Tynan, RJ and McLain, MN

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

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

Rick Tynan
Liverpool John Moores University
Education Health and Community, School of Education
Barkhill Road
Aigburth
Liverpool
L17 6BD
R.J.Tynan@ljmu.ac.uk
0151 231 4644

Matt McLain
Liverpool John Moores University
Education Health and Community, School of Education
Barkhill Road
Aigburth
Liverpool
L17 6BD
M.N.McLain@ljmu.ac.uk
0151 231 4622

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

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

A small group involved in initial teacher education at an 11-16 academy school took part. The academy was in partnership with a higher education QTS provider. Each respondent independently completed an anonymous on-line sorting exercise using a concourse of 41 statements about the assessment of trainees. This was constructed from appropriate literature, national policy and the partnership's documentation and guidelines.

The group included a visiting tutor, two school mentors and five trainees. Their responses were analysed using standard Q-methodology software. Participants demonstrated a clear consensus about prioritising statements concerning compliance with national requirements and local guidance. However, factor analysis identified one group, one pair and two individuals with subjective differences in their levels of agreement with statements about the assessment of trainees' teaching competencies.

A group of four, that included one mentor and his mentee, prioritised statements linked to fairness, validity and quality assurance processes within and external to the academy. The other mentor and her mentee perceived assessment almost solely in terms defined by the Teachers' Standards in England.

This suggested that the different perspectives on assessment were associated with influences other than the respondents' differing roles within ITE.

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

Context and Literature Review

Initial Teacher Education in England

For Robinson (2006) the history of initial teacher education (ITE) in England has been characterised by two important themes. There have been broad cyclical swings between school led models of provision and those based in college or university. These have been influenced by the link between ITE and the politics and provision of state education. Robinson (2006) was writing at the beginning of, what turned out to be, an extended and on-going period of central government control of teacher education with an emphasis upon school-led professional training and the multiplication of work-based routes into teaching. Over time, this approach to ITE in England has developed and extended the mentoring role of school based teacher educators. The assessment of trainees' teaching has become a key part of the mentor's role (Department for Education, 2011).

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

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

Cajkler and Wood (2016, p1) used the term 'reductive models' for ITE approaches such as this. They suggested that the need to gather evidence to meet standards must restrict training experiences and advocated a collaborative approach to planning, observing and evaluating lessons (lesson study) as a more effective way of

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

Assessing teachers and the potential for subjectivity

The Teachers' Standards (2011) give clear guidance that head teachers and their delegates have the first responsibility in schools for assessing teachers. However, there is no guidance on mechanisms or the evidence for assessment. School based ITE mentors are free to choose their methods of assessment subject to internal and external quality assurance procedures. In the early stages of competency based assessment using standards in England, Jones (2001) compared the way 25 English and 25 German ITE mentors perceived their roles. Both groups agreed on their advisory role but English mentors placed more importance on their assessment role than their German counterparts. They were also less likely to perceive difficulties associated with the assessment of trainees or potential conflicts between their assessment and other mentoring roles. Anecdotal experience from our partnerships would suggest mentors perceive their assessment roles pragmatically in much the same way as the mentors from England in this early small-scale study.

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

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

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

Reolofs and Sanders (2007) provide an example of support for the use of performance criteria of the sort associated with competency based teacher assessment In England. In a thorough attempt to provide a framework for the assessment of teaching performance aimed primarily at Dutch teacher educators, Reolofs and Sanders (2007) maintained that applying a reductionist model was more likely than other approaches to result in valid inferences about teaching competence. They emphasised that this allows the assessor to focus on different areas of teaching separately when reaching decisions. However, the adoption of competency-based assessment of teacher performance in England stimulated some academic debate concerning its validity and reliability for assessing the performance of trainees on ITE programmes. Turner-Bisset (1999) raised early concerns over the use of reductionist standards and descriptors and considered these inadequate for providing a model of the subject knowledge demonstrated by teachers. Proposing an alternative model, Turner-Bisset (1999) emphasised that teacher self-knowledge was an important element missing from the descriptors at that time. Again anecdotally, it has been our experience as teacher educators that trainees and mentors have found the performance criteria encompassed in The Teachers' Standards (Department for

Education, 2011) a useful focus for the formative analysis of teaching skills and for guiding mentoring and coaching feedback. Leshem and Bar-Hama (2008) reported that their students expressed similar sentiments during a study of the introduction of competency-based assessments to an ITE programme in Israel. However, their students also preferred assessors to use holistic judgements when making summative assessments.

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

How this Q-method study informs our previous investigations

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

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

2017). The authors accepted that these speculations on the observed patterns in the assessment data needed to be informed by the second qualitative phase of the study in order to investigate the degree and nature of subjectivity about assessment and grading amongst assessors.

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

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

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

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

Methodology and methods

Overview

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

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

Participants for the study were invited from attendees at a mentor training session at a partner secondary school academy. Eight respondents volunteered and represented a range of ITE partnership roles. This captured the perspectives of trainees and subject mentors directly involved in the assessment and grading process and a HE school liaison tutor with a quality assurance role. This was a convenience sample (Etikan *et al*, 2016) of mentors and trainees available to the lead researcher from a wider population of ITE participants at the HE provider. This non-random sampling technique could result in over representation of more numerous categories of participant. However, this is appropriate for the study as inferences are made only about subjective differences in the responses of the actual

participants with no reference to any larger population (Etikan *et al*, 2016) involved in ITE.

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

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

The concourse of statements

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

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

Table 1 Statements in Category A: Individual Teachers' Standards

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

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

Table 2 Statements in Category B: Quality assurance processes

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

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

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

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

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

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

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

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

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

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

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

Table 6 Statements in Category F: OfSTED descriptors

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

Table 7 Statements in Category G: Celebration

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

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

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

Findings

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

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

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

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

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

Table 10 Consensus and distinguishing statements

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

Interpretation of findings

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

Consensus statements

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

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

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

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

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

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

Group 1's Distinguishing Statements Profile: Making it fair

Group 1 consisted of a male history mentor and his male trainee, a female art and design trainee and a male mathematics trainee. The respondents in Group 1 agreed more than Group 2 with most of the statements in the categories:

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

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

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

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

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

Group 2's Distinguishing Statements Profile: Applying the National Criteria

Group 2 consisted of a female mathematics mentor and a female mathematics trainee. The respondents agreed more than their counterparts in Group 1 with most of the statements in the categories:

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

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

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

Summary of main findings

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

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

Discussion

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

Assessment and The Teachers' Standards (Department for Education, 2011)

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

Agreement in these areas might suggest support amongst the participants for the views of Reolofs and Sanders (2007) on the value of reductionist models for improving the validity of measurements of competence and the value of grades for feedback in ITE (Matthews and Noyes, 2016). This might be stronger in Group 2

given their profile of distinguishing statements consistently supporting the importance of individual standards for assessment. However, Group 1 made a distinction between the best-fit application of standards and descriptors and fully holistic judgements reflecting the preference for holistic summative assessment expressed by the students in Leshem and Bar-Hama's (2008) study. Group 1 might be receptive to sort of the change in assessment emphasis towards pedagogic literacy advocated by Cajkler and Wood (2016) or Korthagen's (2017) ideas concerning integrating the teacher as a person into ITE professional development programmes.

Assessment and ITE Partnerships

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

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

Assessment and Subjectivity

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

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

Conclusions

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

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

than the teacher competencies that are currently the only measures of teaching skill in England.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Pruneddu, A. (2014) *QSortWare*, [Online software].
Available at http://www.qsortouch.com (Accessed October 2014).

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

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

Schmolck, P. (2014) *PQ Method (Version 2.35),* [Computer program].
Available at http://schmolck.userweb.mwn.de/qmethod/index.htm (Accessed October 2014).

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

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

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

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

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

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

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

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

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