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 TWO

AN INTRODUCTION TO THE PORTFOLIO OF PUBLISHED PAPERS

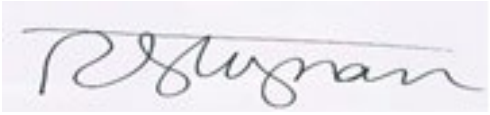
August 2021

A critical evaluation of criteria-based assessment of subject knowledge and other competencies of teachers in training

Richard James Tynan

Declaration

Submission to LJMU Doctoral Academy for PhD by publication by Richard James Tynan. This submission, in two parts, consists of nine outputs and an introduction according to the guidelines and regulations set out by the Doctoral Academy at Liverpool John Moores University. The seven lead author, one second author and one third author articles were published in peer reviewed journals between 2013 and 2019 and have not been used in the submission for another research degree.



18th August 2021

Acknowledgements

I would like to acknowledge the role of colleagues involved in this submission and thank them for their help and encouragement. During the preparation of the portfolio of papers: Michael Inglis for taking the lead in the beginning; Andrea Mallaburn, Robert Bryn Jones and Ken Clays for their contributions as co-authors; Matt McLain for introducing me to Q-methodology and mentoring me as I adapted it to our study; and Andrea Pratt for including me in her study. In preparation for the initial submission, I received much help and guidance from David Putwain and Mark Brundrett. During preparation of the Part 2: Introduction I have relied completely upon my excellent supervisory team for feedback and encouragement: Andrea Mallaburn, Linda Seton and Angela Daly.

I am greatly indebted to all of you and cannot thank you enough!

Tables

Figures

Abstract

This thesis draws together a portfolio of nine peer reviewed papers investigating the assessment of teacher competencies. Earlier papers studied the effectiveness of a particular model for delivering a subject knowledge enhancement course in creating new physical science teachers from those without chemistry or physics first degrees. This programme was attended, year-long, full-time and higher education accredited. Later papers widened the scope to include the assessment of teacher subject knowledge and other competencies in a range of school subjects in the English education system. The portfolio adopted a mixed methods approach with an emphasis on quantitative analysis. In considering the body of research, a critical realism perspective was adopted to describe and interpret the findings in terms of many demonstrable phenomena concerned with assessment as an event. An underlying mechanism is proposed to explain these in terms of the assessment behaviours of teacher practitioners who also act as teacher educator practitioners and may experience role conflict: *conflicted role, professional judgement*. This approach locates the portfolio's contribution to new knowledge against pre-existing work, makes clear unifying and coherent themes and establishes the papers' individual and combined methodological rigour. The research has great local utility in initial teacher education in England. It also demonstrated issues with criteria-based assessment of teaching described through competencies and the problems with adopting number grades to aid quality assurance and accountability. Findings have relevance and may be applicable to all similar assessment systems in any profession.

Contents

Portfolio of published works

1. 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), pp 101-107

2. 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

3. Tynan, R., Jones, R.B., Mallaburn, A. and Clays, K. (2016a) Working towards evidence-based practice in science teaching and learning, *School Science Review*, 97 (361), pp 109-115

4. Tynan, R., Jones, R.B., Mallaburn, A. and Clays, K. (2016b) Subject knowledge enhancement courses for creating new chemistry and physics teachers: the students' perceptions, *School Science Review*, 98(363), pp 109-114

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

6. Tynan, R. and Jones, R.B. (2018) Assessing trainee secondary teachers on school placement: Subject knowledge and overall teaching grades, *Teacher Education Advancement Network Journal*, 10(1), pp 20-34

7. 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? *Teacher Education Advancement Network Journal*, 11(1), pp 4-14

8. Tynan, R. and McLain, M. (2019) Attitudes to assessing trainee teachers on school experience placement within a group involved in an initial teacher education partnership at an 11-16 academy: a Q-methodology approach, *Teacher Education Advancement Network Journal*, 11(4), pp 3-20

9. Pratt, A. and Tynan, R. (2019) The journey from reflection on practice to research informed practice: the contribution of MA practitioner research, *Teacher Education Advancement Network Journal,* 11(4), pp 35-46

1. Introduction

> "… the practitioner researcher maintains a fluid and flexible stance with respect to each domain, behaving sometimes as a professional, sometimes as a researcher and at all times as an author who is making meaning out of the interactions and presenting them to an external audience."
>
> Drake and Heath, 2011, p 2

Drake and Heath (2011) acknowledged that, to qualify for a doctorate, practitioner research must necessarily create new knowledge. Their synthesis argued that practitioner research results in unique additions to knowledge by combining understanding from research, professional practice, and the individual's reflexive project.

A portfolio of nine papers is submitted here for consideration for PhD by publication.

It meets the summary of practitioner research attributes suggested by Goodfellow (2005), based upon a synthesis of research and authors' opinions in the field. As such, the papers are a systematic and critical investigation by initial teacher education (ITE) practitioners intending to improve professional practice, who took the responsibility to make new meanings and share them through publication in peer reviewed journals.

1.1 Research Questions

Investigating a series of linked themes, the studies in the papers sought to address two main research questions:

i. How successful in creating new physical science teachers was a particular model (attended, year-long, full-time, higher education accredited) for delivering subject knowledge enhancement (SKE) courses for non-subject specialist graduates?
ii. What were the factors influencing the assessment of teachers in training whilst on school experience placement?

Both questions were asked in relation to an assessment system involving number graded assessments against either the Professional Standards for Teachers (TDA, 2007) or the Teachers' Standards (Department for Education, 2011). This system was used across all programmes at a large ITE provider in the northwest of England.

1.2 Overview of the portfolio papers

Each paper reported practitioner research by teacher educators. These were initiated to investigate aspects of SKE courses and ITE programmes leading to qualified teacher status (QTS) in England. These can be associated with both of the research questions and were guided by the need for programmes to satisfy scrutiny by central policy makers, often through inspection of ITE partnerships by the Office for Standards in Education, Children's Services and Skills (OfSTED).

The research was undertaken in ITE partnerships in the northwest of England between schools and a Higher Education (HE) provider accrediting QTS. Initial investigations

evaluated the impact of one type of science SKE course on aspiring teachers. These led to questions regarding assessment practices that generated some of the evidential data. The tension between practitioner and researcher purposes led to the widening and developing of the research over six years to encompass a critical evaluation of the assessment of subject knowledge and other teaching competencies. The ITE practitioner might seek to gather sufficient evidence to support plausible explanations in answer to questions posed by stake holders. However, the researcher also sought to establish the validity of the studies and their findings. Ultimately, the research included the full range of teacher competencies described by the Teachers' Standards (DfE, 2011), the other school subjects, and the factors influencing the practice and outcomes of assessing teacher skills during mandatory school experience placements.

The papers constitute a body of teacher educator practitioner research utilizing quantitative and qualitative data with a high degree of local utility. They also demonstrate wider potential issues for teacher educators and others practicing in any professional accreditation system that uses competency-based assessment and grading. During the preparation of this submission the researcher has developed the application of critical realism for mixed methods research to help identify these wider implications (Archer et al, 1998; Bygstad and Munkvold, 2011; Downward and Maerman, 2006; and Zachariadis, Scott and Barrett, 2013). This is discussed further in sections **3. Methodology** and **5. Discussion of a possible underlying mechanism indicated by the portfolio**.

Together the papers contribute to a wider understanding of:

- mixed methods research
- the importance of professional subject knowledge for teachers
- the complexity of professional subject knowledge
- the assessment of teacher competencies during initial ITE programmes
- the influence of quality assurance and accountability management on assessment
- the critically reflexive journey towards research informed practice

1.3 A reflexive teacher educator practitioner research project

Returning to Drake and Heath's (2011) perspective on the generation of new knowledge through practitioner research, the early papers in this portfolio (Inglis *et al.*, 2013; Tynan *et al.*, 2014; Tynan *et al.*, 2016a, Tynan *et al.*, 2016b) provided a set of plausible explanations for patterns and themes in data concerning outcomes for newly qualified physical science teachers who had followed a SKE route, and their perspective of the course. One of the broad aims of the initial SKE paper (Inglis *et al.*, 2013) was to investigate the central policy theme that subject knowledge is a key measure of teacher worth, explicit and not yet superseded in the government white paper, *The Importance of Teaching* (Department for Education, 2010).

Tynan *et al*. (2014) demonstrated that year-long HE accredited SKE courses could create new physical science teachers that were indistinguishable to assessors and employers from those with a degree in their specialist teaching subject. One

interpretation of this was that the SKE course had provided the subject knowledge necessary for non-subject specialists to become effective physical science teachers. This key finding was welcome and useful to practitioners defending the status and funding of year long, face-to-face SKE programmes. However, this was only one possible explanation, amongst several plausible alternatives (Tynan *et al.*, 2014). The same data on equal outcomes could also have arisen because the assessment tools in use during the SKE and ITE programmes differed from those used in subject discipline degree programmes. It was also consistent with the practise of assessors on different programmes assessing different aspects of subject knowledge. Any of these explanations, or a combination of them, could explain why ITE assessors did not differentiate between aspiring physical science teachers who had followed a SKE route from those with a first degree in their teaching subject.

As the early research key findings relied heavily upon assessment data generated during ITE programmes (Tynan *et al*., 2014), establishing its validity and accuracy became a key research objective. This involved a wider investigation of the consistency of assessment outcomes and possible sources of subjectivity (Tynan and Mallaburn, 2017; Tynan and Jones, 2018, Tynan and Jones 2019; Tynan and McLain, 2019). This provided indications of the effectiveness of the assessment of subject knowledge and other teaching competencies. Pratt and Tynan (2019) considered the development of professional subject knowledge for teachers and the transformational learning required to move from reflective trainees to evidence or research based Newly Qualified Teachers (NQTs). This generated new knowledge concerning the validity of assessment practices for trainee teachers on school experience placement that used published criteria (Department for Education, 2011) and locally used guidance descriptors to assess and number grade teaching competencies. This reflexive journey is further expanded and discussed in sections **2-6**.

Four papers were published in The School Science Review (SSR), the Association for Science Education's (ASE) peer reviewed international journal. According to its own promotional material, the ASE (https://www.ase.org.uk/) has around 5000 members and the SSR is widely distributed to members in the United Kingdom, to science teachers, science teacher educators and around 290 libraries. The SSR also claims a sizeable overseas readership. So, it is read by science educators, including those who recruit and employ teachers, who might want to be better informed about the science teachers they encounter who hold SKE qualifications. The other papers were published in the peer reviewed international Teacher Education Advancement Network (TEAN) Journal (https://ojs.cumbria.ac.uk/index.php/TEAN/index) with a wide readership of teacher educators in schools and other institutions, in the United Kingdom and internationally. All papers were subject to rigorous, anonymous peer review.

As lead author for seven papers, I took full responsibility for each project: research design, seeking ethical approval, data gathering and analysis, initial writing, submission, and making amendments after peer review. For Tynan and McLain (2019), McLain mentored me in the use of Q-Methodology (Brown, 1980; van Exel and de Graaf, 2005) and associated software (Schmolk, 2014) and I designed and conducted the research. I would estimate my contribution to lead author papers at 90% or above. As third author for Inglis *et al*. (2013) I contributed to the early planning and design stages, and to the

first and final drafts (5%). For the last paper in the portfolio (Pratt and Tynan, 2019), Andrea Pratt was the lead author and I contributed to the data analysis, writing a section in the findings (10%).

2. Relevant Literature

The portfolio of papers submitted have separate literature reviews in keeping with the editorial policy of the journals that published them. The areas discussed below demonstrate the coherence of the papers and their common themes. They have been useful in setting the context for the studies and interpreting findings.

2.1 Qualifying to be a teacher in England

Robinson (2006) argued that the history of ITE in England has been characterised by two strands: broad cyclical swings between school and higher education led models of provision, and the politics and provision of state education. Two major pieces of legislation (Department for Education, 2010 and 2011) indicate that the research in this portfolio took place during a period of government control over ITE programmes that favoured multiple routes into the teaching profession, school led provision and a subject based school curriculum. Within this context, the role of school-based mentors has expanded to include the assessment of trainees' teaching skills as a key component (Tynan and McLain, 2019).

The system of providing evidence of competence for the recommendation of QTS in England and the influence of OfSTED frameworks for the inspection of ITE partnerships was summarised in Tynan and McLain (2019). To be recommended for QTS in England, aspiring teachers must evidence that they have reached minimum performance levels in eight standards and satisfy descriptors for professional conduct (Department for Education, 2011). The Teachers' Standards (Department for Education, 2011) provide descriptors for the minimum level of performance for specified teacher competencies. However, without prescribing extra criteria, OfSTED (OfSTED, 2018) have held providers accountable according to the number of teachers surpassing minimum expectations. Further, OfSTED's use of a four-point scale to grade partnerships and QTS providers, resulted in some providers using number grades linked to OfSTED categories for some or all formal assessments. The institutional response at the authors' QTS provider, in ascribing numerical grades at both formative and summative review points during ITE programmes, is described in Tynan and MClain (2019) and much of the research in the portfolio submitted investigates assessment outcomes in the form of number grades (Tynan *et al.*, 2014; Tynan and Mallaburn, 2017; Tynan and Jones, 2018; Tynan and Jones 2019).

In 2010, just prior to the start of the research in the portfolio of papers, central policy makers in England signaled reform of the curriculum and its assessment in secondary schools and specified subjects as the basis for curriculum design, and end test examinations as the mode for assessing learning (Department for Education, 2010). With the importance in school curricula of subject specialism established, the same white paper formalized through applicant funding that teachers' specialist subject knowledge should be considered a key indicator of potential teaching ability. Teachers in all but some shortage subjects were expected to hold at least a 2:2 Honours

classification in their first degree (Department for Education, 2010), but the government was willing, with some rules, to accept a SKE qualification for programmes that enhanced subject discipline knowledge to teach shortage subjects (Inglis *et al.*, 2013).

2.2 Subject knowledge for teachers

Inglis *et al.* (2013) identified the nature of teacher subject knowledge as a contested area but selected Shulman's (1986) influential model as a useful way to describe aspects of teacher subject knowledge easily recognisable to education and teacher education practitioners. Elements of subject knowledge for teachers covered by performance descriptors in the Teachers' Standards (Department for Education, 2011) can be identified using Shulman's categories (1986), subject matter content knowledge (SMCK), pedagogical content knowledge (PCK) and curriculum knowledge (CK) (Shulman, 1986). Inglis *et al.* (2013) gave a clear explanation of Shulman's (1986) model and its relationship to other similar models such as Banks, Leach and Moon (2005). All the papers in this portfolio utilised Shulman's (1986) model whilst accepting there are other influential models. For instance, Banks, Leach and Moon (1999) included school knowledge in their consideration of professional knowledge for teachers and authors such as Turner-Bisset (2006) and Korthagen (2017) have argued for the inclusion of teacher self-knowledge.

Although elements of subject knowledge can be linked to all Teachers' Standards descriptors (Department for Education, 2011), at that time of data gathering for Inglis *et al.* (2013) and Tynan *et al.* (2014) its elements were mainly covered by descriptors Q14 and Q15 of the Professional Standards for Teachers (TDA, 2007). From September 2012, after the standards descriptors were revised, they were found mainly in standards 3 and 4 of the Teachers' Standards (Department for Education, 2011). All the subsequent papers refer to the revised descriptors. Shulman's model (1986) was retained throughout this portfolio of papers as a useful conceptual framework for investigations. It could be argued that, despite different perspectives on subject knowledge for teachers, the Teachers' Standards (Department for Education, 2011) directed teacher and teacher educator practitioners towards subject knowledge in the titles of standards 3 and 4. Standard 3 refers to subject and curriculum knowledge, which can be linked to SMCK and CK. Standard 4 includes PCK when it refers to planning and delivering effective lessons (Department for Education, 2011). However, in Standard 4, it should be noted that the Teachers' Standards (Department for Education, 2011) are referring generically to learning and teaching strategies but for Shulman (1986) PCK was a subject specific construct.

Literature indicated that Shulman's model (1986) continued to be current, cited and further developed by researchers (Tynan and Jones, 2018). The inter-relationship between the three categories suggested by Shulman (1986) and later sub-categories (Lucero, Petrosini and Delgado, 2017) were useful in attempting to explain the patterns in assessment outcomes observed (section **4. Outcomes of the research portfolio**). Inglis *at al.* (2013) described the government's intention that SKE courses should only deal with SMCK, to be followed by a suitable ITE programme that developed PCK and CK. Inglis *at al.* (2013), Tynan *et al.* (2014), and Tynan *et al.* (2016b) established the appreciation of students for components of the SKE course that touched upon PCK and

CK as well as SMCK, and for the mutual peer and tutorial support associated with face-to-face SKE courses.

2.3 Good teachers and ITE programmes

No matter how it is described, secure professional subject knowledge is only one attribute of successful teachers. Discussing dominant discourses on good teachers, Moore (2004) identified three publicly held perceptions also used by education policy makers. Moore (2004) called these the competent craftsperson, the reflective practitioner, and the charismatic subject. According to Moore (2004) the competent craftsperson was perceived as demonstrating definable, transferable knowledge and skills that an apprentice might learn, and the reflective practitioner responded reflexively to their own teaching in order to learn by experience. Only the charismatic subject was born to teach and, therefore, beyond the scope and range of ITE programmes. Moore (2004) argued that ITE programmes understandably sought to draw on discourses that allow for professional learning, such as the competent craftsperson and reflective practitioner, and that these were then difficult to develop or displace. These discourses still have currency in discussions of teachers' professional learning, school-led ITE programmes (Philpott, 2014; Door,2014) and the role of teacher educators (Czenerniawski, 2018).

However, they need to be considered in the context of requirements for QTS in England. The Teachers' Standards (2011) in England describe teaching competencies using criteria for minimum performance in eight teaching standards and three professional standards that are acceptable for the recommendation of QTS. This analytical approach to describing teaching is associated with the expectation that most of a teacher's training takes place in school under the guidance of school-based teacher educators and mentors (OfSTED, 2018). This supports the idea that central policy makers perceive teaching in terms of measurable skills that can be learned from more experienced colleagues and, to some extent, by trial and error. Previously, more holistic ways of evaluating teachers were acceptable and widely utilized (Leshem and Bar-Hama, 2008).

Indeed, Cajkler and Wood (2016, p1) used the term 'reductive models' for competency-based approaches to organizing and assessing ITE programmes. They advocated a more literate use of criteria and a collaborative lesson planning, observation and evaluation strategy as more effective and less restrictive for developing teaching skills. Korthagen (2017) argued that good teachers simply made consistently good decisions, inside and outside of the classroom, that result in learning. Korthagen's Onion Model for Reflection (Korthagen, 2017) acknowledges that these decisions are affected by the environmental context as well as based upon underlying intrinsic teacher attributes affected by a teacher's values and motivations (Figure 1). In considering ITE programmes and their assessment, Korthagen (2017) maintained that teacher competencies have a place in ITE programmes but only comprise the middle layers of a much wider range of teacher attributes (Figure 1) that contribute to this decision making.

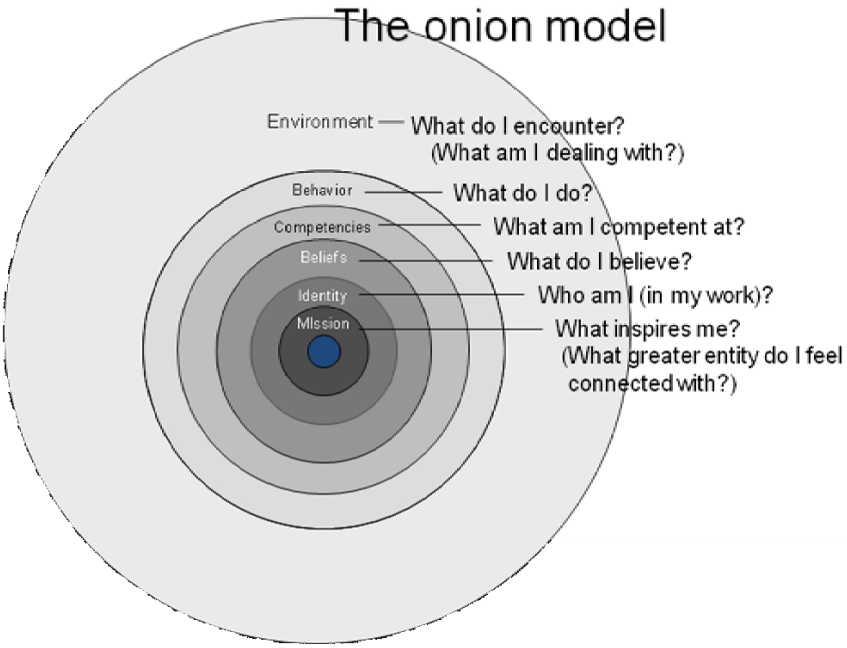


Figure 1. The Onion Model: Levels of Reflection Korthagen (2004) cited in Korthagen (2017, p395)

Throughout the submitted portfolio of papers, a theme developed questioning the trustworthiness of graded assessments of teacher competencies using reductive standards descriptors as a measure of teacher worth. Did mentors assess their trainees holistically or analytically even when using descriptors? What were the mentors' benchmarks? How did mentors manage the requirements for recommending QTS, adhering to agreed partnership assessment practices, and quality assurance procedures? It is, perhaps, not possible to directly measure the true role of the standards descriptors (Department for Education, 2011) when mentors evaluated trainees and reached assessment judgements.

2.4 A model for professional learning and its assessment

Inglis *et al.* (2013) and Tynan *et al.* (2014, 2016b) established that a face-to-face, year-long HE accredited SKE course was perceived as appropriate by participants, who expressed high levels of confidence in their ability to teach their SKE course subject. These papers used Shulman's (1986) model for teacher subject knowledge. However, Inglis *et al.* (2013) observed that Shulman's model (1986) could be criticized for emphasizing individual cognitive development and not social aspects of professional learning. Tynan *et al.* (2014) introduced Martin and Cloke's (2000) application of Hager and Butler's (1996) treatment of professional learning that includes two different assessment models for different stages of professional development. This model describes professional development in three phases: acquiring knowledge, practicing in a safe environment, and demonstrating competence in the real world. However, assessment during these stages requires different approaches. Hager and Butler (1996) and Martin and Cloke (2000) argued that preparatory knowledge is easier and more

likely to be assessed using tools such as continuous assessment and examinations that are usually considered more objective and scientific. Assessing the application of knowledge requires a more qualitative and judgmental approach often based upon observation (Hager and Butler, 1996; Martin and Clokes, 2000).

A consideration of Martin and Cloke's (2000) application of Hager and Butler's (1996) model raised the possibility that similar assessment outcomes for teachers with SKE and subject specialist degree qualifications might not reflect equal performance. It was possible that the assessments were accurate yet arrived at through differing assessment tools. Alternatively, assessors may have placed their emphasis on other aspects of teacher subject knowledge to subject discipline knowledge gained through first degree or SKE courses. Similar assessment outcomes might also be explained by a change to a broader and less discriminating qualitative, judgmental assessment model for the later stages of professional learning.

In later portfolio papers (Tynan and Mallaburn, 2017; Tynan and Jones, 2018; Tynan and Jones; 2019; Tynan and McLain, 2019) a recurring theme developed that there was less than expected variability in assessment data. Hager and Butler (1996) and Martin and Cloke's (2000) model predicted many opportunities for subjective differences between assessors, and their interpretation and application of assessment practices. However, these were not demonstrated quantitatively in graded assessment data. In several portfolio papers Hager and Butler (1996) and Martin and Cloke's (2000) model was applied to the structure of ITE programmes at the HE QTS provider and Figure 2 shows the final version constructed (after Tynan and Jones, 2018). This simple proposed process model does not assume individual cognitive or social models of professional learning and development, but acknowledges the social element implied by qualitative, judgmental assessment. Steps were taken by the HE provider in the studies to standardize assessment practices within and across ITE programmes (Tynan and Mallaburn, 2017) but the assessment of trainee teachers remained qualitative rather than scientific (Tynan and Jones, 2018; Tynan and Jones, 2019; Tynan and McLain, 2019). This led repeatedly to a consideration of potential issues with assessment against criteria referenced competencies and number grading.

2.5 Institutional Ethnography (IE) and Actor Network Theory (ANT)

Tynan and Mallaburn (2017), Tynan and Jones (2018, 2019) and Tynan and Mclain (2019) suggested strongly that the management of assessment and its quality assurance had influenced assessment outcomes in a way that could be explained by applying IE and ANT models. These interrelated models aim to understand how people interact with the organisations they work for. The organisation needs to communicate its position and requirements clearly through documentation (IE) but human actors must work with this to implement it (ANT). Current ideas about IE are based upon the work of Dorothy Smith (Smith, 2005) whilst those around ANT have been attributed to Bruno Latour, John Law and Michael Callon (Fenwick and Edwards, 2010). Smith (2010) explained IE analysis as finding an institution's standpoint to direct and provide a relevant framework for a sociological study. It starts with the local realities for people working within an institutional process. Applying ANT to education, Fenwick and Edwards (2010) described it as an interrelated mix of natural, social and technical phenomena enacted through webs of human activity i.e., human actors working within their context and environment.

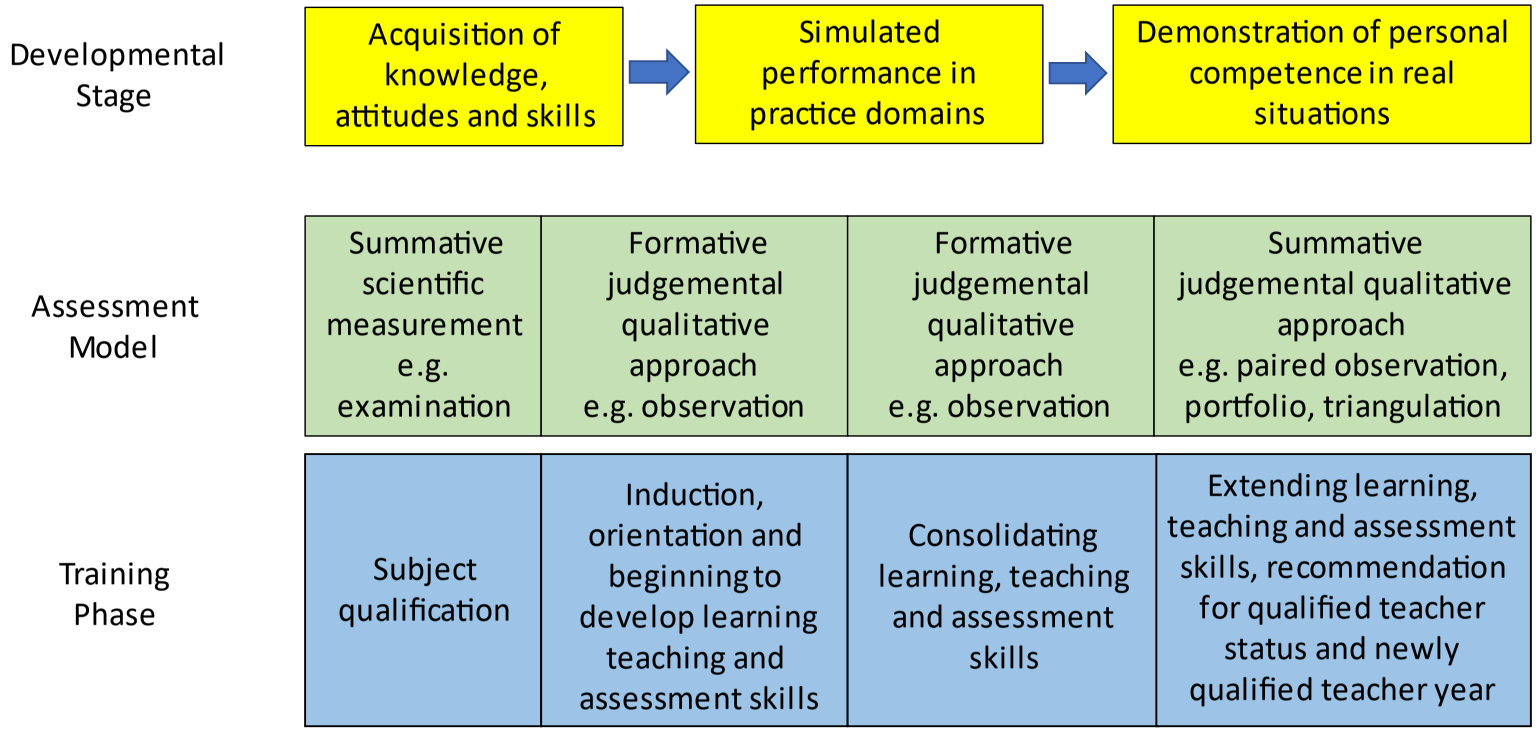


Figure 2.
A model for the professional development of trainee teachers after Tynan and Jones (2018), Martin and Cloke (2000) and Hager and Butler (1996).

For Tummons (2010) and Tynan and Jones (2019) these were interpreted pragmatically as the documentation pertaining to assessment, and the network of practitioners acting as advocates for the agreed approaches in the documentation. Tynan and Jones (2019) provided some quantitative evidence that the HE QTS provider and its partnerships had established a clearly recorded and communicated standpoint on assessments and grades (IE) that was successfully implemented by human advocates (ANT). Qualitative evidence reinforced this strongly with compliance issues and quality assurance processes uppermost in the minds of a small group of ITE participants (Tynan and McLain, 2019).

2.6 Summary

The literature discussed above demonstrate the themes common to this portfolio of papers and models found to be useful during the studies. It considers:

- the criteria-based assessment of teaching competencies required for the recommendation of QTS in England
- professional subject knowledge for teachers and subject knowledge enhancement
- wider perspectives on the attributes of successful teachers
- a proposed model for professional development and assessment and
- the possible impact of IE and ANT and the management of assessment and quality assurance on assessment outcomes.

These have all been found to be relevant when attempting to answer the research questions posed in section **1. Introduction (1.1)**.

3. Methodology

3.1 Researcher standpoint

The portfolio of papers submitted constitute a series of small-scale practitioner research investigations published over six years. The research utilised mixed methods with an emphasis on quantitative approaches involving statistical analysis. In the process of gaining acceptance for publication, the research paradigm was pragmatic (Morgan, 2017). Each paper was subject to rigorous peer review according to journal editorial policy and was accepted upon the basis of the validity and rigour of each paper's methodology. This was argued from a positivist or constructivist perspective as required, adopting the pragmatic perspective that ontological considerations could be set aside temporarily whilst trying answer questions about what it means to conduct research and experience the consequences (Morgan, 2017).

However, this portfolio has also been considered from Zachariadis, Scott and Barrett's (2013) standpoint. They argued that adopting a critical realism perspective (Bhaskar, 1997) alters considerations of design, measurement and inferential validity for both positivist and constructivist research, allowing both to contribute to the overall

methodological rigour of mixed methods investigations. Whilst investigating some very specific practitioner research questions, new knowledge was generated that indicates and identifies an underlying mechanism concerned with assessing teachers.

3.2 Research paradigms and mixed methods research

Kuhn (1970) first developed the concept of paradigms, describing science as a social activity and demonstrating that some scientific theories persisted as paradigms long after contrary objective evidence became available. This could be for a variety of expedient sociological, political, religious or economic reasons unrelated to scientific investigation (Kuhn, 1970). He proposed the term paradigm shift for the eventual revolution in thinking needed to replace them. Kuhn (1970) used historical examples mostly from the natural sciences and thought this concept was unsuitable for describing other types of research. However, the word paradigm now has common usage as an idiomatic way of expressing the concept of any normal or acceptable way of thinking about something. In the context of methodology, research paradigms are widely accepted approaches to systematic investigations that, unlike scientific theory paradigms, can co-exist simultaneously without resolution because they derive from different philosophical standpoints on ontology (study of being), epistemology (study of knowledge) and axiology (study of ethics), together with their associated methods (Aliyu *et al.*, 2015; Kivunja and Kuyini, 2017).

Authors describe the major research paradigms using different terminology. Aliyu *et al.* (2015) discussed ontology, epistemology, and axiology with respect to quantitative and qualitative research methodologies by applying them to positivist, interpretivist, and critical theory research paradigms. Kivunja and Kuyini (2017) similarly discussed locating research in positivist, post positivist, constructivist (interpretivist), critical (transformative) and pragmatic approaches through their associated methodologies. Aliyu *et al.* (2015) and Kivunja and Kuyini's (2017) papers summarised the three major research paradigms, describing positivist methodologies as emphasising the search for intransitive, external realities existing independently of the researcher. The authors argue that for constructivists, knowledge is fully transitive, created and amended by individuals acting socially even when behaving scientifically, and that the critical theory research paradigm emphasises socio-political realities. In their analysis of research paradigms, Aliyu *et al.* (2015) explored the methods used within paradigms in detail and, whilst accepting the shortcomings of positivist research methods first highlighted by Kuhn (1970), they rejected constructivism as its replacement. They considered it an incomplete response, over-emphasising the contribution of perception and communication to generating knowledge. Aliyu *et al.* (2015) preferred a participatory world view of research, which acknowledges that the researcher is one subject amongst many and integrating their action with their knowing. This appears to share some attributes of the pragmatist researcher paradigm (Morgan, 2017) without addressing the ontological tensions between realist and constructivist methodologies (Kivunja and Kuyini, 2017).

The mixed method portfolio of papers submitted in Part 1 investigated the assessment of subject knowledge and professional learning of aspiring teachers. It is probably best

described as a sociological or social psychological study of a peopled and, therefore, open system periodically closed by assessment events (Danermark, Ekström and Karlsson, 2019). The individual papers generated findings appropriate to the qualitative and quantitative methods summarised in this section. Separately, these findings were valid, thought-provoking, and beneficial, providing a record of regularities and patterns in data at certain points in time. However, when considering the portfolio as a coherent body of work, maintaining a pragmatic research position made final abstractions, conclusions, and theorisations from these separate findings difficult to justify. During the later stages of this submission another perspective, critical realism (Archer *et al.*, 1998), was considered and applied to the portfolio of papers to help unify the findings. Adopting a critical realism standpoint with a mixed methods portfolio also reduced the ontological and epistemic tensions between positivist and constructivist positions and their associated quantitative and qualitative approaches to research (Zachariadis, Scott and Barrett, 2013). Critical realism (Archer *et al.*, 1998) moves beyond the pragmatic approach to research that adopts a positivist, constructivist, or other research paradigm as the research question requires it. Critical realism assumes that knowledge has both intransitive and transitive domains and argues that there are underlying structures and mechanisms external to the subject that cause events, some of which can be experienced as phenomena by humans and investigated; Bhaskar's (1997) Real, Actual and Empirical Domains. On the other hand, research is a human and interactive social activity constrained by our ability to perceive or measure events. So, science and research generate knowledge regarding structures and mechanisms that is also undeniably psychologically and socially constructed, understood and shared (Archer *et al.*, 1998).

Bhaskar (1997) explained that critical realism could provide an ontology and epistemology that accounted for the existence of scientific investigation. Archer *et al.* (1998) expanded the argument to include sociological research. Danermark, Ekström and Karlsson (2019) summarised Bhaskar's (1997) and Archer *et al.*'s (1998) philosophical position with the goal of explaining critical realism's ability to reconcile ontological and epistemological differences between quantitative and qualitative methodologies for mixed methods approaches and multidisciplinary investigations. For Danermark, Ekström and Karlsson (2019) a pragmatic research approach leaves a dichotomy between quantitative and qualitative methodologies that obscures the critical realism perspective on the stratification of reality, and intransitive and transitive elements of ontology and epistemology.

Applied to mixed methods research, this means that critical realism allows a change in perspective from searching for causal relationships between phenomena to identifying the effect of underlying structures and mechanisms on them, some of which may be social and psychological (Zachariadis, Scott and Barrett, 2013). In this analysis (Zachariadis, Scott and Barrett, 2013), the quality of research can be gauged by considering internal validity (design and measurement), external design validity (or transferability), and inferential validity. It can be argued that adopting a critical realism perspective renders these similar for both quantitative and qualitative methodologies (Zachariadis, Scott and Barrett, 2013). Both qualitative and quantitative approaches are

effective in their domains and mutually interact to generate knowledge about potential relational causal powers triggered in certain contexts (Archer *et al.*, 1998; Danermark, Ekström and Karlsson, 2019.) Danermark, Ekström and Karlsson (2019) also argue that critical realism helps address issues with sociology research paradigms associated with quantitative and qualitative approaches that focus on either social structures or agency. This means that both quantitative and qualitative data have the possibility of contributing to the identification of underlying structures and mechanisms without making assumptions about knowledge from positivist, constructive or other standpoints (Zachariadis, Scott and Barrett, 2013).

The submitted portfolio used quantitative and qualitative approaches to demonstrate the phenomena associated with an event: the number graded assessment of subject knowledge and other teacher attributes. Adopting a critical realism approach in order to explain the research journey and the coherent themes linking the phenomena involved abduction and retroduction, in addition to other forms of inference associated with qualitative and quantitative approaches (Downward and Mearman, 2006; Bygstad and Munkvold, 2011; Danermark, Ekström and Karlsson, 2019). In the papers submitted in Part 1, both quantitative and qualitative approaches contributed to abduction, or theoretical re-description of the findings, and to retroduction, the identification of a possible causal mechanism.

Critical realists perceive causal powers as natural necessities (Archer et al., 1998), structures, powers, mechanisms, and tendencies that have intrinsic properties with the potential to influence events (Danermark, Ekström and Karlsson, 2019). Counterfactual reasoning (Archer et al., 1998; Danermark, Ekström and Karlsson, 2019.) has been found useful for thinking about causality across disciplines (Pearl, 2009; Levy, 2015). This involves thought experiments that consider assumptions derived from models and empirical findings and theorise on the possible consequences of omitting elements in order to speculate on possible alternative outcomes. For example, the portfolio papers submitted adopted specific models for describing teacher subject knowledge (Shulman, 1986) and a process for professional development and assessment (Hager and Butler, 1996; Martin and Cloke, 2000). Accepting these models predicted many opportunities for assessors to demonstrate subjective differences. These should have been observed as variability in assessment outcomes but were not. Considering the elements of the models that would need to be removed to explain the empirical mixed study findings and theorising on possible explanations for the apparent anomalies between these and model predictions guided thinking about possible causal mechanisms as described in **Section 5 Discussion**.

3.3 Overview of Methods

In the planning phase, all the papers were classified as pedagogical studies by the HE institution and were submitted for approval by a research ethics committee because of the intention to publish in a peer reviewed journal. The individual portfolio papers address their research ethics considerations.

Overall, the participants in the research represented all roles within the ITE partnership for the ITE programmes investigated. The papers' central theme was the assessment of

trainees on school experience placement. Mentors in school subjects made initial judgements that involved their trainees, and these were moderated by professional mentors, mentors from other schools and liaison tutors from the QTS provider. All aspects of the partnerships were externally scrutinised through Ofsted inspections. The participants in the quantitative studies focussing on number graded assessments were all the partnership members that contributed to the process of assessment. The data was collected to facilitate trainees' progress through ITE programmes and ultimately their recommendation for QTS. For these investigations it was only necessary to obtain gatekeeper permission to access the numerical data with the understanding that any raw or analysed data was reported anonymously.

For Tynan *et al.* (2014), Tynan *et al.* (2016b) and Tynan and McLain (2019) individuals were asked for permission to publish their anonymous responses to questionnaires or a Q-method task. The researcher followed the HE QTS provider's regulations and guidelines for obtaining informed, implied consent and protecting participants and researchers. The participant information emphasised that they did not have to take part and could withdraw their data at any time up to the publication of the papers. Either course of action would not lead to any detrimental effect on their working relationship with the ITE partnership or the researchers. For SKE programmes participants this was very important as they were invited to take part prior to continuing onto their PGCE course.

Returning a questionnaire or completing the Q-method on-line task implied consent. Participants were provided with the researchers' contact details should they wish to withdraw their data. These studies did not involve questions regarding controversial or potentially damaging areas, nor did they ask for personal data. However, they were of intrinsic interest to the respondents as well as the researchers. A few did not want their responses included in the questionnaire studies, but no requests were received from respondents to withdraw data from the study. The Q-method convenience sample constituted a small proportion of those originally invited to participate but was valid for the purpose of the paper (Tynan and McLain, 2019) and there were no requests to withdraw data.

The data gathering, and analysis used in the submitted portfolio are summarised in Table 1.

3.3.1 Data gathering

To clarify what is meant in the table headings under data gathering in Table 1, none of the papers constitute a meta-study of literature (Cooper, 2017) but all required a systematic literature review related to each individual study.

Programme document review refers to institutional documents (e.g., mentor guidance materials, formal review forms, assessment data spreadsheets) used as a source of qualitative or quantitative data. These were produced during the normal activities related to the SKE and ITE programmes involved in the studies.

Responses to end-of-course questionnaires were routinely invited from SKE programme students to enhance and expand on the university student feedback systems existing at that time. Informed implied consent was obtained from SKE students for the results to be used anonymously in articles and papers.

With gatekeeper permission, the assessment data sets used were taken from institutional summary spreadsheets recording formative and summative student assessments. Subject mentors used number graded assessments at formal review periods throughout and at the end of all the ITE programmes at the HE provider.

Q-methodology (Brown, 1980; van Exel and de Graaf, 2005) utilised an anonymous, on-line, constrained choice prioritisation exercise. This required participants to consider the relative importance of a concourse of 42 statements about the assessment of trainees on school experience placement. This is explained further below in section **3.5 Q-Methodology**.

3.3.2 Qualitative analysis

To clarify the table headings in Table 1 under qualitative analysis, reporting refers to the citing or description of information in text with minimal interpretation.

Thematic analysis refers to the process of categorising data within text and identifying overarching themes and sub-categories, followed by their interpretation and summary (Saldaña, 2009). All papers in the portfolio have an element of qualitative thematic analysis through the literature reviews and document analysis. However, Tynan *et al.* (2016b) and Tynan and McLain (2019) were concerned specifically with qualitative primary data gathering, analysis, and interpretation. Tynan *et al.* (2016b) used thematic analysis of anonymous open written responses from an end of programme questionnaire to investigate student attitudes and confidence as they started their Post Graduate Certificate in Education (PGCE) programme. Co-authors reviewed the paper and acted as critical friends for the author's coding. However, coding is a subjective inductive process, and the results were interpreted with reference to the researcher's practitioner activity (Saldaña, 2009). Tynan and McLain (2019) utilised Q-methodology, which is discussed more fully below in section **3.5 Q-Methodology**.

3.3.3 Quantitative analysis

Where quantitative data was gathered, the most appropriate methods of analysis and presentation was selected from several available. This depended on the nature of the data set and the purpose of the investigation and is explained further in this section and in section **4. Outcomes of the research portfolio**.

In Table 1 under quantitative analysis, description of numerical data is a verbal or visual presentation of raw data with minimal further analysis.

Simple calculations included the reporting of percentages, means and standard deviations without employing further statistical analysis.

Both parametric and non-parametric statistical methods were employed as described in basic statistical texts (Langley, 1968; Upton and Cook, 1996 and Hinton, 2014). Parametric methods used to test statistical significance were the visual comparison of means and their 5% confidence limits, one-tailed analysis of variance (ANOVA), and Pearson's Correlation Coefficient (Langley, 1968; Upton and Cook, 1996 and Hinton, 2014). The non-parametric tests of significance were Chi Squared analysis (with and without the use of contingency tables) and Spearman's Rank Correlation Coefficient (Langley, 1968; Upton and Cook, 1996 and Hinton, 2014). The effect size metric used for Tynan and Jones (2019) was Proportion of Variance (POV) expressed as a percentage (Ellis, 2010). It is worth noting that the factor analysis in Q-methodology is based upon Spearman's Rank Correlation Coefficient (Brown, 1980; van Exel and de Graaf, 2005).

3.4 The use of statistical analysis

The statistical methods employed are established statistical methodology. The arguments and explanations behind the statistical approaches adopted in the quantitative papers can be found in basic starter texts on quantitative and statistical analysis such as Langley (1968), Upton and Cook (1996), Ellis (2010), Hinton (2014) and Cooper (2017). The suitability of the approaches used is expanded upon below.

The quantitative papers in the portfolio demonstrate the three interrelated aspects of statistical analysis discussed by Ellis (2010) and Cooper (2017): tests of significance, statistical power, and effect sizes. Some papers in the portfolio utilised statistical tests of significance to provide an objective measure of the likelihood that observed phenomena could be explained by random events. The 5% error statistical benchmark was rigorously applied (Ellis, 2010; and Cooper, 2017). Thus, concluding that a pattern in the data was non-random, with a five percent or less probability of error, was taken as an indication that this was worthy of further investigation. This reflects the critical realism perspective that other evidence of possible underlying structures and mechanisms should be sought rather than implying that causal relationships between phenomena had been demonstrated (Zachariadis, Scott and Barrett, 2013).

The second aspect of statistical analysis is the statistical power of a test of significance to detect non-random patterns from samples taken from whole populations (Ellis, 2010; Cooper, 2017). This is very much improved by a large sample size (Ellis, 2010; Cooper, 2017). However, the quantitative studies in the portfolio (Tynan *et al.*, 2014; Tynan and Mallaburn, 2017; Tynan and Jones, 2018; Tynan and Jones, 2019) utilised statistical analysis not statistical inference. They investigated patterns in data sets without seeking to extrapolate statistical conclusions to any larger population of trainee teachers. The statistical investigations utilised data from full populations of trainee teachers for statistical analysis, or as close to full as would make no difference to the statistical conclusions. The studies may be described as small-scale because they utilise local not national data. However, because the statistical tests of significance were used analytically with defined populations, they have the maximum statistical power to identify patterns in data that have a low probability of being randomly generated. For this reason, statistical power is unlikely to have caused any issues concerning the statistical conclusions reached.

Further, tests of significance inherently allow for the effects of sample or population size in the probability tables used when reaching statistical conclusions (Lindley and Miller, 1953). The chance of falsely concluding that there is a non-random pattern in the data is reduced by increasing the critical values needed to demonstrate significance as population or sample size decrease (Langley, 1968; Upton and Cook, 1996; Ellis, 2010; Hinton, 2014; and Cooper, 2017). The rigorous application of the protocols for each test used ensured that conclusions about significant differences were associated with the same accepted benchmark probability of error (5% or less) regardless of population size.

The cohorts participating in the statistical and effect size studies ranged from 8 to 238. It is important to note that, although these defined populations would have been small and localised for an inferential study, for analytical studies most of the populations defined were quite large (Lindley and Miller, 1953). So, the statistical conclusions regarding the populations defined in the studies were deduced with confidence that the statistical power of tests were sufficient to identify non-random patterns in the data with a quantified chance of error. Also, it should be re-emphasised here that, adopting a critical realism perspective for mixed methods investigations (Zachariadis, Scott and Barrett, 2013) allowed the findings, interpretations and conclusions from the separate studies to contribute to the overall findings. This, in turn, informed the wider implications of the portfolio for assessors of professional learning working with similar assessment tools.

Finally, statistical tests only indicate that an event is significant on the basis that it is unlikely to be a random occurrence (Ellis, 2010; Cooper, 2017). They do not indicate the scale (large or small) of the practical implications of any non-random differences or correlations identified (Ellis, 2010; Cooper, 2017). Effect size metrics can be used to indicate the practical impact of being in one group compared to another without assuming any significant differences or correlations (Ellis, 2010; Cooper, 2017). Tynan and Jones (2019) explored the correlation data linking grades ascribed for overall teaching and grades for individual Teachers' Standards (Department for Education, 2011) using the POV effect size metric expressed as a percentage (Ellis, 2010). This gave a percentage score indicating how closely assessors associated grades for different individual standards with a trainee's overall teaching grade based upon the amount of shared variance for a large ITE programme over a three-year period. This suggested the emphasis placed upon individual standards when evaluating a trainee's overall teaching ability based upon all eight standards.

The use of statistical analysis caused much discussion during the peer review process. Using statistical tests for inference and direct extrapolation to larger populations appeared more widely known than their use for analysis. The original use of Chi Squared tests to compare observed counts with results predicted by modelling was well understood, but less well known was the development of its multiple uses with contingency tables. This appears to be a cyclical problem experienced by quantitative researchers. Norman (2010) discussed and strongly challenged frequent criticisms of the use of parametric tests of significance (that assume data has a normal distribution) on the grounds of small sample size, non-normal distribution of the data, or their use

Table 1. Data gathering and analysis

Paper	Data gathering method					Qualitative Analysis				Quantitative analysis						
	Literature review	Programme document review	Questionnaire	Assessment data set	Q method	Report	Thematic analysis	Q: factor analysis	Description of numerical data	Simple calculation (%, mean, standard deviation)	5% confidence limits	ANOVA	Chi squared	Pearson's correlation coefficient (r)	Spearman's Rank correlation coefficient (r)	Proportion of Variance (POV) (%)
Inglis et al., 2013	✓	✓				✓	✓			✓						
Tynan et al., 2014	✓	✓	✓						✓	✓			✓			
Tynan et al., 2016a	✓	✓				✓	✓		✓	✓						
Tynan et al., 2016b	✓		✓			✓	✓		✓							
Tynan & Mallaburn, 2017	✓			✓							✓	✓	✓	✓	✓	
Tynan & Jones, 2018	✓			✓					✓				✓			
Tynan & Jones, 2019	✓			✓										✓		✓
Tynan & McLain, 2019	✓				✓	✓	✓	✓		✓					✓	
Pratt & Tynan, 2019	✓	✓				✓	✓									

with ordinal data. Early in the portfolio, to gain editorial acceptance of Tynan *et al.* (2014) the analysis was amended to avoid using means and 5% confidence limits (a parametric approach) with Likert (Edmondson, 2005) and other ordinal scales in favour of the Chi Squared (non-parametric) test of significance (Langley, 1968; Upton and Cook, 1996 and Hinton, 2014). If Norman (2010) was correct this was an unnecessary change. This led to a data-led consideration of the suitability of statistical testing with ordinal numerical data such as number graded criteria-based assessments.

Tynan and Mallaburn (2017) tested some of Norman's arguments and confirmed that using parametric or non-parametric (not assuming normal distributions) correlation coefficient calculations on ordinal assessment data made no difference to the conclusions reached about the significance of the correlation coefficient. After this paper was published, it was possible to argue more confidently and successfully during the peer review process for all the chosen statistical methodologies. During the preparation of Tynan and Jones, (2018) the method for using Chi Squared as a test of association by calculating the expected values utilising contingency tables was questioned. Citing a very old paper by Dice (1945), it was established that this was a widely accepted approach with considerable longevity used by biologists investigating if two organisms were found together (associated) or apart (dissociated) more often than would be expected if they were both randomly distributed. This was the same approach adopted in Tynan and Jones (2018) to question if grades for individual teaching competencies were associated or dissociated with grades ascribed for overall teaching.

Norman (2010) was responding to a need to counter misconceptions about the use of statistical analysis that appear to arise and be perpetuated through psychology, social science and medical research publications (Gaito, 1980). These are addressed by statistics experts only to be forgotten and recur cyclically (Gaito, 1980). Bacchetti (2002) argued that the peer review system for medical papers often faced criticism for missing flaws in research but was equally guilty of identifying flaws where none existed. This error was often linked to the use of statistics and was time consuming and demoralising for researchers to counter (Bacchetti, 2002). It can be argued that the use of quantitative analysis in this portfolio has extended and contributed to the discussion concerning the uses and possible misuses of statistical analysis in small scale social and psychological research.

3.5 Q-methodology

Quantitatively, Q-method factor analysis (Brown, 1980; van Exel and de Graaf, 2005) involves using a matrix of non-parametric rank correlation coefficients, mean ranking scores, and standard deviations to identify clusters of subjectivity in participants' responses. Q-methodology is an adaptation of the R-methodology that may be more familiar to those who work with correlations analysis (Brown, 1980; van Exel and de Graaf, 2005). In R-methodology multiple comparisons are made between sets of data paired to individual cases to find those that vary together. Q-methodology rotates the focus of the analysis and searches for clusters of participants (even when there is only a low number) sharing the same profile of responses to many alternatives (Brown, 1980; van Exel and de Graaf, 2005).

Q-methodology provides a quantitative approach to qualitative analysis (Brown, 1980; van Exel and de Graaf, 2005). Qualitatively, it depends upon a concourse of statements offered to respondents to prioritise. This is analysed statistically to demonstrate subjective differences between groups of participants at a single point in time. In Tynan and McLain (2019) a bank of 42 statements was derived from national and local policies, programme training materials and guidance, and programme documentation. The programmes referred to were all offered by a single HE QTS provider in the northwest of England. These statements were selected to offer distinct choices to participants about standpoints concerned with the assessment of aspiring teachers whilst on school experience placement.

The analytical software (Schmolck, 2014) used factor analysis, mean priority scores and standard deviations to estimate the degree of agreement and disagreement on individual statements between groups displaying subjective differences. It was then possible to construct a qualitative profile for each group based upon the statements indicated in the consensus and disagreement categories and the scale of those disagreements. The potential for researcher subjectivity arises in this approach from the construction of the concourse of statements. Further, the method is an inductive process that cannot give any information about ideas and perspectives not included in the statements put forward for prioritisation. For Tynan and McLain (2019) the purpose was to investigate the possibility that subjective differences existed between mentors, tutors and trainees in the ITE partnership with respect to the assessment of trainees on school experience placement. For this purpose, it was sufficient to utilise a small self-selecting 'convenience sample' rather than a random or representative sample (Etikan, Musa and Alkassim, 2016). The profiles demonstrated were interesting, in that they resonated with the findings of the quantitative studies in the portfolio. This indicates that these findings may be more widely applicable than in that partnership only.

3.6 Summary

Small-scale practitioner research may be criticised for the difficulties posed when trying to infer general conclusions with more general utility from its findings (Shaw 2005). The preceding discussion surveys the methods used in the submitted portfolio papers and seeks to establish the rigour of the mixed methods employed within a framework consistent with a critical reality (Bhaskar, 1997) research perspective. Each was submitted and accepted for publication in a peer reviewed journal. This established that the methods were individually fit for immediate purpose. Adopting the critical realism perspective (Bhaskar, 1970) allowed the contribution of the papers to the whole portfolio to be integrated and inform wider questions about underlying structures and mechanisms associated with the assessment of professional performance. In this context, the findings and conclusions were valid locally but also have wider applicability to other areas nationally and internationally where the assessment of professional performance is based upon descriptions of competence.

4. Outcomes of the research portfolio

4.1 Professional Subject Knowledge and Subject Knowledge Enhancement

Inglis *et al.* (2013) reviewed ideas about subject knowledge for teachers as a key attribute impacting their teaching ability. This paper discussed SKE courses as one approach to filling vacancies for teacher training in shortage subjects such as the physical sciences. The journal article described the design and thinking behind one course developed at an HE QTS provider in the northwest of England. Other SKE providers offered a range of programmes with differing modes of delivery over various lengths of time. However, the SKE course discussed by Inglis *et al.* (2013) was attended, year-long, full-time, face-to-face and HE accredited, as this was considered most appropriate for developing science theory and science investigation subject knowledge. Participant observation and document analysis were used to review lessons learned from the early cohorts. Initial indications were that SKE trainees were positive about the course and were not hampered in any way by their lack of a chemistry or physics first degree during the PGCE year.

Tynan *et al.* (2014) analysed quantitatively the impact of the SKE course upon outcomes for students. The paper dealt with student perceptions reported through Likert style items from an end-of-programme feedback survey, observation feedback data, assessment grading data and early employment returns. This paper introduced statistical analysis to remove subjectivity from the conclusions reached about quantitative data. This confirmed rigorously and objectively observations by Inglis *et al.* (2013) that assessors and employers in the study had not perceived a difference between physics and chemistry graduate trainees and their SKE counterparts when grading their subject knowledge, teaching, or when offering jobs.

Tynan *et al.* (2016b) investigated qualitatively the attitudes and perceptions of SKE students at the end of their programme. SKE courses were introduced to address subject discipline knowledge issues but inevitably looked forward to the next stage, the ITE programme. As they looked forward to teacher training, most students accepted the continuing challenge involved in further expanding their professional subject knowledge, but they valued the time given to pedagogical considerations in the SKE programme. Perhaps predictably for students who had chosen to follow this style of SKE programme, most participants responded very positively to questions about the face-to-face components, and the support received from peers and tutors. They also expressed high levels of confidence that the SKE programme had prepared them well for the ITE programme to follow. This was congruent with national data on stakeholder perceptions of SKE programmes (Gibson *et al.*, 2013a, 2013b).

Tynan *et al.* (2016a) explored the nature of subject knowledge for science teachers through an investigation of a particular science pedagogy, the socio-economic approach (Cavagnetto, 2010), and possible sources of evidence that allow ITE students to demonstrate their subject knowledge. It did so in the context of encouraging science teachers to recognise the importance of their everyday informal practitioner research. Cavagnetto's (2010) earlier meta-study had indicated that the socio-economic approach was not the most effective pedagogy for developing science literacy and scientific

argumentation in secondary age learners. However, Tynan *et al.* (2016a) had modeled the approach as subject specific PCK (Shulman, 1986), appropriate for learning and teaching complex scientific concepts, and were surprised when the participating aspiring science teachers learned more about debate and assertion than scientific argumentation. The anonymous voting behaviour of the students during the activities and subsequent discussions suggested nuanced evidence about the pedagogy and its relationship to teacher subject knowledge (Tynan *et al*., 2016a) that demonstrated a complexity difficult to describe using only a four-point scale.

The Teacher's Standards (2011) not only describe teaching competencies for ITE programmes but for all teachers in England throughout their career. Pratt and Tynan (2019) further explored the theme of the complexity of professional subject knowledge. The paper indicated the way teachers develop beyond ITE and how they can demonstrate their professional subject knowledge through engagement with research literature and by conducting their own teacher-practitioner research. It discussed the transformational role of early completion of a master's qualification in education practice. It indicated, through their research interests, the areas of professional subject knowledge important to a group of newly and recently qualified teachers (N/RQTs). Their research project titles had centred most often on pedagogy, education psychology, assessment of learning and inclusion issues. Pratt and Tynan (2019) explored transformational learning and the early career development of teachers and in the process indicated the inadequacy of the Teacher's Standards (2011) when attempting to evaluate professional teacher subject knowledge.

The SKE related papers in the portfolio (Inglis *et al.*, 2013; Tynan *et al*., 2014, 2016a, 2016b) questioned whether SKE-route teachers would, in one year, gain sufficient breadth and depth of subject discipline knowledge to teach their SKE subject. They also questioned subject discipline knowledge as the key teacher attribute emphasized by central policy makers (Department for Education, 2010). The research associated with the SKE course for creating new physics and chemistry teachers appeared to answer these questions for the science SKE-route trainees at one HE provider. It indicated that an appropriate one-year HE Certificate course could render SKE PGCE science trainees indistinguishable to assessors and employers from those with first degrees in their specialist subject (Tynan *et al.*, 2014).

4.2 The factors influencing the assessment of teachers in training whilst on school experience placement

During the preparation of the SKE related papers (Inglis *et al.*, 2013; Tynan *et al*., 2014, 2016a, 2016b) it became clear that the findings regarding similar assessment outcomes for the SKE and specialist first degree cohorts (Tynan *et al.*, 2014) needed careful consideration. Several possible reasons for this key finding were investigated further by papers in this portfolio:

- The simplest interpretation of the findings from Tynan *et al.* (2014) was that, regardless of the assessors' perspective on teacher subject knowledge (Shulman 1986), the subject knowledge capability possessed by SKE route and specialist first-degree route trainees was the same. This seemed unlikely as the shorter time the SKE trainees had spent on their specialist teaching subject studies reduced the depth of coverage it was possible to achieve compared with a degree

course. However, it was also possible that degree subject specialists had more subject knowledge than was required to teach 11–16-year-old learners and both groups were judged to be equal with respect to the subject knowledge needed for secondary school teaching (Tynan *et al.*, 2014; Tynan and Jones, 2018).
- Another possibility was that ITE assessors had used different measurement tools (Figure 2) that did not discriminate sufficiently between levels of subject discipline knowledge (Tynan *et al.*, 2014; Tynan and Mallaburn, 2017; Tynan and Jones, 2018).
- It was possible that ITE assessors had perceived subject knowledge for teachers differently to the subject discipline knowledge assessed during specialist first degrees. According to Shulman's (1986) model different assessors could place different emphasis on the balance between SMCK, CK and PCK when trainees evidenced teacher subject knowledge (Tynan and Jones, 2018).
- A further possible factor became apparent when increased measures were adopted by the HE QTS provider to quality assure and manage the assessment of trainee teachers and QTS compliance issues. This may have reduced variation in the grading data (Tynan and Mallaburn, 2017; Tynan and Jones, 2016b) and guided assessment outcomes (Tynan and Jones, 2019; Tynan and McLain, 2019).

Tynan and Mallaburn (2017) investigated several statistical checks for monitoring consistency in assessment grades between partnerships across and within ITE programmes at a HE QTS provider. They recommended the use of Chi Squared Pearson's Correlation Coefficient. The paper demonstrated consistency across five ITE programmes in the frequency of different number grades ascribed for overall teaching in final summative assessments. Findings suggested that this could have been linked to HE QTS provider interventions across programmes to improve consistency in assessment practices, such as confirming the final summative teaching grade at a rigorous and structured meeting chaired by a HE liaison tutor. This was an early indication of the possible impact on assessors' ascribed grades of the QTS provider's IE (Smith, 2005), set out in is documentation, and actor network (ANT) (Fenwick and Edwards, 2010) exemplified by tutors' role in quality assuring assessment processes. Tummons (2010) had previously suggested IE and ANT as important in HE ITE assessments at another QTS provider in the northeast of England.

Tynan and Mallaburn (2017) found the correlations between grades ascribed for individual standards and overall teaching were almost always positive and statistically significant. This was entirely predictable if agreed partnership procedures were followed and a profile of grades for individual standards had been used to arrive at an overall grade for teaching. A negative correlation would have meant that the grades for a particular standard were low when the grades for overall teaching were high. To do this, assessors would have to think that performing well in that standard was detrimental to overall teaching performance and *vice versa*. A non-significant positive correlation would be an indication that the positive association was so small that there was a good chance that it was just random occurrence. During the monitoring process, any outcome other than a strong positive correlation between a teacher standard and overall teaching grade would indicate that assessors working on a particular ITE programme had possibly adopted a more holistic approach to assessment than the agreed partnership procedures. This is supported by Leshem and Bar-Hama (2008) who investigated the introduction of analytical assessment to ITE programmes in Israel,

noting the alternative approach of assessors who arrived at overall teaching grades holistically then ascribed grades for competencies that fitted with their overall judgement. Tynan and Mallaburn (2017) identified Chi Squared and Correlation Coefficients as suitable ways of monitoring consistency in assessment outcomes. Such consistency in assessment outcomes supported the idea that mentors had followed agreed partnership practices and helped evidence this as a performance criterion in the ITE partnership inspection framework (OfSTED, 2018).

For Chi Squared, a statistical conclusion that there was a difference between observed and expected counts indicated a lack of consistency. However, for correlation coefficients a significant correlation was associated with consistency. Tynan and Mallaburn (2017) were, therefore, able to quantify the probability of error for conclusions indicating consistency or inconsistency when they occurred. So, methodologically, the two statistical tests identified were useful because, used together, they can identify quantified Type 1 Errors (explained in Tynan *et al.*, 2014) when rejecting Null Hypotheses to demonstrate consistency or inconsistency in assessors' ascribed grades. The high level of consistency in assessment data demonstrated by Tynan and Mallaburn (2017) and Tynan and Jones (2018, 2019) was not predicted using the Hager and Butler model (1996) applied to ITE teaching by Martin and Cloke (2000) and mapped to an HE QTS provider ITE programme by Tynan and Jones (2018) (Figure 2). This suggested that there would be multiple opportunities for subjective differences between assessors applying the agreed assessment practices to the Teachers' Standards (2011) and OfSTED guidelines (OfSTED, 2018) that would result in more variability in outcomes than was observed.

Tynan and Jones (2018) reported the further use of the Chi Squared test of significance to look for subjective differences in the grading of different aspects of subject knowledge for teachers linked to Standards 3 and 4 and overall teaching grades based upon standards 1-8 (Department for Education, 2011). Chi squared had been identified as the most sensitive statistical test considered by Tynan and Mallaburn (2017). Tynan and Jones (2018) started to explore assessors' application of agreed partnership guidelines for arriving at grades for standards and overall teaching. These required grades for individual standards to be ascribed first and the profile obtained then used to arrive at a grade for overall teaching. The study compared grades ascribed for all subjects grouped together and English, mathematics and science separately. This utilized assessment data from one large ITE programme at the HE QTS provider. Although not a linear study, data from two consecutive years were analysed. Tynan and Jones (2018) found a high level of consistency for grades ascribed for standards 3, 4 and overall teaching but in mathematics and science there were some differences. In these instances, using Shulman's (1986) categories, assessors were more likely to give higher grades for the standard associated with SMCK and CK than for overall teaching based on standards 1-8 or the standard associated with pedagogy that included PCK. This indicated at least a slight tendency by science and mathematics school-based teacher educators to favour ability to teach their subject over other aspects of teacher subject knowledge when evaluating teaching ability.

For trainees teaching English, the grades for both aspects of subject knowledge were equivalent to grades for overall teaching over both years. However, where they did find subjectivity, Tynan and Jones (2018) indicated that teacher assessors in science and

mathematics were more likely to associate grades for pedagogy with their judgements of overall teaching ability than grades for subject and curriculum knowledge. This indicated that subjective differences were possible but not typical between assessors along subject specialism lines and that these appeared fluid and dynamic from one year to the next. Science teachers had displayed a tendency to favour pedagogy over subject content and curriculum knowledge as an indicator of overall teaching skill. This supports one of the alternative explanations the authors considered for the parity between SKE and subject specialist first-degree route trainees observed by Tynan *et al.* (2014). Assessors grading subject knowledge of SKE and first-degree subject specialists may have given similar grades to both based upon their pedagogical knowledge, which might be assumed to be generally similar. The effect observed by Tynan and Jones (2018) was too fleeting to establish this argument as a major or single influence, but it could be indicative of what a larger scale study might find.

Tynan and Jones (2019) used an r-family effect size metric (Ellis, 2010), POV expressed as a percentage, to explore the relative scale of the correlations between grades ascribed for individual standards and overall teaching. They used assessment data from a single large ITE programme over three consecutive years. Using the POV effect size metric as a percentage is intended to allow the reader to grasp the scale of associations intuitively (Ellis, 2010). The larger the percentage, the greater the shared variation between grades ascribed for a particular standard and the grades ascribed for overall teaching. This could be interpreted directly as the assessors associating higher grades for that standard more closely with higher grades for overall teaching. A consistent pattern in assessment grades emerged over three years. The standard most closely identified with subject knowledge and curriculum knowledge overlapped least with overall teaching grades (Tynan and Jones, 2019). At the same time, grades for three of the eight standards were associated most strongly with overall teaching grades. Following guidance from an OfSTED consultant, these were standards that mentors had been advised, during mentor training and quality assurance meetings, should be more indicative of overall teaching skill (Tynan and Jones, 2019). This suggested once again that the QTS provider's stance demonstrated by its IE (Smith, 2005) and implemented by its network of advocates (ANT) (Fenwick and Edwards, 2010) were a strong influence on grading, and this led Tynan and Jones (2019) to further question the use of number grades for valid assessments of competencies.

The final lead author paper (Tynan and McLain, 2019) used Q-methodology to look for subjective differences in attitudes to assessment amongst a small group of eight ITE participants at a secondary school. The sufficiency of this sample for the aim of this study was explained in **3.5 Q-methodology**. The method successfully identified a group of 4 and a pair of participants with significant differences in their profiles of responses. These profiles were not associated with the participants' role in ITE. Two mentors held dissimilar perspectives on the assessment of trainees, but each shared the same profile of responses with their own mentees. These instances would be consistent with a social model of professional learning (Wenger, 1998) operating so that mentors and mentees working closely together adopted similar perspectives on assessment. Two other respondents fitted neither profile nor formed a third cluster of subjectivity. However, Tynan and McLain (2019) found that, despite subjective differences in their profiles, the group and pair also shared a consensus on the importance of statements relating to QTS compliance issues and following agreed partnership guidelines when

assessing teaching. This suggested that, beyond agreeing with the need to follow the procedures necessary to recommend trainees for QTS, that subjective differences between assessors were easily demonstrated. This was implied by Martin and Cloke's (2000) application of Hager and Butler's (1996) model for professional development and assessment, but not apparent in assessment grade outcomes (Tynan and Mallaburn, 2017; Tynan and Jones, 2018).

One profile of responses (Tynan and McLain, 2019) emphasised the importance of the individual standards and OfSTED criteria almost exclusively. Assessment was perceived as a matter for mentors and mentees that only other teaching practitioner peers could usefully help with. The other group of respondents perceived assessment as a process, emphasising fairness and a greater degree of agreement with quality assurance activities originating within and externally to the school. They were also more in favour of considering evidence from other sources such as alternate school placements and even university training days when arriving at grades. The study established that subjectivity concerning assessment could be demonstrated qualitatively together with a strong consensus about the need for compliance with QTS requirements and agreed assessment practices (Tynan and McLain, 2019).

4.3 Summary

Parity between outcomes for SKE and subject specialist first-degree participants in a secondary ITE programme was observed by Inglis *et al.*, 2013 and Tynan *et al.*, 2014. Further, SKE ITE trainees expressed great confidence in their level of preparation to teach their SKE subjects (Inglis *et al.*, 2013; Tynan *et al.*, 2014; Tynan *et al.*, 2016b). Pragmatically the SKE courses to create new physical science teachers had worked but it seemed unlikely that there was an explanation for this as simple as a one-year SKE course successfully providing equal subject discipline knowledge as a specialist first degree. Tynan and Mallaburn (2017), Tynan and Jones (2018, 2019) and Tynan and McLain (2019) investigated the factors affecting the competency-based assessment of trainees on school experience placement and highlighted assessment issues associated with the complexity of subject knowledge for teachers (Tynan *et al.,* 2016a; Pratt and Tynan, 2019). These called into question the validity of assessment of subject knowledge and other teacher competencies when using descriptors and number grades (Tynan and Mallaburn, 2017; Tynan and Jones, 2018; Tynan and Jones 2019; Tynan and McLain, 2019). Assessments might be influenced by many factors including personal subjectivities of assessors (Tynan and McLain, 2019), that could be linked to their subject specialism (Tynan and Jones, 2018) but this was not evident in assessment outcomes expressed as number grades. These showed high levels of consistency between and within the ITE progammes investigated (Tynan and Mallaburn, 2018; Tynan and Jones, 2018). There was a strong indication that these had been influenced by quality assurance and management of assessment processes (Tynan and Jones, 2019) and this was supported by the emphasis on compliance despite their personal subjectivities demonstrated within a small group of ITE participants (Tynan and McLain, 2019).

5. Discussion of a possible underlying mechanism indicated by the portfolio

5.1 Adopting a critical realism approach

The critical reality perspective (Bhaskar, 1997) has been applied to a portfolio of nine papers with separate and individual purposes and research methods. Downward and Mearman (2006) argued that mixed methods triangulation could be understood in terms of retroduction and critical realism. Downward and Mearman (2006) accepted and discussed Sayer's (1992) definition of retroduction as a type of inference that is neither deductive or inductive but seeks to explain events by identifying and exploring underlying mechanisms that could produce them. There was reason to believe that a search for an underlying structure or mechanism (Bhaskar, 1997) amongst the portfolio's theoretical explanations and findings could be productive.

The process for this was described and discussed by Bygstad and Munkvold (2011, p5) under these headings:

1. Description of events
2. Identification of components
3. Theoretical redescription (abduction)
4. Retroduction: Identification of candidate mechanisms
5. Analysis of selected mechanisms and outcomes
6. Validation of explanatory power

The first three stages are already described by the papers in the portfolio and the previous sections **1-4**. These discuss the phenomena observed that were associated with themes that run through the portfolio of papers, together with the models found useful in trying to explain them. Adoption of a critical realism approach (Bhaskar, 1997) aligned the themes developed in the nine portfolio papers. Both quantitative and qualitative data contributed to separate findings that can be interpreted as observed phenomena and represent new knowledge in Bhaskar's (1997) empirical domain. These were discussed in relation to pre-existing theoretical models relevant to the phenomena during the process of academic research and publication. This describes what was detectable or measurable during investigations and their publication, which is affected by personal and social psychology and is the transitive portion of the resultant new knowledge (Bhaskar, 1997). These phenomena were all associated with an elusive process - the assessment of trainee teachers' subject knowledge, other teaching skills and overall teaching ability. That assessment was necessary for the recommendation of QTS in England (Department for Education 2011), but the phenomena described in the portfolio illustrated the difficulty in demonstrating the factors contributing to this process. Some may be beyond measurement, and this identifies assessment as an event in Bhaskar's (1997) actual domain (Figure 3).

The findings described phenomena about an event that indicated the possibility of an underlying structure or mechanism, the intransitive portion of Bhaskar's (1997) real domain. Structures and mechanisms have enduring attributes and the potential to affect

phenomena, even if not exercised (Zachariadis, Scott and Barrett, 2013). Bygstad and Munkvold (2011) described them as invisible, interacting in open systems, and highly contextual. Therefore, Bygstad and Munkvold (2011) argued that mechanisms and structures should be considered useful in hindsight for explaining phenomena, rather than using them to predict future effects. The following discussion summarises the search for any underlying structures or mechanisms indicated by the portfolio of papers based upon Downward and Mearman's (2006) and Bygstad and Munkvold's (2011) approaches.

5.2 Professional Subject Knowledge and Subject Knowledge Enhancement

The phenomena considered by Tynan *et al.* (2014; 2016a; 2016b) were complexity of subject knowledge, parity of outcomes and confidence to teach (Figure 3). Explanations for these were suggested by considering Hager and Butlers' (1996) model of professional development and assessment applied to education by Martin and Cloke (2000) and mapped to an ITE programme by Tynan and Jones (2018). This indicated a qualitative judgemental approach to assessment in the later stages of ITE programmes (Figure 2). However, the use of criteria and grades when assessing trainees would suggest to the casual observer the continuance of an analytical scientific measurement approach more often associated with undergraduate degree courses. Furthermore, it could be argued that this has been encouraged by frameworks for the inspection of ITE partnerships (OfSTED, 2018) that have held QTS providers and partnerships accountable (often with consequences) for the numbers of NQTs working at levels above the minima described in the Teachers' Standards (Department for Education, 2011). It was not possible to know for certain the individual mental processes followed by assessors or their students during the two-year programme and its assessment. This identifies the assessment of trainees in the context described in the portfolio, as an event in Bhaskar's (1997) actual domain (Figure 3).

Tynan and Jones (2018) demonstrated that subject specific subjectivity in assessment grades was possible. For science and mathematics mentors, pedagogy could be more important than specialist subject content knowledge when considering overall teaching ability. Further, Tynan and Jones (2019) reported that across subjects at an HE QTS provider, grades for subject content knowledge were least likely to be correlated with grades for overall teaching consistently over several independent cohorts. Some SKE students expressed awareness of the limitations of their subject knowledge going forward into their ITE programme, yet generally they expressed their confidence to teach their SKE subjects. They were already aware of the importance of pedagogical knowledge for teachers (Inglis *et al.*, 2013; Tynan *et al.*, 2014; 2016b) and favoured more emphasis on this aspect of teacher subject knowledge. This suggests a weaker association between specialist content knowledge in the minds of aspiring teachers and their assessors than policy makers might have intended and an additional influence on assessment not linked to guidance and direction. However, the SKE programme was found to be effective in creating new physical science teachers although the portfolio suggests a social process of accepting them into a community of practice (Wenger, 1998) in addition to simple grading outcomes.

5.3 The factors influencing the assessment of teachers in training whilst on school experience placement

The portfolio also evidenced observable phenomena associated with competency based, graded assessment of trainee teachers. These can be listed as compliance, adherence, consistency, subjectivity, IE and ANT (Figure 3). The portfolio demonstrated assessors' compliance with OfSTED QTS requirements and adherence to agreed partnership assessment and quality assurance practices quantitatively through the high levels of consistency found in assessment grades across and within ITE programmes over time (Tynan and Mallaburn, 2017; Tynan and Jones, 2018). Tynan and Jones (2018; 2019) provided some quantitative evidence that assessors' grades were influenced over time by IE and ANT. Qualitative data collected (Tynan and McLain, 2019) were congruent with this and demonstrated, through the consensus amongst a small group of ITE participants, the importance of compliance and adherence.

Tynan and Jones (2018) demonstrated that subjectivity was also possible between ITE assessors in some core subjects concerning the emphasis they placed on different aspects of subject knowledge for teachers described in the Teachers' Standards (Department for Education, 2011) when considering overall teaching ability. Furthermore, Tynan and McLain (2019) demonstrated that ITE participants expressed significant subjective differences concerning their assessment priorities despite sharing consensus on compliance and adherence issues. Although this might be expected to work antagonistically to compliance and adherence, it did not appear to reduce the level of consistency observed in graded assessment data (Tynan and Mallaburn, 2017; Tynan and Jones, 2018). The influence of the QTS provider's stance on assessment and grading expressed through its documentation (IE) and the effectiveness of its network of teacher educators supporting this (ANT) were considered as joint candidates for an underlying mechanism to explain this. However, although they were not straightforward to demonstrate, it could be argued that these were an observable influence on assessment (Tynan and Jones, 2019; Tynan and McLain, 2019). It is likely that the portfolio findings are suggesting an underlying mechanism for assessment that allows for compliance and adherence even when working antagonistically to reflect assessors' true grading decisions.

5.4 Good teachers

The location of a mechanism to explain all the phenomena demonstrated within the portfolio may lie within the partnership mentors' own ITE practitioner discourse on good teachers and their attitude towards competency-based assessment involving the grading of aspiring teachers. Teacher educators would have a variety of potential sources to draw on in current and less recent literature (Wenger, 1998; Door, 2004; Moore, 2004; Andrews, Richards and Templin, 2012; Jones and White, 2014; Caikler and Wood, 2016; Clarke, 2016; Korthagen, 2017; Czerniawski, 2018). The portfolio indicates that ITE practitioners and trainees complied with, but did not always share, central policy makers' assumptions about subject discipline knowledge as a key indicator of good teachers (Inglis *et al.*, 2013; Tynan *et al.*, 2014; Tynan and Jones, 2018, Tynan and Jones, 2019; Tynan and McLain, 2019; Pratt and Tynan, 2019). Most strikingly, over three consecutive years, the teaching standard (Department for Education, 2011) least

strongly associated with overall teaching ability was the standard most associated with subject discipline knowledge. In addition, Pratt and Tynan (2019) indicated that, for one year, approximately three fifths of education practice master's degree research projects (n=38) conducted by early career teachers focused on pedagogy, its contributary aspects, assessment or inclusion. Teacher educators, trainees and newly qualified teachers all appear to be aware of the issues associated with developing subject discipline knowledge in the absence of other aspects of professional subject knowledge.

This portfolio of work suggests that a model of social learning involving acceptance into a community of practitioners (Wenger, 1998) would be one recognised by many of the participants in ITE during these studies. It is consistent with the respondents' views, recorded by Tynan and McLain (2019), that the mechanism affecting assessment and causing the phenomena observed could lie in the professional identity of the teachers mentoring the trainees as individuals within a community of teaching and teacher educator practitioners (Door, 2004; Czerniawski, 2018). It is conceivable that assessment decisions could have been reached holistically then performance descriptors for competencies used analytically to justify and support those decisions (Leshem and Bar-Hama, 2008).

Practitioners in the studies would then need to balance many considerations to arrive at an overall judgement of a trainee's suitability to teach. They would need to comply with QTS requirements and adhere to agreed partnership practices. They would also need to merge analytical approaches based upon performance criteria and more holistic judgements based upon other undefined criteria. The success of this depends upon the shared norms of professional standards in teaching at the training and recruitment phase. I would argue that assessment decisions would be influenced by underlying psychological structures and reflect an individual assessor's experiences of teaching and teacher education. Simultaneously, this could be influenced indirectly by social psychological and sociological mechanisms. In the context of the assessment approach described in the portfolio studies, this could be indirectly through the influence of general expectations placed upon assessors by other educators, teacher educators, stakeholders and society in general concerning teachers, and/or directly through interaction with the QTS provider, OfSTED training and their quality assurance processes.

Arguably, the Teachers' Standards (Department for Education, 2011) were intended to replace holistic professional judgements by assessors with a more transparent, analytical and accountable approach. However, it could be argued that this portfolio demonstrated the difficulty of achieving this aim particularly when the assessment of complex teaching skills is reduced to categories and grades to satisfy the purpose of accountability. A variety of influences on assessment are described by the phenomena indicated in the portfolio papers. A candidate for an intransitive underlying mechanism (Bhaskar, 1997) should be highly contextual, invisible and operating in an open system (Downward and Maerman, 2006; Bygstad and Munkvold, 2011) and, above all, capable of producing and therefore explaining the phenomena described. I propose the term *conflicted role, professional judgement* (Figure 3) satisfies these criteria and covers the characteristics of a mechanism that can be retroduced from these phenomena.

Mentors act simultaneously as teachers and teacher educators and both roles are integrated to varying degrees in different schools. In England, teachers and teachers acting as teacher educators are also currently charged with the task of recognising good teachers. Teacher educators in England must comply with QTS requirements and adhere to agreed practices. Logic dictates that, sometimes, the teacher, teacher educator and assessor roles will conflict during the process of trainee assessment. Andrew, Richards and Templin (2012) explored the role conflict experienced by most physical education teachers who also acted simultaneously as sports coaches and suggested that often one role (often coaching) dominated the other. When faced with the conflicting requirements of formative feedback during mentoring and summative assessment for accountability, it can be assumed that the teachers in the portfolio studies performed their duties with professional regard to making a fair and accurate evaluation of the candidate for the teaching profession. They balanced the needs of all the stakeholders for fairness, accuracy and accountability. Zsargo and Palmer (2019) considered factors involved in the conflicted role for teachers acting as ITE mentors assessing trainees in primary school in England and theorised a 'mentor assessment identity' (p79) that guided judgements. This resonates with the idea of the underlying biological and psychological mechanism suggested by this portfolio. *Conflicted role, professional judgement*, based upon self-perceptions of status and professionalism would drive assessment decisions in this context (Figure 3) and recognises the dual role of mentors as teachers and teacher educators.

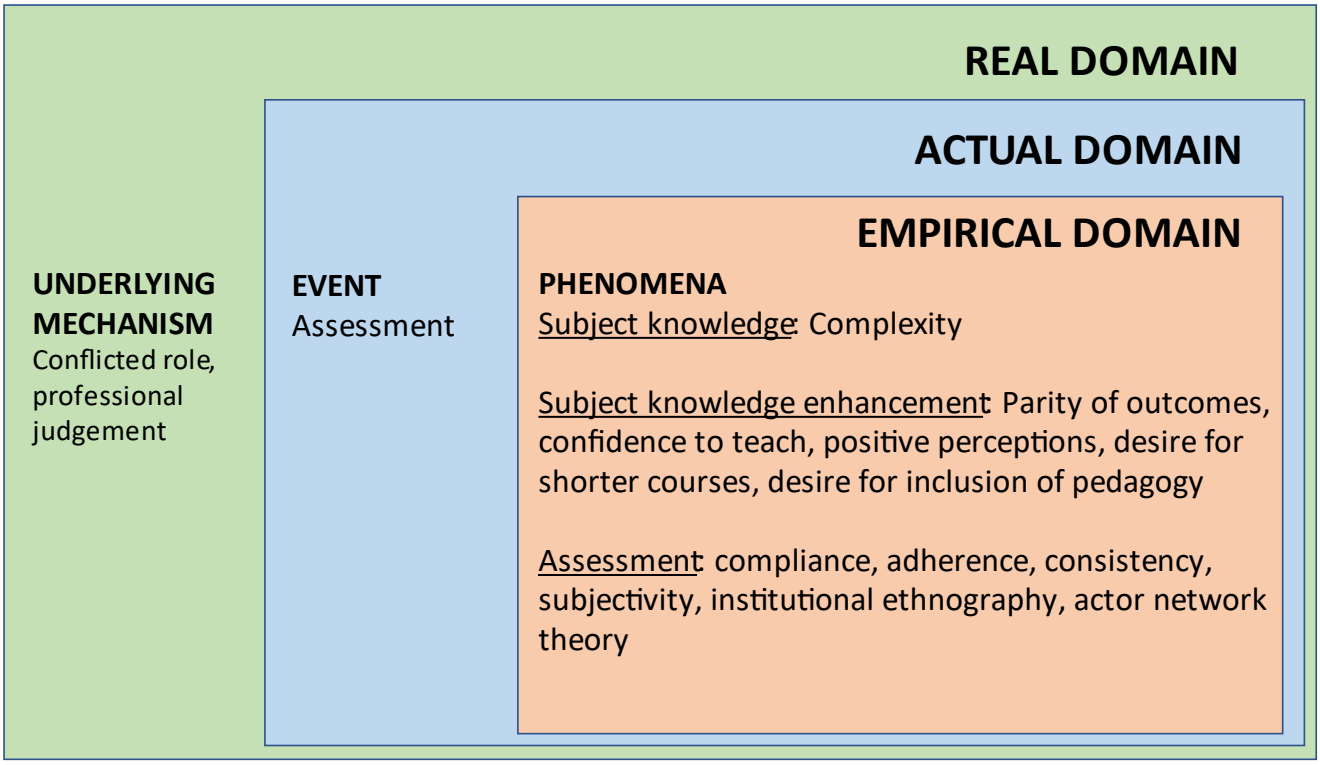


Figure 3 Proposed critical realism perspective (Bhaskar, 1997) on the assessment of trainee teachers on school experience placement.

6.1 Conclusion and Recommendations

6.1 Research Question i: How successful was a particular model for delivering subject knowledge enhancement (SKE) courses for non-subject specialist graduates in creating new physical science teachers?

Evaluation studies of a year-long HE accredited SKE course for creating new chemistry and physics teachers indicated parity in student outcomes between SKE physics and chemistry teachers and those with a first degree in these subjects. This included equal confidence to teach, teacher educator assessments of trainee subject knowledge and early employment rates (Inglis *et al.*, 2013; Tynan *et al.*, 2014; Tynan *et al.*, 2016b). The participants in Inglis *et al.* (2013), Tynan *et al.* (2014), Tynan *et al.* (2016a) and Tynan *et al.* (2016b) were a large proportion of teachers trained in the NW of England and all of those trained at a large HE provider. The findings added to the debate concerning SKE courses and were used by the Royal Society of Chemistry when considering accrediting SKE courses in Chemistry.

The government sponsored study published around the same time (Gibson *et al.*, 2013a, 2013b) was a large scale but self-selecting survey of perceptions and opinions about SKE course provision. The local studies in this portfolio considered the outcomes for students as well as perceptions of SKE-route and specialist first degree route PGCE trainees. The qualitative findings reflected the findings of the national survey (Gibson *et al.*, 2013a, 2013b) including an expressed preference by some students for shorter distance learning SKE courses reported in *Tynan et al.* (2016b). The government briefly suspended funding for SKE courses in 2013-14 before brokering their re-instatement in shorter and often more distance learning formats (Tynan *et al*., 2016b).

SKE courses were perceived as successful by stakeholders and SKE route teachers were not perceived as disadvantaged by taking them. However, it is unlikely that they could provide equivalent subject discipline knowledge to an undergraduate degree programme. Confidence may have been an important factor as SKE route trainee teachers engaged in self-evaluation, sought to convince assessors in school that they had sufficient subject knowledge and competed for science teacher jobs. Turner-Bisset (2006) and Korthagen (2017) have argued for the inclusion of teacher self-knowledge in the description of professional subject knowledge, and the ability to self-evaluate would contribute to this confidence. Korthagen (2017) called for teacher competencies to be located amongst other important teacher attributes when planning teacher training and assessment (Figure 2) to acknowledge their importance. Cajkler and Wood (2016) expressed this in terms of a more literate use of standards descriptors. Further research is needed on the effectiveness of shorter and distance learning SKE courses particularly concerning the outcomes for students including their longer-term progress in the teaching profession. However, this portfolio supports the pragmatic case for year-long, face-to-face SKE programmes and the inclusion of other aspects of professional subject knowledge for teachers in addition to subject discipline knowledge.

Recommendation 1: *SKE programs should continue to be resourced and funded as an important strategy in contributing to teacher recruitment in shortage subjects e.g. physics and chemistry.*

Recommendation 2: *In certain subjects there is a strong case for longer on-site programmes and the inclusion of PCK and CK in the SKE programme content.*

6.2 Research Question ii: What were the factors influencing the assessment of teachers in training whilst on school experience placement?

The quantitative papers (Tynan *et al.*, 2014; Tynan *et al.*, 2016a; Tynan and Mallaburn, 2017; Tynan and Jones, 2018; Tynan and Jones, 2019) and qualitative papers (Inglis *et al.*, 2013; Tynan *et al.*, 2016b; Tynan and McLain, 2019 and Pratt and Tynan, 2019) demonstrated the usefulness of statistical analysis, Q-Methodology and thematic analysis in contributing to mixed methods investigations.

The portfolio demonstrates the complexity of subject knowledge and other teacher competencies for trainee teachers, and their assessment. It also identifies potentially negative influences on criteria-based assessment of teachers in training when assessment has a dual function and is used for accountability and quality control measures as well as describing teachers' capabilities.

Number grading against performance criteria may have been useful to trainees and mentors in the formative analysis of teaching to develop subject knowledge and other competencies. However, the quantitative studies Tynan and Mallaburn (2017), Tynan and Jones (2018) and Tynan and Jones (2019) together with Tynan and McLain (2019) suggested that, despite subjective differences between assessors, the need to comply with QTS requirements and partnership agreed assessment procedures was an over-riding factor in ascribing number grades. This resulted in a high level of consistency in the assessment grades ascribed that was easily demonstrated statistically and a pattern in the data over time that was congruent with highly effective IE and ANT (Tynan and Jones, 2019; Tynan and McLain, 2019).

Recommendation 3: *Assessment for accountability should be separated from assessment for professional development.*

Recommendation 4: *Assessment against performance criteria should be accepted as a qualitative and judgemental process.*

Recommendation 5: *The Teachers' Standards (2011) describe minimum performance levels. Assessment against this type of performance criteria should be more nuanced and detailed than simple categories or grades.*

6.3 Good teachers

Discourses on good teachers differ in the emphasis they place on the innate ability of teachers, their ability to learn from training and their ability to learn from experience and reflection (Moore, 2004). Central education policy makers in England have placed a

high level of importance upon subject knowledge derived from undergraduate degrees (Department for Education, 2010) as a key attribute of good teachers. Whatever attributes are assumed make a successful teacher, to be recommended for QTS, trainee teachers must convince mentors that they have evidenced eight trainable teaching competencies and several non-negotiable professional attributes described by the Teachers' Standards in England (Department for Education, 2011). Tynan *et al.* (2014) indicated that NQTs in physical sciences with first degrees and SKE qualifications in a physical science subject were perceived similarly. Perhaps this was because teacher attributes other than subject discipline knowledge were more important to assessors and employers as indicators of teacher worth. Teachers and teacher educator practitioners have questioned the dominance of teaching competencies, their ability to describe teachers and their effect on ITE programmes (Turner-Bissett, 2006; Cajkler and Wood, 2016; Korthagen, 2017). This portfolio supports the idea that the ITE practitioners involved in these local studies assessing subject knowledge, other competencies and overall teaching ability of trainees can adopt a different assessment stance to central policy makers whilst complying with the current system for the recommendation of QTS.

This portfolio of papers highlights many reasons to critically consider the use of performance-criteria based assessment of teaching competencies for summative judgements. There is a stronger case (Cajkler and Wood, 2016) for the formative use of such approaches, typified by the Teachers' Standards (2011), to help trainees and mentors analyse and give feedback on lessons. The use of number grades to categorise teachers cannot be recommended, particularly when central guidance is available only for the pass/fail boundary. A recent critique of performance criteria based ITE systems by Korthagen (2017) does not seek to remove competencies from the description of teaching but to place them amongst other important teacher attributes, such as teacher identity and core beliefs (Figure 2), that are currently overshadowed by the need to provide evidence of competence. This portfolio supports the position that teachers cannot be described simply as the sum of their profile of competencies. Initiatives by authors such as Cajkler and Wood (2016) and Korthagen (2017) suggest the possibility of incorporating more meaningful ways of assessing teacher worth into ways of preparing aspiring teachers that also give emphasis to their core values and self-knowledge.

Recommendation 6: *ITE programmes should address the complexities of professional learning and its assessment. This is difficult to achieve whilst the current discourse on good teachers in England comprises a restrictive list of competencies and the evidence that a minimum performance level has been achieved.*

6.4 Conflicted role, professional judgement

Figure 3 proposes my model to explain the findings and conclusions of the body of research represented by the portfolio of papers submitted. This is based upon the three domains described by Bhaskar's (1997) critical realism perspective, diagrammatically showing the relationship between them. In Bhaskar's (1997) view of ontology and epistemology, invisible intransitive underlying structures and mechanisms exist

independently of researchers in the real domain. These have enduring attributes and influence events. Events occur in the actual domain where they may be, at least, partially observed or measured. Phenomena are fully detectable to researchers and can be the subject of science and other forms of investigation to generate knowledge in the empirical domain. Figure 3 summarises the numerous phenomena associated with assessment as a specific event These are reported in section **4. Outcomes of the research portfolio** and discussed in section **5. Discussion of a possible underlying mechanism indicated by the portfolio.** The model names an underlying mechanism, *conflicted role, professional judgement* that has been postulated and identified according to the approaches advocated in the literature (Archer et al, 1998; Downward and Maerman, 2006; Bygstad and Munkvold, 2011; Zachariadis, Scott and Barrett, 2013; Aliyu *et al.*, 2015; and Kivunja and Kuyini, 2017).

Recommendation 7: *Although Bygstad and Munkvold (2017) suggested that underlying mechanisms were context specific and, therefore, more useful for explanation than prediction. However, the model proposed in Figure 3 highlights demonstrable phenomena associated with the assessment of professional learning as an event. These and the underlying mechanism suggested have wider application to systems of professional learning and development, especially if they rely on mentors who also have a major role as assessors. As such it is worthy of further consideration.*

6.5 Suggested further research

The portfolio of research raised many new questions. Arising from the early research, what has been the impact of shorter on-line SKE programmes on outcomes for students and trainee teachers? Associated with this, what has been the impact of shorter on-line SKE programmes on student and trainee teacher perceptions and attitudes towards their ITE programmes?

The later papers highlight the need for further research into the impact of different assessment strategies on the design and conduct of ITE programmes. Further, what is the effect of on teaching and learning in schools of describing teachers solely in terms of competencies whilst ignoring the impact of teaching environments and teacher attributes such a teacher self-knowledge and their underpinning values?

Finally, considering the underlying mechanism identified by this thesis, the next area for study would be the impact of role conflict on mentors practicing both as teachers and teacher educators with a mentoring role that includes assessment.

References

Aliyu, A. A, Singhry, I. M., Adamu, H. and Abubakar M. M. (2015) Ontology, epistemology and axiology in quantitative and qualitative research: elucidation of the research philosophical misconception, *Proceedings of The Academic Conference: Mediterranean Publications and Research International on New Direction and Uncommon*, (2), 1

Andrew, K., Richards, R. and Templin, T. J. (2012) Toward a multidimensional perspective on teacher-coach role conflict, *Quest*, 64, pp 164-176

Archer, M., Bhaskar, R., Collier, A., Lawson, T. and Norrie, A. (1998) *Critical Realism: Essential Readings*, London: Routledge

Bacchetti, P. (2002) Peer review of statistics in medical research: the other problem, *BMJ*, 324(7348), pp 1271–1273

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

Bernard, H. R. and Ryan, G. W. (2010) *Analysis of Qualitative Data, Systematic Approaches*, SAGE Publishing Inc.

Bhaskar, R. (1997) *A Realist Theory of Science*, London: Verso

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

Bygstad, B. and Munkvold, B. E. (2011) In search of mechanisms. Conducting a critical realist data analysis, *Proceedings of the International Conference on Information Systems*, 7, Association for Information Systems
Available at https://aisel.aisnet.org/icis2011/proceedings/researchmethods/7

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

Clarke, L. (2016) *Teacher Status and Professional Learning*, Critical Publishing
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), pp 336–371.

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

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

Danermark, B, Ekström, M. and Karlsson, J. Ch. (2019) *Explaining Society: Critical Realism in the Social Sciences (2nd Edition)*, Routledge: London and New York

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

Door, V. (2004) *Developing creative and critical education practitioners*, St Albans: Critical Publishing Ltd

Downward, P. and Mearman, A. (2006) Retroduction as mixed-methods triangulation in economic research: reorienting economics into social science, *Cambridge Journal of Economics*, 2006, pp 1-23

Drake, P. and Heath, L. (2011) *Practitioner Research at Doctoral Level: Developing coherent research methodologies*, London and New York: Routledge

Edmondson, D. (2005) Likert scales: A history, *CHARM 2005*, pp127-133

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

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)

Fenwick, T. and Edwards, R.C. (2010) *Actor Network Theory in Education*, London and New York: Routledge

Gaito, J (1980) Measurement Scales and Statistics: Resurgence of an Old Misconception, Psychological Bulletin, 87(3), pp 564-567

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) *Technical Report, Analysis of Survey Data -2011-12: Research Report*, Department for Education/CooperGibson Research

Goodfellow, J. (2005) Researching with/for whom? Stepping in and out of practitioner research, *Australian Journal of Early Childhood*, 30(4), pp 48-57

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

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

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), pp 101-107

Jones. K. and White, E. (2014) *Developing Outstanding Practice in School-based Teacher Education*, Critical Publishing

Kivunja, C. and Kuyini, A.B. (2017) Understanding and Applying Research Paradigms in Educational Contexts, *International Journal of Higher Education*, 6 (5), pp 26-41

Korthagen, F. (2004) In search of the essence of a good teacher: Towards a more holistic approach in teacher education, *Teaching and Teaching Education*, 20, pp77-97

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

Kuhn, T. (1970) *The Structure of Scientific Revolutions, 2nd Edition*, University of Chicago Press

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

Levy, J. S. (2015) Counterfactuals, Causal Analysis, and Historical Analysis, *Security Studies*, 24(3), pp 378-402

Leshem, S. and Bar-Hama, R. (2008) Evaluating Teaching Practice, *ELT Journal*, 62(3), pp 257-265

Lindley, D.V. and Miller, J.C.B. (1053) *Cambridge Elementary Statistical Tables*, 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

Morgan, D. (2017) *Integrating Qualitative and Quantitative Methods: A Pragmatic Approach*, SAGE Publishing Inc.

Moore, A. (2004) *The Good Teacher-Dominant Discourses in Teacher Education*, London and New York: Routledge

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

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

Pearl, J. (2009) Causal inference in statistics: An Overview, *Statistics Surveys*, (3) pp 96-146

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

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

Robinson, S. (2006) Teacher training in England and Wales: Past, Present and Future Perspectives, *Education Research and Perspectives*, 33(2), pp19-36

Saldaña, J. (2009) *The coding manual for qualitative researchers*, CA: SAGE Publishing Inc.

Sayer, A. (1992) *Method in Social Science: A realist approach*, CA: SAGE Publishing Inc.

Shaw, I. (2005) Practitioner Research: Evidence or critique? *British Journal of Social Work*, 35, pp 1231-1248

Schmolk, P. (2014) *PQ Method (version 2.35)*, [Computer program]

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

Smith, D. E. (2005) *Institutional Ethnography: A sociology for people*, Altamira Press

TDA (2007) *Professional Standards for Teachers*, Training and Development Agency for Schools/gov.uk

Tummons, J. (2010) Institutional ethnography and actor network theory: a framework for researching the assessment of trainee teachers, *Ethnography and Education*, 5(3), pp345-357

Turner-Bisset, R. (2006) The knowledge bases 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, *Teacher Education Advancement Network 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? *Teacher Education Advancement Network Journal*, 11(1), pp 4-14

Tynan, R., Jones, R.B., Mallaburn, A. and Clays, K. (2016a) Working towards evidence-based practice in science teaching and learning, *School Science Review*, 97 (361), pp 109-115

Tynan, R., Jones, R.B., Mallaburn, A. and Clays, K. (2016b) Subject knowledge enhancement courses for creating new chemistry and physics teachers: the students' perceptions, *School Science Review*, 98 (363), pp 109-114

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

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

Tynan, R. and McLain, M. (2019) Attitudes to assessing trainee teachers on school experience placement within a group involved in an initial teacher education partnership at an 11-16 academy: a Q-methodology approach, *Teacher Education Advancement Network Journal*, 11(4), pp 3-20

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

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

Zachariadis, M., Scott, S. and Barrett, M. (2013) Methodological implications of critical realism for mixed-methods research, *MIS Quarterly*, 37(3), pp 855-879

Zsargo, E. and Palmer, J. (2019) Common understanding or 'hodgepodge'? The consistency and accuracy of school based mentors' assessment of trainee primary teachers in England, Teacher Education Advancement Network Journal, 11(2), pp72-81

Appendix

The Methods and Results sections of a designated paper converted to American Psychological Association Publication Style and Standards (American Psychological Association, 2020):

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

Methods Section

Design

A comparative survey of parametric and non-parametric statistical methods with graded assessment data was conducted to identify approaches to monitoring consistency across and within five Initial Teacher Education (ITE) programmes at a Higher Education Institute (HEI), Qualified Teacher Status (QTS) provider. This also constituted a single, action research cycle (Burton & Bartlett, 2009) that aimed to evaluate the impact of recently agreed assessment practices. It was assumed that consistency in number graded assessments between the constituent ITE partnerships reflected consistency in assessment practices.

Participants

Assessment data was analysed from five ITE programmes: Primary Post Graduate Certificate in Education (PGCE), Secondary PGCE, Secondary Salaried School Direct, Primary Education Honours degree with QTS (3 Year) and Primary Education Honours degree with QTS (4 Year). The Secondary PGCE cohort included trainees following the Non-Salaried School Direct route.

All those involved in the number graded assessment of trainee teachers on school experience placement for the ITE partnerships within these programmes during the academic year 2014-15 were participants in the study. Subject mentors were ultimately responsible for assessments on school experience placements. However, trainees were able to discuss their evidence for meeting the Teachers' Standards (Department for Education, 2011). Also, assessment

practices and outcomes were subject to quality assurance, internally by professional mentors and externally by visiting mentors and tutors from the HEI, QTS provider.

The number of trainees following each programme varied according to their recruitment quotas and final uptake by applicants, and varied during the study period as trainees left, took leave of absence, or returned to studies (see results.) However, in all cases the data used for statistical analysis constituted the full population of assessment grades collated at each review point. Typically, subject mentors worked with one trainee, but professional mentors and university liaison tutors worked with varying numbers of subject mentors and trainees.

The study analysed data gathered naturally during the normal course of the ITE programmes and reported anonymous findings in a fully analysed format. It was impossible to identify any institution or individual participant from the collated data. No personal or demographic information was collected. As such, the HEI provider categorised this as a pedagogical study with minimal ethical risk. However, as the authors intended to publish the findings outside the institution, the study was submitted for proportionate consideration and approved (15/EHC/083) by a Research Ethics Committee. Informed consent to analyse the assessment data and report findings was given by the appropriate HEI gatekeeper.

Materials

The researchers were given access to number graded assessment data for five ITE programmes at four formal review points during 2014-15. The grades were collated onto a spreadsheet by programme administrators.

Blank Microsoft 2010 Excel Spreadsheets were used to conduct and compare the results from different statistical tests using pre-existing spreadsheet formulae or cell formulae constructed from first principles.

Procedure

Aiming to maximise consistency of assessment outcomes and practice, several agreed partnership practices were operating when the 2014-15 assessment data was gathered. Attendance and participation in Mentor Training had been increased by including a training element in all HEI provider liaison visits and supplementing the HEI provider programme of training meetings by on-site training in partner schools. All assessments were referenced to the minimum performance descriptors set down in the Teachers' Standards (Department for Education, 2011) and criteria set down in the evidence tracking document completed by trainees during their programme. The rigour of the final assessment process was secured by adopting a longer, more structured triangulation meeting chaired by a HEI Tutor.

The assessment data analysed statistically were drawn from five programmes, at three formative review points and the final summative assessment during the academic year 2014-15.

Analytic Plan

The trial included three ways of analysing overall teaching grades at different review points across programmes.

The parametric statistical analyses used to compare overall teaching grades across programmes and assessment points were the visual presentation of mean grades and their 95% confidence limits and one way analysis of variance (ANOVA) on counts for grades across programmes. These were compared with the non-parametric Chi squared analysis on counts for grades across programmes. A parametric (Pearson's) and non-parametric (Spearman's) correlation coefficient were used to compare grades ascribed for individual standard grades with overall teaching grades.

Pre-existing Microsoft 2010 Excel Spreadsheet formulae were used to calculate means, standard deviations for populations, one way analysis of variance (ANOVA), and Pearson's

correlation coefficients. The procedures and formulae inputted into Microsoft Excel Spreadsheet cells to calculate 95% confidence limits, Chi squared based upon contingency tables, and Spearman's Rank correlation coefficient were taken from standard statistical texts (Hinton, 2014; Langley, 1968; Upton & Cook, 1996.)

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 the Null Hypothesis (H^0), that there was no difference between means or counts, indicated consistency in assessment outcomes and rejecting it demonstrated inconsistency. For the investigation of correlations, it was the opposite. Accepting H^0, that there was no correlation between grades for individual teachers' standards and the overall teaching grades awarded demonstrated inconsistency. Rejecting H^0 and finding a significant positive correlation suggested an association between the standard and overall teaching performance in the minds of the assessor, and therefore consistency. However, a negative correlation or random association would indicate grading decisions about individual standards and overall teaching that were inconsistent with the guidance and training the assessors had received.

Results

Comparison of mean grades and 95% confidence limits

The use of means and 95% confidence limits (Table 1) to demonstrate assessment data pictorially was easily understood by non-mathematically minded teacher educators. It indicated the progression in overall teaching grades throughout the year awarded across all the programmes. The 95% confidence limits of the mean overall grades awarded by programmes at the end of each phase of training overlapped there was no significant difference between the mean grades for all but one pair of programmes at the third formative assessment point (Table 1). This suggested that there was a high degree of consistency between programmes.

Table 1

Mean Overall Teaching Grades Recorded For Different Initial Teacher Education (ITE) Programmes At A Higher Education Institute (HEI) Qualified Teacher Status (QTS) Provider

ITE Programme	Formal Review Point	Mean grades			
		n	*M*	*SD*	*95% CI*
Secondary PGCE	First Formative	94	2.5	0.7	[2.4, 2.6]
	Second Formative	91	2.3	0.6	[2.2, 2.4]
	Third Formative	86	2.1	0.6	[2.0, 2.2]
	Summative	81	1.6	0.7	[1.5, 1.7]
Secondary Salaried	First Formative	17	2.2	0.5	[1.9, 2.5]
	Second Formative	16	2.1	0.6	[1.8, 2.4]
	Third Formative	16	1.6	0.6	[1.3, 1.9]
	Summative	12	1.2	0.4	[0.9, 1.5]
Primary PGCE	First Formative	19	2.4	0.6	[2.1, 2.7]
	Second Formative	23	2.3	0.5	[2.1, 2.5]
	Third Formative	29	1.8	0.5	[1.5, 2.1]
	Summative	21	1.4	0.6	[1.1, 1.7]
Primary Undergraduate Year 3	First Formative	44	2.3	0.6	[2.1, 2.5]
	Second Formative	48	2.1	0.5	[1.9, 2.3]
	Third Formative	48	2.0	0.4	[1.9, 2.1]
	Summative	47	1.3	0.5	[1.1, 1.5]
Primary Undergraduate Year 4	First Formative	40	2.3	0.4	[2.2, 2.4]
	Second Formative	39	1.9	0.6	[1.7, 2.1]
	Third Formative	36	1.8	0.6	[1.6, 2.0]
	Summative	38	1.3	0.5	[1.1, 1.5]

One way analysis of variance

One way ANOVA indicated no significant differences in the distribution of grades across the five ITE programmes at any assessment point suggesting consistency between programmes (F (4, 15) = 1.02, p = 0.4; F (4, 15) = 0.94, p = 0.5; F (4, 15) = 0.64, p = 0.6; F (4, 15) = 1.02, p = 0.4)

Chi squared analysis

Tables 2-5 show the observed distribution of grades across the five ITE programmes at each formal review point during the year. These were used to calculate expected numbers and Chi squared values calculated using Yates' Correction if appropriate.

Table 2

Distribution Of Grades At The First Formative Review Point

	Programme				
Grade	Secondary PGCE	Secondary Salaried	Primary PGCE	Primary Undergraduate Year 3	Primary Undergraduate Year 4
1	0	0	0	0	0
2	45 (53)	13 (10)	16 (21)	27 (25)	30 (23)
3	49 (41)	4 (7)	21 (16)	17 (19)	10 (17)
4	0	0	0	0	0
No Grade	0	0	0	0	0

Note: In all columns the observed count is followed by the expected count in brackets

Table 3

Distribution Of Grades At The Second Formative Review Point

	Programme				
Grade	Secondary PGCE	Secondary Salaried	Primary PGCE	Primary Undergraduate Year 3	Primary Undergraduate Year 4
1	5 (9)	1 (1)	1 (4)	5 (4)	10 (3)
2	86 (88)	13 (11)	36 (36)	43 (41)	29 (31)
3	0	0	0	0	0
4	0	0	0	0	0
No Grade	0	0	0	0	0

Note: In all columns the observed count is followed by the expected count in brackets

Table 4

Distribution Of Grades At The Third Formative Review Point

	Programme				
Grade	Secondary PGCE	Secondary Salaried	Primary PGCE	Primary Undergraduate Year 3	Primary Undergraduate Year 4
1	7 (11)	7 (2)	3 (4)	5 (6)	6 (5)
2	79 (75)	7 (12)	28 (27)	43 (42)	30 (31)
3	0	0	0	0	0
4	0	0	0	0	0
No Grade	0	0	0	0	0

Note: In all columns the observed count is followed by the expected count in brackets

Table 5

Distribution Of Grades At The Summative Review Point

	Programme				
Grade	Secondary PGCE	Secondary Salaried	Primary PGCE	Primary Undergraduate Year 3	Primary Undergraduate Year 4
1	44 (54)	9 (6)	21 (20)	33 (30)	27 (24)
2	34 (28)	1 (3)	10 (10)	13 (15)	11 (12)
3	8 (4)	0 (0)	1 (2)	1 (2)	0 (2)
4	0	0	0	0	0
No Grade	0	0	0	0	0

Note: In all columns the observed count is followed by the expected count in brackets

For the first three formative assessment points, Chi squared analysis indicated inconsistency in the distribution of grades across ITE programmes. This was demonstrated by significant differences between the observed and calculated expected frequencies of grades ascribed for overall teaching ability by different programmes (χ^2 (4) = 14.24, p = 0.007; χ^2 (4) = 14.73, p = 0.005; χ^2

(4) = 15.53, p = 0.004). There was no such significant difference for the summative assessment (χ^2 (8) = 9.51, p = 0.30) indicating consistency across the programmes.

Pearson's correlation coefficient

All correlations were positive, and the majority were significant with a 5% or less chance of error (Table 6). The correlation study did not indicate any 'rogue' standards (not positively associated with overall teaching grade) in any programme at any assessment point. Overall, a high level of consistency was indicated across programmes at all assessment points. Standards and programmes with weaker positive correlations are indicated in Table 6 by a probability note.

Spearman's Rank correlation coefficient

All correlations were positive, and the majority were significant with a 5% or less chance of error. The correlation study did not indicate any 'rogue' standards (not positively associated with overall teaching grade) in any programme at any assessment point. Overall, a high level of consistency was indicated across programmes at all assessment points. The use of this non-parametric methods yielded almost identical statistical conclusions to the parametric approach.

Table 6

Pearson's Correlation Coefficient (r) Between Grades For Overall Teaching And Individual Standards

ITE Programme	Teacher Standard	Formal Review Point			
		First Formative	Second Formative	Third Formative	Summative
Secondary PGCE[a]	1	0.78	0.69	0.69	0.81
	2	0.71	0.73	0.79	0.84
	3	0.62	0.51	0.37	0.58
	4	0.76	0.79	0.78	0.81
	5	0.65	0.71	0.73	0.72
	6	0.6	0.66	0.66	0.69
	7	0.68	0.67	0.61	0.73
	8	0.57	0.59	0.6	0.64

Secondary Salaried[b]	1	0.68	0.72	0.85	1
	2	0.69	0.85	0.81	0.63
	3	0.57	0.59	0.74	0.40**
	4	0.81	0.92	0.77	0.94
	5	0.53	0.73	0.85	0.78
	6	0.23**	0.81	0.58	0.76
	7	0.65	0.57	0.68	0.4**
	8	0.25**	0.09**	0.19**	0.67
Primary PGCE[c]	1	0.68	0.59	0.62	0.8
	2	0.86	0.81	0.62	0.87
	3	0.74	0.77	0.81	0.7
	4	0.84	0.35**	0.85	0.76
	5	0.85	0.71	1	0.87
	6	0.68	0.5	0.66	0.61
	7	0.76	0.33**	0.74	0.8
	8	0.8	0.56	0.52	0.5
Primary Undergraduate Year 3[d]	1	0.78	0.86	0.58	0.64
	2	0.67	0.82	0.66	0.52
	3	0.66	0.73	0.72	0.69
	4	0.72	0.77	0.67	0.73
	5	0.65	0.75	0.54	0.59
	6	0.57	0.49	0.57	0.58
	7	0.55	0.61	0.57	0.51
	8	0.58	0.58	0.57	0.65
Primary Undergraduate Year 4[e]	1	0.67	0.76	0.71	0.74
	2	0.64	0.75	0.55	0.71
	3	0.37	0.74	0.73	0.43
	4	0.57	0.82	0.22**	0.67
	5	0.42	0.69	0.85	0.59
	6	0.45	0.63	0.62	0.58
	7	0.46	0.83	0.62	0.45
	8	0.42	0.85	0.49	0.68

Note: All correlation coefficients are significant unless indicated with **

[a] n = 94, 91, 86, 80 [b] n = 17,16, 16, 12 [c] n = 19, 23, 23, 21

[d] n = 46, 48, 48, 47 [e] n = 40, 39 37, 38

**$p > 0.05$

Summary

All the statistical methods used to compare the overall teaching grades ascribed across the five programmes indicated consistency between partnerships for the final summative review point.

However, Chi squared tests demonstrated differences in the distribution of grades ascribed across programmes for formative review points. The Chi-squared test is recommended for future monitoring exercises at the HEI QTS provider.

Both parametric and non-parametric methods for comparing grades ascribed for individual standards and overall teaching gave similar results and statistical conclusions. The more rapid Pearson's correlation coefficient is recommended for future monitoring exercises.

This quantitative survey approach demonstrated consistency in assessment outcomes across and within ITE programmes at the HEI QTS provider. At this point, this is assumed to indicate consistency in assessment practices.

References

American Psychological Association (2020) *Publication Manual of the American Psychological Association – 7th Edition*, American Psychological Association

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).

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

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

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

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

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