a male mathematics trainee. It also shows a distinct pairing (Group 2) consisting of a female mathematics mentor and her female mentee. A male liaison tutor and a female design and technology trainee were not associated with either profiles. Both groupings contained both trainees and subject mentors and, as such, subjective differences could not be associated with their different roles in the assessment process. It is worth noting that two trainees had independently arrived at similar profiles of responses to their own mentors during the sorting exercise. Using Tables 1-8, 9 and 10 it is possible to describe the profiles of responses associated with Group 1 and Group 2:

Consensus statements
Both Group 1 and Group 2 demonstrated a consensus concerning thirteen out of forty-one statements across most of the statement categories.

They were most likely to have similar levels of agreement about statements in the categories:

- C. Compliance and following local guidelines.
- E. Support for mentors and its source.

They most closely agreed that some standards were more important for assessment in the final stages of training than others. Both valued paired observations with peers when grading and favoured grading individual standards before the overall teaching grade. Neither placed a high priority on consistency in a trainee's performance over time when grading, considered an OfSTED performance indicator for higher grades.

Both Group 1 and Group 2 also tended to agree that grading should take into account a trainee's stage of training and that neither the professional mentor's quality assurance role nor a best-fit approach necessarily helped make grading more accurate. Other middle ranking consensus statements supported or were more neutral towards the use of the criteria contained in a trainee tracking document and the Teachers' Standards (Department for Education, 2011) and the role of mentor training in assessment for feedback and grading. Similarly, both groups assigned middle ranks to statements concerning taking account of a trainees' ability to respond to advice when considering grades and the triangulation meeting constituting a celebration of successful mentoring.

The consensus statements, for Group 1 and Group 2 identified by factor analysis show a keen awareness of compliance and guidance issues around the assessment and grading of ITE trainees with some qualifications. It also suggests supportive awareness of the trainee's perspective on potential issues with the grading process.

Group 1's Distinguishing Statements Profile: Making it fair
Group 1 consisted of a male history mentor and his male trainee, a female art and design trainee and a male mathematics trainee. The respondents in Group 1 agreed more than Group 2 with most of the statements in the categories:

- D. Differences in assessment priorities by training phase,
- B. Quality assurance processes
- H. Sources of evidence for grades.

15

Their shared profile emphasises their identification with mechanisms for ensuring fairness in the assessment process.

When thinking about assessing trainee teachers on school experience placement Group 1 most strongly prioritised the importance of a triangulation meeting, of at least an hour, to ensure the accuracy of final grades. They favoured the inclusion of evidence from university training days when deciding final grades and considered holistic assessments more accurate than those arrived through reference to descriptors.

Middle ranked statements were concerned with the emphasis given to standards at different stages in training, the structure of the final assessment triangulation meeting and the importance of the range of measures put into place to quality assure assessment process. They supported paired observations with university liaison tutors, the inclusion of evidence from the Alternate Placement in final grading decisions and the celebration of partnership during the triangulation meeting.

They gave less priority to several statements linked to specific standards or aspects OfSTED guidance: setting high expectations, which inspire, motivate and challenge pupils; demonstrate good subject knowledge; plan and teach well-structured lessons and demonstrate an impact on learning.

Group 2's Distinguishing Statements Profile: Applying the National Criteria
Group 2 consisted of a female mathematics mentor and a female mathematics trainee. The respondents agreed more than their counterparts in Group 1 with most of the statements in the categories:

- A. Individual Teachers' Standards
- F. OfSTED descriptors

Their shared profile emphasises the application of national criteria for assessment.
When thinking about assessing trainee teachers on school experience placement Group 2 most strongly prioritised having high expectations for learners that inspire, motivate and challenge, taking responsibility for all learners' progress and outcomes, and the impact of trainees' teaching on learning. Group 2 together with Group 1 prioritised paired observations with peers for improving accurate grading. However, Group 2 also prioritised paired assessment with peers for general assessment and feedback purposes.

Middle ranked statements prioritised all but one of the eight areas for the Teachers' Standards and several parameters set by OfSTED for assessing trainee teachers. Group 2 gave low priority to statements that the university Liaison Tutor's paired observations helped with the accuracy of grading or formative assessment for feedback. Group 2 least strongly valued ideas that all standards were equally important in the final stages of training or that considering evidence taken from university training days was important when grading in the final stages of training.

Summary of main findings
Q-methodology identified one group, one pair and two individuals with subjective differences in their levels of agreement with statements about the assessment of trainees' teaching competencies. Different profiles did not appear to be linked to a respondent's role in the ITE partnership. Both profiles contained a mentor and their own mentee and this is consistent with the suggestion that the mentoring relationship had encouraged similar perceptions of assessment and grading.

The participants demonstrated consensus over statements concerning compliance and guidance. They supported taking into account a trainee's stage of training during assessment but supported less the importance of consistency over time when grading. One group demonstrated more concern about the use of evidence and supported external checks to the mentor-mentee assessment arrangement. The other profile of responses demonstrated a lack of confidence in quality assurance processes and a preference for the independent application of criteria and guidance by practitioners in school. These respondents supported the use of the Teachers' Standards (Department for Education, 2011) for assessment.

Discussion
No extrapolation to a larger population nor general utility is claimed for the similarities and differences in the levels of agreement concerning statements about assessment that originated from this small sample of respondents. However, they demonstrate the subjectivity in thinking in a small group involved in an ITE partnership at a secondary school about the assessment of teaching at one point in time. These reflect potential tensions inherent in current ITE provision and assessment and have implications for the maintenance of successful partnerships.

Assessment and The Teachers' Standards (Department for Education, 2011)
Group 2 consistently assigned more importance to statements about individual standards descriptors than Group1. This may indicate different levels of confidence in the descriptors as valid measures of teaching competency. However, there was less separation concerning the use of The Teachers' Standards (2011) and criteria contained in the partnership individual tracking document. Both Group 1 and Group 2 agreed with the process of grading individual standards before arriving at an overall grade for teaching and on the usefulness of the tracking document criteria. Both agreed that best-fit approaches could be inaccurate in some contexts.

Agreement in these areas might suggest support amongst the participants for the views of Reolofs and Sanders (2007) on the value of reductionist models for improving the validity of measurements of competence and the value of grades for feedback in ITE (Matthews and Noyes, 2016). This might be stronger in Group 2 given their profile of distinguishing statements consistently supporting the importance of individual standards for assessment. However, Group 1 made a distinction between the best-fit application of standards and descriptors and fully holistic judgements reflecting the preference for holistic summative assessment expressed by the students in Leshem and Bar-Hama's (2008) study. Group 1 might be receptive to sort of the change in assessment emphasis towards pedagogic literacy advocated by Cajkler and Wood (2016) or Korthagen's (2017) ideas concerning integrating the teacher as a person into ITE professional development programmes.

Assessment and ITE Partnerships
Robinson's (2006) described the oscillations between school and HE provider situated ITE provision in England. These preceded and developed into the current diversity of routes into teaching and QTS providers based upon school-led partnerships. The respondents in the two groups identified by the Q-Sort analysis shared a consensus concerning statements associated with compliance with requirements for QTS and following the agreed partnership assessment guidelines. However, the groups expressed different levels of acceptance of ITE practitioners other than teachers in helping ensure consistency and accuracy of number graded assessments. Group 1 was more willing to accept the utility of evidence from the HE provider and alternative placement, and accept feedback on the accuracy of assessments from external ITE practitioners. Group 2 were very confident in their independent use of standard based criteria.

This may indicate different levels of confidence amongst respondents in assessors' ability to deliver fair and valid assessments without controls and checks. It is interesting that both Group 1 and Group 2 contained a mentor and their trainee suggesting that, in these cases, differences in perspectives were not associated with a respondent's role in the ITE partnership. This was more consistent with a social view of professional learning such as Wenger (1998) describes. However, the consensus and differences recorded still suggest an overall acceptance of the established system of assessment and machinery in place for its implementation. This is congruent with ideas about IE and ANT applied to assessment in Higher Education by Tummons (2010) and supported by Tynan and Jones (2019).

Assessment and Subjectivity

Martin and Cloke's (2000) application of Hager and Butler's (1996) process model of professional learning and its assessment to ITE proposes a simple progression for professional learning involving two assessment models. It assumes teachers have to acquire the knowledge necessary to teach prior to and during training. They then practice teaching in a protected environment before finally demonstrating their competence in real life situations. Unlike the assessment of knowledge, which can be examined, the assessment of competence was seen as judgemental and based largely upon qualitative evidence. This predicts the potential for a range of subjective differences between assessors based upon their interpretation of performance criteria and the selection and interpretation of evidence used to judge competence. This might be more evident during formative assessments when the purpose is developmental rather than to report summatively to external stakeholders. Tynan and Jones (2018; 2019) discussed some sources of subjective variability inherent in the application of number grades during summative assessments of trainees at an HE provider.

Previous quantitative findings (Tynan and Mallaburn, 2017; Tynan and Jones, 2018) and this Q-method study indicate that it is possible to demonstrate subjective differences between assessors particularly for formative assessment. However, this subjectivity was not easily demonstrated in the assessment outcomes at the HE provider in this study. Tynan and Mallaburn (2017) demonstrated consistency in summative grades for overall teaching across five ITE programmes at the same HE provider. They speculated that this was due to a number of measures adopted by the partnerships to improve consistency of assessment practice and outcomes across assessors. Tynan and Jones (2019) findings supported this speculation. Consistent patterns in summative assessment data indicated the possibility that information given to mentors during training had influenced grading decisions over a three-year period. This would be an illustration IE and ANT in action in an HE provider similar to that suggested by Tummons (2010). Like Tummons (2010) our findings give cause for concern that complex assessment issues could be masked by the management and quality assurance of the assessment process.

Conclusions

Subjectivity amongst a group of participants in ITE was clearly demonstrated by the Q-methodology. However, the distinguishing profiles identified by the factor analysis are located within a framework of consensus of about the importance of compliance and following agreed partnership guidelines. This is congruent with previous quantitative findings on consistency in assessment outcomes from ITE partnerships at the same HE QTS provider (Tynan and Mallaburn, 2017; Tynan and Jones, 2018; Tynan and Jones, 2019.) Consistency in assessment outcomes does not necessarily guarantee their accuracy nor validity. Further, when the degree of subjectivity observed is low whilst consistency in assessment data is high it suggests that compliance issues and quality assurance of the assessment process are influencing assessment decisions. The subjectivity profiles identified and described by Q-methodology support this possibility in this one instance. Acknowledging potential issues with numerical grading and that The Teachers' Standards (2011) only describe minimum performance criteria would suggest they are not an appropriate basis for numerical grading for partnership schools at the HE QTS provider.

It would seem more appropriate to use them for formative assessment and feedback during training and only to inform recommendations for QTS.

However, for those with responsibility for quality assuring partnerships, the discussion of accuracy and validity of assessments necessarily centres upon OfSTED inspection frameworks for ITE partnerships (OfSTED, 2018). OfSTED inspectors currently act as sole arbiters of the accuracy and validity of the assessments tools they observe and there is little opportunity for practitioners or quality assurers to influence the parameters for this. Inspection will ultimately govern the assessment of teachers in schools and the need for summative grades. The danger of potentially inaccurate or invalid assessment is that it fails to retain teachers who would be assets to the profession and does not identify teachers who need to address further professional development. However, authors such as Cajkler and Wood (2016) and Korthagen (2017) are active in proposing alternative approaches to professional learning and its assessment. These are interesting approaches to professional learning that could enhance ITE provision and include but extend further than the teacher competencies that are currently the only measures of teaching skill in England.

References

Brown, S. R. (1980) *Political Subjectivity: Applications of Q Methodology in Political Science*, Yale: Yale University.

Cajkler, W. and Wood, P. (2016) 'Lesson study and pedagogic literacy in initial teacher education: Challenging Reductive Models'*, British Journal of Educational Studies*, 64(4), pp 503-521.

Carter, A. (2015) *Carter review of initial teacher training (ITT)*, Crown copyright.

Czerniawski, G. (2018) *Teacher educators in the twenty-first century*, Critical Publishing Ltd.

Department for Education (2011) *The Teachers' Standards*, Crown copyright.

Department for Education (2016) *National Standards for school-based initial teacher training (ITT) mentors*, Crown copyright.

Door, V. (2014) *Developing creative and critical educational practitioners*, Critical Publishing Ltd.

Etikan, I., Musa, S. A. and Alkassim, R. S. (2016) 'Comparison of Convenience Sampling and Purposive Sampling', *American Journal of Theoretical and Applied Statistics*, 5(1), pp 1-4.

van Exel, N. Job A. and de Graaf, G. (2005) *Q methodology: A sneak preview*, Available at: www.jobvanexel.nl] or visit http://www.qmethodology.net (Accessed 23 November 2016).

Hager, P. and Butler, J. (1996) 'Two models of educational assessment', *Assessment and Evaluation in Higher Education*, 21(4), pp 367–378.

Jones, K. and White, E. (2014) *Developing outstanding practice in school-based teacher education*, Critical Publishing Ltd.

Jones, M. (2001) 'Mentors' Perceptions of Their Roles in School-based Teacher Training in England and Germany', *Journal of Education for Teaching*, 27(1), pp 75-94 .

Korthagen, F. (2017) 'Inconvenient truths about teacher learning: towards professional development 3.0', *Teachers and Teaching*, 23(4), pp 387-405.

Leshem, S. and Bar-Hama, R. (2008) 'Evaluating teaching practice', *ELTJournal,* 62(3), pp257-265.

Martin, S. and Cloke, C. (2000) 'Standards for the Award of Qualified Teacher Status: reflections on assessment implications', *Assessment and Evaluation in Higher Education*, 25(2), pp 183–190.

Matthews, R. and Noyes, A. (2016) 'To grade or not to grade: balancing formative and summative assessment in post-16 teacher trainee observations', *Journal of Further and Higher Education*, 40(2), pp 247-261.

Mentoring and Coaching CPD Capacity Building Project (2005) *National Framework for Mentoring and Coaching*, CUREE.

Moore, A. (2004) *The Good Teacher – Dominant Discourses in Teacher Education*, Routledge.

Ofsted (2018) *Initial teacher education inspection handbook*, Crown copyright.

Philpott, C. (2014) *Theories of professional learning: A critical guide for teacher educators*, Critical Publishing Ltd.

Pruneddu, A. (2014) *QSortWare*, [Online software].

Available at http://www.qsortouch.com (Accessed October 2014).

Reolofs, E. and Sanders, P. (2007) 'Towards a framework for assessing teacher competence', *European Journal of Vocational Training*, 40(1) pp1-17.

Robinson, S. (2006) 'Teacher Training in England and Wales: Past, Present and Future Perspectives', *Education Research and Perspectives*, 33 (2), pp 19-36.

Schmolck, P. (2014) *PQ Method (Version 2.35),* [Computer program].

Available at http://schmolck.userweb.mwn.de/qmethod/index.htm (Accessed October 2014).

Stiggins, S. (1999) 'Evaluating Classroom Assessment Training in Teacher Education Programs', *Education Measurement Issues and Practice*, Spring 1999.

Tummons, J. (2010) 'Institutional ethnography and actor network theory: a framework for researching the assessment of trainee teachers', *Ethnography and Education*, 5 (3), pp 345-357.

Turner-Bisset, R. (1999) 'The knowledge base of the expert teacher', *British Educational Research Journal*, 25(1), pp 39-55.

Tynan, R. and Jones, R. B. (2018) 'Assessing trainee secondary teachers on school placement: subject knowledge and overall teaching grades', *TEAN Journal*, 10(1), pp 20-34.

Tynan, R. and Jones, R.B. (2019) 'Can effect sizes give any clues to the way mentors ascribe numerical grades when assessing secondary trainee teachers against the Teachers' Standards in England?', *TEAN Journal*, 11(1) pp 4-14.

Tynan, R. and Mallaburn, A. (2017) 'Consistency counts – or does it?', *TEAN Journal*, 9(1), pp 90-99.

Watts, S. and Stenner, P. (2012) *Doing Q Methodological Research: theory, method and interpretation*, London: Sage Publications Ltd.

Welsh Government (2014) *Principles of mentoring and coaching*, Crown Copyright.

Wenger, E. (1998) *Communities of Practice: Learning, Meaning and Identity*, Cambridge University Press.

The journey from reflection on practice to research informed practice: the contribution of MA practitioner research

Andrea Pratt and Rick Tynan
Liverpool John Moores University

Abstract

This paper examines how one higher education institute (HEI) in the Northwest of England has devised and developed an innovative Master of Arts (MA) in Education practice with a curriculum designed to meet the needs of newly and recently qualified teachers. The research considers that in the early stages of initial teacher education (ITE), learning to teach may involve an apprenticeship model where the teacher educator must model learning, teaching and assessment strategies and inspire their mentees to find out those that will work for them through models of reflection and policy in practice. Later, successful teachers must be able to choose and critically evaluate strategies and pedagogies for themselves, a defining factor in the design of the programme and curriculum. Emerging from a course review, we consider evidence based practice and practitioner research through a progression route (MA) from ITE. Furthermore we determine that the research focus of early career teachers in this context enables research at a personal level for professional development of emerging and focused targets for context dependent research. This paper concludes that the current programme does indeed meet serving teachers' needs, and represents the move from reflection on practice to research informed practice evident through a consideration of the contemporary issues facing beginning teachers and their foci of study. This research also informs how we develop our recruitment strategy and next steps for the future to encourage a more sustained approach to research as teachers. It highlights clear steps for taking this research further and tracking beginning teachers' research journey over time.

Key words

Professional practice; impact; reflection; transformation; professional development; teacher researchers, early career, Masters, profession, curriculum design.

Background

This paper focuses on the introduction and innovation of a professional Masters award, the MA in Education Practice at an HEI (Higher Education Institute) in the North West of England (2017-18). The design of the course aimed to address the professional development needs of those entering the teaching profession, encompassing the theory of early research approaches, as newly or recently qualified teachers (NQT/RQT). The rationale for the programme built on the existing provision of a Postgraduate Diploma in Education (PGDE) with Qualified Teacher Status (QTS). In September 2014 the PGDE replaced the School's Postgraduate Certificate in Education (PGCE) programme as the main vehicle for its postgraduate Initial Teacher Education (ITE) provision although the PGCE remained through partnership with school consortia and the School Direct (getintoteaching.education.gov.uk) route in teacher education. The vision for the programme was for emerging teachers to contribute to the development of approaches to knowledge in the field of education practice and for the institution to support the early professional learning needs of NQTs and RQTs. The early career framework introduced in 2019 clearly states that *'early career teachers should be entitled to 'learn about' and learn 'how to do' based on expert guidance and the best available research evidence'* (DfE 2019:5).

Citation

Pratt, A. and Tynan, R. (2019) 'The journey from reflection on practice to research informed practice: the contribution of MA practitioner research', *TEAN journal, 11(4), pp. 35-46.*

Introduction

The aim of this research project was:

> To consider newly and recently qualified teachers' (NQTs and RQTs) research foci during one year of an Education Practice MA and early completion of Masters. It also considers their engagement with and performance of practitioner research through an innovative programme and curriculum design.

Currently those entering the teaching profession following initial teacher education programmes predominantly carry Masters level credits in the UK, principally a PGCE award with 60 credits. However this does exclude some school centred initial teacher education (SCITT) programmes (UCAS 2019) which have qualified teacher status (QTS) only route. The HEI undertaking this review in the north west of England has a successful record in developing and delivering such programmes, with large numbers of teachers graduating each year in both primary and secondary teacher education across the partnership (approximately 500 in 2018). In September 2014, the School of Education introduced a PGDE award with 120 credits at level seven (QAA) as the normal route for its postgraduate ITE programmes. It was determined that new entrants to the teaching profession for the first time in our institution had the potential for completion of a full MA award within their first years in the teaching profession. Demand for such an award was evident through student consultation prior to validation and consultation with school partners. Locally there was a positive demand from school partners to provide an award targeting the needs of NQT/RQTs. Indeed the original proposal came from school partners working in partnership and exploring the progression routes for beginning teachers. It was determined through a rationale for the programme that this HEI was well positioned to meet a growing need, having a base of experience and expertise gained from a successful record of traditionally delivering 60 credit PGCE programmes and involvement in the Masters in Teaching and Learning (MTL 2009). This initiative was introduced in 2009 in response to a call for 'all teachers to achieve a Masters qualification over the course of their career' the rationale being to bring parity with the 'highest performing education systems in the world' (DfCSF, 2008 4.2.2). Our institution also supported the Teach First initiative recruiting the highest calibre teachers into the profession to schools facing the biggest challenges (teachfirst.org.uk). A new programme design and implementation brought new challenges in working with NQT's in their first teaching posts. This review considers the research areas of early career teachers having completed two cycles of practitioner action research within their school context and presented as two research reports.

The journey from reflection on practice (graduate) to research informed practice (teacher) informed by curriculum design

The PGDE programme design, like many other teacher education programmes is underpinned fundamentally by reflective practice and is a practice that is well-established (Zeichner and Liston 1987). Reflection however is not to be presumed as an indicator of a good teacher (McLaughlin, 1999). Teacher education programmes can be criticised for superficial reflection and a lack of understanding of the foundation of reflective practice (Hebert 2015). Valli (2000) acknowledges the positive contribution universities are making in developing skills of enablement through critical reflection on practice in supporting teachers in applying structured reflective models. What is clear, is that the curriculum needed careful consideration and an approach that works alongside the demands of early career teaching and the constraints that still remain current as stated here by Ross (1987):

> The demands of first-year teaching and the structural constraints of most teaching contexts make it highly unlikely… that graduates or novice teachers, will conduct formal action research projects during their beginning years of teaching
>
> (Ross, 1987:147).

Before designing the curriculum, the vision needed to be clear; would trainees value M-Level study further beyond the PGDE year and value evidence based practice and practitioner research? Are both only useful on a personal level for professional development? In addition, would students make the move from structured reflection on practice to research in action as teachers driven by personal constructs?

The design of the MA Education practice programme was an opportunity therefore to shift the focus to early career teacher research approaches and evidence based practice through practitioner enquiry as a natural progression for early career teachers and researchers, thus building on and developing research methods acquired during the PGDE programme. Evidence-based research is highlighted as an area of significant research potential (Mujis and Reynolds (2017). Darling-Hammond (2008) however debates a quandary in teacher education through the assimilation of theory in institutions and the difficulties student teachers face in relating abstract concepts to their own classroom context. The challenge for the programme was to encourage co-construction and meaning to theory and research in practice. Murray and Passy (2014) stress that theoretical perspectives are required to develop practice. However, these are not always valued by student teachers early in their practice (Knight 2015). There was therefore an opportunity to support the natural progression of research skills and value, to maintain the momentum for student teachers working at Masters level and further develop a richness of practice through evidence-based practice of what matters to the teachers in the appropriate context, shifting the focus to teacher research. McLaughlin, Black-Hawkins and McKintyre (2004) describe the three purposes of teacher research founded in practice; the resolve from a personal, political and school improvement perspective are identified as key drivers. McLaughlin alludes further to the fact that these are entwined and are not, as may be assumed, without complications. This course design sought to embrace the move from a focus on directed assignment structured reflective practice to a more critical engagement and immersion in personal research driven wholly by a personal resolve and motivation that would surely encourage study alongside the demands of the teaching year (Ross 1987). Therefore moving the focus from student teachers to teachers needed to consider how to bridge the gap between the academic and teacher (Flores 2018) and to encourage transformation in practice.

The challenges were clear; our student teachers were now teachers and may still have a predisposition to reject educational research (Knight 2015). Cain (2015) asks us to consider the rationale for the research we present to teachers and this would be a foundation on which we needed to build through practitioner research and community of practice. The responsibility of our core team is reflected below and highlights an obligation to support and respond:

> As teacher educators, we need to be open to shifting our purposes and our designs in response to changing institutional, political, and practical circumstances. Doing so in principled ways and in dialogue with partners is central to responsible innovation
>
> (Jurow, Horn and Philip, 2019:94).

The approach above was exemplified through the consultation with all partners prior to validation and the changes and shifts in the political landscape and the inception of the programme to ensure a responsible approach. Korthagen (2017) is clear in stating that sometimes teacher research just does not work and therefore suggests there are still challenges in teacher research. The climate and support needs to be there and that is why the responsibility of the University is one that was not taken lightly and needed careful consideration.

Blended learning approach

The curriculum approach to learning for the programme is a blended approach (Graham 2013), a method predominantly found in the higher education context and a familiar model to most in this setting. The blending of face to face teaching and online materials and support were provided through a virtual learning environment (VLE) and thoughtfully approached (Garrison and Kannuka 2004). Teaching and module assessment were designed to meet the challenges and development of intellectual, analytical and research skills relevant to the needs of the emerging professional in education practice. With this in mind the VLE needed to support this through rich resources, and create a purposeful online community of practice. Three focused teaching days and associated materials were designed to expand the knowledge and understanding of students to further consider teaching and learning in education. Through learning based on critical engagement with current education theory, research, policy and practice, the focus was determined by the teachers' own early career targets and personal focus (McLaughlin, Black-Hawkins and McKintyre 2004).

The blended approach supports independent study to encourage beginning teachers to make an original contribution to research applied to teaching, and learning relevant to their professional learning needs and educational settings. The longitudinal aim was to enable students to participate in lifelong professional learning leading to further specialised study (PhD/EdD) or indeed continue with research practice throughout their career.

Module and Curriculum design

The main rationale for the programme builds on the foundation of practitioner inquiry through one 60 credit module with two component parts equivalent to a traditional Masters level dissertation. Timperley and Alton Lee (2008) recognise that practitioner inquiry is complex. However the shared goal is improvements in curriculum and teaching methods. Practitioner research/inquiry is part of the *'family of action research'* (Kemmis and McTaggart, 2005:560). Furlong and Oancea (2009) further define practitioner research as connected by academic activities and informing practice through research. Groundwater-Smith and Mockler (2006:107) state that '*those involved in practitioner inquiry are bound to engage with both theoretical and practical knowledge moving seamlessly between the two.'* Therefore the role of the University is to enable beginning teachers to navigate through a complex landscape balancing the day to day demands of early career teaching. Through the practitioner inquiry module the action research model is used to examine problems that the early career teachers identify and encourages further critical engagement with solving those problems (Darling-Hammond 2010) in depth.

Baumfield, Hall and Wall (2013) highlight the foundation for research emerging from questions from the classroom and experience does indeed find itself in-between reflection and action research. The move from reflective practitioner to research informed practice mirrors the progression from PGDE students to the education practice programme. Connected to and exemplified by the rationale of British Educational Research Association (BERA 2014) who clearly establish the importance of the role of universities to develop a more systematic approach to teachers' continued professional development this is reflected in the course design. BERA (2014:12) also recommended that '*every teacher should have the confidence, ability and capability to engage in research and enquiry'.* Aspirational in nature, the report further exemplifies the need for a 'self-improving education' system that is research informed and research inquisitive, and the need for cultural change to create a new paradigm that is 'research rich.' The DfE (2016) states that it is indeed time to embrace evidence based practice. Our shared goal is reflected by the Department for Education in the UK and we establish a foundation on which to build for our early career teachers:

> Effective teaching requires considerable knowledge and skill, which should be developed as teachers' careers progress

> (DfE, 2016:1).

Leat, Lofthouse and Reid (2014) identify that the path to research engagement is not always progressive and can be perceived negatively, and adversely linked to a target driven agenda, and concede that this can inhibit creativity. While tensions can exist, effective leadership of the process is key. However, BERA (2014) provides further affirmation of the 'innovation and change' being very 'powerful' as a result of successful commitment to research. The continuum of the programme as a bridge for the teacher researcher also ties in with the agenda of evidence informed teaching (DFE 2019) and is a clear benefit of the programme design. Biesta (2007) challenges the principle of 'effective practice' and highlights that just because something is effective, the moral practice should be more important than the initiative. Biesta (2007) therefore places here a value judgement and consideration for paying due diligence to the work. Also in this case thinking about the moral responsibility of improving and 'what is appropriate for these children in these circumstances' Sanderson (2003:340) forms a core value at the heart of the programme.

Beginning teachers as researchers
Acknowledging the demands on the wider teaching community to engage in research as teachers (BERA, 2014; Leat, Lofthouse and Reid, 2014; Goldacre, 2013 and DFE, 2019) is by no means a new concept (Campbell and Jaques 2004). The continuity design of our programme enables our students to situate their emerging research in a new context and this can be a complex and indeed 'new' environment. In doing so early in their career, beginning teachers can continue to build on the foundations of research practice and academic momentum from the PGDE while extending their knowledge of evidence based practice and practitioner enquiry.

The role of the university is crucial and recognised as significant in supporting the transition of teacher researchers (Ginns et al 2001), and how this can support professional development in the NQT/RQT year. Timperley (2008) highlights the benefits of those engaged in professional learning and emphasises the benefits of effective practice and observable implications when teachers can see the impact in the classroom. Expertise is also something that Timplerley also acknowledges is essential and the course does indeed allow us as a group to '*challenge existing assumptions and develop the kinds of new knowledge and skills associated with positive outcomes for students'* (Timperley 2008:20). Moving beyond a one year PGDE course also allows for an opportunity to consider how this can impact further on pupil outcomes however does require '*sound theoretical knowledge evidence informed inquiry skills, and supportive organisational conditions'* (Timperley 2008:24). Time to conduct research and reading is a key concern identified in many an introductory session on the Masters programme. Leat, Lofthouse and Reid (2014) further highlight that research can be experienced as:

> A burden, a conflicting agenda or a contradiction or at the very least an addition to overcrowded workloads. Given this, teachers need some surety in the quality of relationship
>
> (Leat, Lofthouse and Reid, 2014:4).

Transforming practice
Our course aim is not small or transitory, the goal is to sustain effective practice and embed a culture of enquiry beyond the formative years in early career teacher research. This aim may be aspirational in nature. However, it is something we intend to consider in the research and subsequent research beyond the remit of this paper. This goal is in line with Zeichner (2003) in considering school based research can make a difference but is dependent on thinking together, therefore allowing for meaningful discourse in a collegiate and supportive environment that questions and challenges existing practice for the benefit of the learners in the classroom. In asking beginning teachers what they perceive the teaching concerns are as they commence their career, we are, as an institution, mindful of their emotional development and their resilience and we consider 'their potential' (Korthagen 2017:399). A responsibility in the course design that we take seriously moves the focus

from ITE and what they need to know to be teachers. A focus on personal development is a definitive and perhaps obvious shift in the paradigm. We commence teacher education with a focus on what we think student teachers should know and determine the content of the course. The shift in what is important to the teacher becomes the priority and the research tools required become the emphasis of delivery.

It is clear in the conception of this programme that the skills embedded in early career teachers' research practice need to continue to develop beyond reflection to both engage and respond; to *'make sense of their practice and problems they face.'* (Darling, Hammond and Snyder 2000:540). This is a reassuring aspect of the programme and design, however the shift in the teaching and personal foci determined by what is important to them relies on balancing the blended learning approach to a community of practice and support network and it is this that will encourage transformation (Hargreaves 2003). Timperley and Parr (2009) further reinforce that to lead to improvement there must be the space to make errors, and supervision and personal contact will aim to counteract this. A focus on pupil learning has been established on the PGDE programme and the Education practice programme design and is a fundamental requisite for any innovation in teaching and learning and changes in school teaching cultures (Wall 2008).

Data gathering
Data was collected through a mixed methods approach, employing the combination of qualitative data to facilitate explanations and quantitative data. It was identified that the current cohort of students would help to substantiate research findings to give a context to the research and contribute to a balanced line of enquiry. The type of data collection is documentary evidence (Bowen 2009), programme leader course evaluation, analysis of the foci of the practitioner research conducted by students and detailed descriptions (Creswell 2009), and falls within the ethical considerations set out by BERA (2011) for anonymity required. Documentary research and analysis of the data enabled the research to consider the conception of information (Cohen, Manion and Morrison 2011). This piece of research analysed a range of documents produced for course evaluation and analysis excluding the evaluation of students and views and opinions. These documents included student foci and external examiner comments following scrutiny of the student research undertaken. The research projects of the 2017-18 cohort (n=38) were scrutinised to find out the attributes of their investigations and the research approaches employed. One investigation constitutes 3% of the sample in the findings. However, no pre-existing models or conceptual frameworks were applied to the qualitative data. As such, we used categories or codes and adopted a grounded theory approach (Saldana, 2015) to identify if a theoretical link is evident in the research. Ethical adherence was considered at all stages (BERA 2011).

Presentation of findings
Foci of research for NQTs and RQTs
The majority of studies attempted action research cycles but a few approached full case study proportions. The majority of interventions (58%) were tightly focused and very specific but there were a significant number of studies (42%) involving the implementation of strategies that were really combinations of discrete interventions. The majority of the investigations (68%) employed mixed methods for data gathering but some were solely quantitative (13%) or qualitative (18%) in their approach.

Investigations could have multiple categories associated with the research area. The 38 investigations were described using 210 categories grouped under 11 major themes (Table 1.).

Table 1. Major themes for research projects.

Themes	% of statements coded under each theme
Pedagogy	22
Psychology	20
Assessment	17
Inclusion	14
Classroom management	8
Methods	7
Literacy	4
Curriculum	3
Social Psychology	2
Cognition	2
Health	1

Table 1 indicates that the NQT and RQT research projects were most concerned with pedagogies and psychology in the classroom followed by assessment and inclusion issues.

General pedagogical approaches investigated included examples such as: intervention classes for new arrivals, problem based learning activities, scaffolding, self-regulation and small group interventions. A number of so called 'hot topics' were of interest to NQT and RQT practitioner researchers who sought to interrogate and validate these in their classrooms. Such areas included accelerated learning, context dependent memory cues, flipped learning, Kagan strategies, manipulatives, mastery, reciprocal teaching styles, repeated retrieval practice and the Singapore bar model.

Personal psychology and cognition studies together equalled pedagogies in the frequency they were the focus of studies (Table 1). Confidence, engagement, motivation, attitudes and perceptions of learners were the common psychological categories. Together with inclusion studies these are examples of the learner centred concerns that formed the basis of many research projects and demonstrated great engagement with whole learner issues. Inclusion projects raised important issues such as: disadvantaged learners and the use of pupil premium, EAL learners, gender issues for both boys and girls in the classroom, and closing attainment gaps. Investigations into underachieving and disaffected learners focused on both high and low attaining learners.

Nine studies used learners' attainment to evaluate the impact of interventions and sometimes linked this to other psychological measures such as engagement or motivation. Others attempted to use less easily defined measures e.g. progress, effectiveness, achievement and proficiency. In some projects, specific new assessment approaches were part of the intervention, e.g. a literature extract question in modern languages. Other themes included formative and summative assessment, ipsative assessment, grading as feedback, pre- and post-the testing and assessment of non-academic skills. Classroom management issues were also of concern in 16 investigations. Behaviour management themes included: behaviour for learning, disengaged learners, low-level disruption, and talking out of turn. Studies also looked at the role of teaching assistants in the classroom and models for their deployment. Others looked at teacher behaviours such as the consistent implementation of praise, support, high expectations and routines. A small number of projects (7) focused on curriculum

innovations and several projects examined reforms to the GCSE curriculum and assessment. Literacy was a focus for 8 studies.

Table 2 indicates the school subjects represented in the profile of research areas chosen, reflecting also the subject specialisms of the beginning teachers recruited to the programme. Core subjects comprise 50% of the foci for projects.

Table 2. Subject focus for the action research projects undertaken (n=38).

Subject	%
None specified	11
Art	3
Art and Design	3
English	16
Food Technology	8
Geography	5
Mathematics	18
Modern Languages	8
Music	3
Physical Education	11
Science	16

Course review data

External examiner comments stated that:

> The work shows that they [NQT/RQT's] have been equipped and encouraged, by the programme, to interrogate their practice, relate this to relevant, current models and discourse, and undertake focussed action research
>
> (Examiner A).

> The programme is very well-designed and obviously based on a deep understanding of the professional needs and circumstances of NQTs. It is well understood throughout the sector that this group of learners requires a high level of scaffolding and support and it is evident that the architecture of the programme successfully provides these
>
> (Examiner A).

In terms of student feedback the following quotes captured through one of the teaching sessions recognise that:

> The MAEP Programme will enable us as new practitioners to utilise the momentum of the PGDE course and further develop our personalised approach to tackle relevant classroom tensions for the benefit of our student progress
>
> (Student A).

> An excellent opportunity to continue professional development through academic research in your educational setting. Another advantage is the chance to discuss practice with peers
>
> (Student B).

Quantitative data responses:

The percentage of students recruited to the programme from the secondary sector is around 67% and from the primary sector 33%. The cohort completion and pass rate has been around 97% with cohorts

of between 40 and 60 since the start of the programme. The beginning teachers asked to consider their rationale for doing the course cited their professional development (75%) and personal development (83%) as a key aspect of their decision to participate in further studies after their PGDE or PGCE. When considering educational research, 92% of the respondents stated that the course had resulted in their valuing educational research more than when on the PGDE training year. 8% of the cohort stated that this had remained the same but that was because they had already valued educational research highly during the ITE programme. 83% of respondents stated the research they had undertaken had made a positive impact on their or others practice. The other 17% considered that it was too early to claim this but that they expected this to be the case in due course.

Discussion

Through the research presented in this paper, empirical evidence suggests that although it is an aspiration to create evidence based practice, in practice it just may not work (Korthagen 2017). However our rationale was about creating a course community that is supportive and functions well and a place where *'effective learning can take place'* (Korthagen 2017:399) and the role of the HEI is vital (Ginns et al 2001). Therefore our programme relies on the sharing of professional expertise and equality (Newman and Mowbray 2012) whilst maintaining support for the demands of the NQT through a blended learning approach (Graham 2006). With this in mind one of the perceived benefits of the programme in its inception was the quality of sustained relationships from the progression of the PGDE. Leat, Lofthouse and Reid (2014) distinguish the importance of having the support and space to engage in dialogue exploring a range of perspectives is vital. Students in this study value space provided in a network of peer support in their early years of teaching and the environment is a safe space to make mistakes and engage with colleagues (Timperley and Parr 2009).

What is clear through the course review (Examiner A) is there is an observed difference in research informed approaches provided by course design and this is evident in student outcomes of confident *'criticality and challenge of existing theory to inform practice'* (Programme leader) and the necessity to structure that support (Leat 2014). The focus of research and attitudes shifts in line with the research to become more valued as time goes on (Knight 2015) to less of a transactional and technical focus (reflective practice). The personal and context specific engagement with research demonstrates an immersive understanding of the research in school through action research and intervention (deliberate practice) and is motivated by personal goals and early career target setting.

Classroom management and focus on behaviour is identified as a significant area of research for the beginning teachers in this study (n=15). This concurs with research that classroom teacher instruction and pupil and teacher communication are a focus of continued teacher research (Beycioglu, Ozer, and Ugurlu, 2010). Considering the motivation to complete the course, the data supports that beginning teachers in this study do indeed have a personal motivation for completing the course, linking to the personal drive (McLaughlin, Black-Hawkins and McKintyre 2004) and clearly focus on a personal perspective, not driven by politics or the wider school improvement agenda. Beginning teachers in this study demonstrate their research is not linked to the target driven agenda (Leat 2006).

There are many challenges identified in the research of teacher research (Goldacre 2013; BERA 2014, DFE 2019). By curriculum design and thoughtful design (Garrison and Kanuka 2004), it is perceived that we can negate the issues in navigating teacher research informally (BERA 2014) to encourage structure and the climate for success. It is also acknowledged that this study has limitations in terms of exploring the challenges in beginning teachers and is a recommendation for further research, however is acknowledge that it may reflect the constraints identified by Leat, Lofthouse and Reid (2006).

Conclusion

This paper aimed to share how an innovative MA in Education practice was designed to meet the needs of newly and recently qualified teachers. It reported on trainee teachers' focus of study in line with their own personal educational research (McLaughlin, Black-Hawkins and McKintyre 2004) and motivation. The paper explored the transition from reflection on practice to conducting in-depth action research through practitioner inquiry. The course design encouraged exploration of teacher research and informed practice to affect and perhaps transform learning in the classroom designed to consider the demands of early career teaching. This paper concludes that through thoughtful curriculum design our early career teachers are more equipped to critically evaluate strategies and pedagogies for themselves and demonstrate and increased engagement and value of research and a choice to continue study motivated by both a personal and professional level linked to meaningful contexts in a community of practice.

Research within this paper explored the tensions and challenges in conducting teacher research, however demonstrated how the programme was able to harness opportunities to design an innovative programme to maximise potential for transforming practice. From the literature and acknowledgement of a small-scale context of early career teachers' foci, this paper acknowledges that the programme does indeed meet the needs of early career teachers. Furthermore it exemplifies that the programme successfully supports personal professional development, and makes a significant link to the effective practice of the fundamental skills (progression) learned on the PGDE programme and is an appropriate progression route. Finally the limitations of this study do highlight there is limited research on emerging and beginning teacher research and therefore we would like to explore the limitations of this study in a longitudinal study. This study would look to the impact of research in the classroom, the likelihood of teachers adopting research approaches throughout their career, and the value this holds through intrinsic motivation rather than a wider school improvement agenda (McLaughlin, Black-Hawkins and McKintyre 2004).

Challenges in teacher research are widely explored in existing research (Leat, Lofthouse and Reid 2014) so we make the recommendation for continued research beyond the initial reflection on the initial stages of the programme in order to consider the progression of early career teacher research and the position of teachers' research after a number of years in the profession. Through research beyond this initial study of 2017/18 we will consider the alumni of the course to gather further research data from the programme over time. The authors acknowledge this paper is a foundation of research that needs to be developed in more depth in terms of the impact in the classroom and tangible outcomes for pupils.

Recommendations and next steps
- It is the role of the HEI to consider closing the gap between primary and secondary beginning teachers and the recruitment of primary early career teachers and address the low male recruitment and increased engagement of foundation subjects in secondary education.
- The core team will explore in more depth the link to the course design and impact in practice and tangible outcomes for classroom practice. Through purposeful research of early career teachers experiences of conducting research in the classroom.
- The researcher will explore attitudes to research in more depth, the challenges and motivation to continue research practice and the challenges by widening the scope of participants over time following completion of the programme.

References
Baumfield, K., Hall, E., Wall, K. (2013) Action Research in Education: *Learning Through Practitioner Enquiry*. Sage Publications, London
BERA (2011) BERA Ethical Guidelines for Educational research
https://www.bera.ac.uk/publication/bera-ethical-guidelines-for-educational-research-2011

BERA (2014) *Research and the teaching profession* accessed through:
https://www.thersa.org/discover/publications-and-articles/reports/research-and-the-teaching-profession-building-the-capacity-for-a-self-improving-education-system

Beycioglu, K., Ozer, N., and Ugurlu, C,T., (2010) *Teaching and Teacher Education: An International Journal of Research and Studies*, v26 n4 p1088-1093

Biesta, G. 2007. "Why 'What Works' Won't Work: Evidence-based Practice and the Democratic Deficit in Educational Research." Educational Theory 57 (1): 1–22.

Bowen, G. A., 2009, 'Document Analysis as a Qualitative Research Method', *Qualitative Research Journal,* vol. 9, no. 2, pp. 27-40.

Cain, T. (2015) Teachers' engagement with published research: addressing the knowledge problem. *Curriculum Journal, 26*(3), 488-509.

Campbell, A., Jaques, K. (2004) Best Practice researched: teachers expectations of the impact of doing research in their classrooms and schools. *Teacher Development* 7:1

Cohen, L., Manion, L. and Morrison, K. (2011) *Research Methods in Education* 7th Edition. London: Routledge.

Creswell, J.W. (2009) *'Research Design'.* Sage Publications. London, pg. 148

Darling-Hammond, L., and Snyder, J. (2000). Authentic assessment of teaching in context. *Teacher and Teacher Education, 16*, 523-545

Darling-Hammond, L. (2008). Introduction: Teaching and learning for understanding. *Powerful Learning: What We Know About Teaching for Understanding* (pp. 1-9). San Francisco, CA: Jossey-Bass.

Darling-Hammond, L. (2010). Teacher Education and the American Future. *Journal of Teacher Education*, *61*(1–2), 35–47

DfE (Department for Education) (2016) Standard for teachers' professional development
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/537030/160712_-_PD_standard.pdf [Accessed 02.10.19]

DfE (Department for Education) (2019) Early Career Framework
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/773705/Early-Career_Framework.pdf [Accessed 02.10.19]

Flores, M, A. (2018) Linking teaching and research in initial teacher education: knowledge mobilisation and research-informed practice, *Journal of Education for Teaching*, 44:5, 621-636.

Furlong, J. and Oancea, A. (2005) A framework of quality for the assessment of applied and practice-based, Oxford University Department of Educational Studies.

Garrison, D.R., and Kanuka, H. (2004) Blended learning: Uncovering its transformative potential in higher education. Internet and Higher Education 7 (2004) 95–105

Ginns, I. Heirdsield, A. Atweh, B. and Watters, J. (2001) Beginning teachers becoming professionals through action research, Education Action Research 9 (1) 11-134

Goldacre, B. (2013) Building Evidence in Education. Department of Education.
https://www.gov.uk/government/news/building-evidence-into-education

Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), *Handbook of distance education*, (3rd ed., pp. 333–350). New York: Routledge.

Groundwater-Smith, S. and Mockler, N. (2006) Research that counts: practitioner research and the academy. In: *Counterpoints on the Quality and Impact of Educational Research, Special Edition of Review of Australian Research in Education,* Number 6, pp. 105-117.

Hargreaves, D, H. (2003) *Education epidemic: transforming secondary schools through innovation networks*

Hebert, C. (2015) Knowing and/or experiencing: a critical examination of the reflective models of John Dewey and Donald Schön. *Reflective Practice*, 2015 Vol. 16, No. 3, 361–371

Jurow, S. L.S. Horn, T, M. Philip (2019) Re-mediating knowledge infrastructures: a site for innovation in teacher education, Journal of Education for teaching 45:1, 82-96.

Kemmis, S. and McTaggart, R. (2005). Participatory action research: Communicative action and the public sphere. In N.K. Denzin and Y.S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed. 559-603). Thousand Oaks: Sage.

Knight, R. (2015) Postgraduate student teachers' developing conceptions of the place of theory in learning to teach: 'more important to me now than when I started', *Journal of Education for Teaching*, 41:2, 145-160

Korthagen (2017) Inconvenient truths about teacher learning: towards professional development 3.0, *Teachers and Teaching*, 23:4, 387-405.

Leat, D. (2006) Conditions for teachers' collaborative learning: The importance of learning space. Paper given at the *European Conference on Educational Research* Geneva, Switzerland. September

Leat, D., Lofthouse, R., and Reid, A. (2014). *Teachers' views: Perspectives on Research Engagement. Research and Teacher Education: The BERA-RSA Inquiry* London: British Educational Research Association.

McLaughlin, T. (1999). Beyond the reflective teacher. *Educational Philosophy and Theory,* 31, 9–25.

McLaughlin, C. Black-Hawkins,K., and McIntyre, D (2004) Researching Teachers, researching schools, researching networks: A summary of the literature. Cambridge, University of Cambridge.

Mujis, D, and Reynolds, D. (2017). *Effective Teaching: Evidence and Practice* (5th edition). London: SAGE Publications Ltd.

Murray, J., and R. Passy. (2014) "Primary Teacher Education in England: 40 Years on." Journal of Education for Teaching: *International Research and Pedagogy 40* (5): 492–5

Newman, L. and Mowbray, S. (2012) We were expected to be equal.' Teachers and academics sharing professional learning through practitioner inquiry. *Teachers and teaching:* Theory and practice, Vol. 18 (4), pp 455-468.

QAA (The Quality Assurance Agency for Higher Education) https://www.qaa.ac.uk/docs/qaa/quality-code/master's-degree-characteristics-statement.pdf

Ross, D. (1987) Action research for preservice teachers: A description of why and how, *Peabody Journal of Education,* 64:3, 131-150,

Saldana, J. (2015) Thinking Qualitatively, SAGE

Sanderson, I. (2003) Is it 'what works' that matters? Evaluation and evidence-based policy-making, *Research Papers in Education*, 18:4, 331-345 https://www.teachfirst.org.uk

Timperley H.S. and Alton Lee, A. (2008) *Reframing Teacher Professional Learning:* An alternative policy approach to strengthening valued outcomes for diverse learners, review of research education, 32:328-369.

Timperley H. S. (2008) Teacher professional learning and development; International academy of education. *Education practices.* Series 18

UCAS (The Universities and Colleges Admissions Service) (2019) https://www.ucas.com/teaching-option/school-centred-initial-teacher-training-scitt

Valli, L. (2000) Connecting teacher development and school improvement: Ironic consequences of a preservice action research course. Teaching and *Teacher Education*16 (7) 715-730.

Wall, K. (2008) Understanding metacognition through the use of pupil views templates; Pupil views of learning to learn. *Thinking skills and creativity* pg.23-33.

Zeichner, K. (2003) Teacher research as professional development for P-12 education in the US. *Education Action Research,* 11 (2), 301-25.

Zeichner, K. M. and Liston, D.P. (1987) Teaching student teachers to reflect. *Harvard Educational Review 56* (1), 23-48.