Tynan, RJ and Jones, RB

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

Tynan, RJ and Jones, RB (2018) Assessing trainee secondary teachers on school placement: Subject knowledge and overall teaching grades. Teacher Education Advancement Network Journal, 10 (1). pp. 20-34. ISSN 2054-5266

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

Teacher Education Advancement Network Journal

Vol 10(1) pages 20-34

Rick Tynan and Robert Bryn Jones
Liverpool John Moores University

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

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

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

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

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

Citation
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.

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

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

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

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

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

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

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

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

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

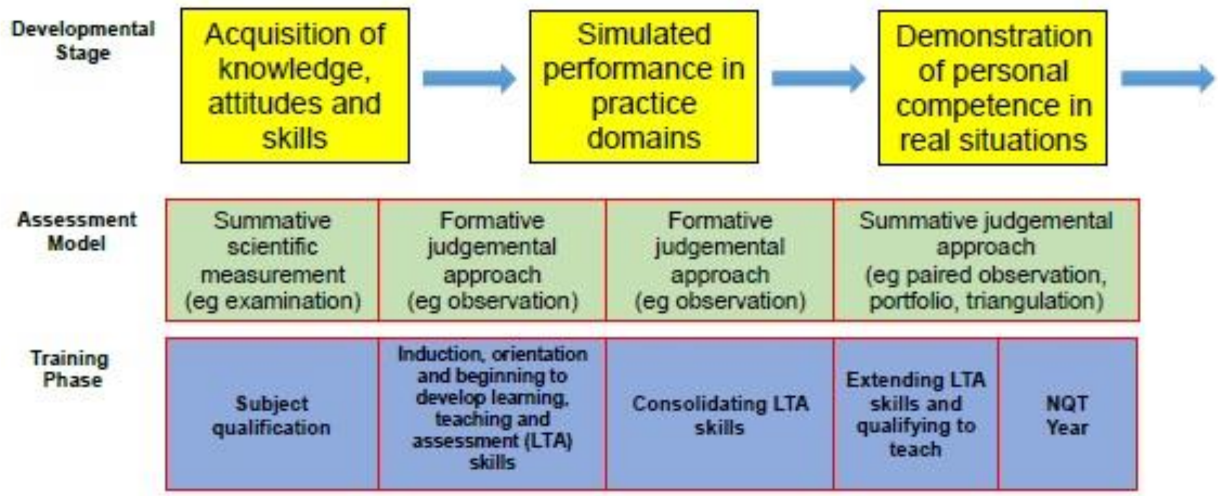


Figure 1. A model for the professional development of trainee teachers (developed from Hager & Butler (1996) and Martin & Cloke (2000).

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

Methodology and methods

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

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

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

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

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

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

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

Simple comparison of the observed and expected counts in Table 1 would suggest that the two sets of data are varying independently of each other. More Grade 1s than expected were awarded for Teachers' Standard 3 and fewer than expected for the overall teaching grade. Conversely, mentors awarded fewer Grade 3 and 4 than expected for Teachers' Standard 3 and more than expected for the overall teaching grade. This approach compares both data sets with the overall frequencies calculated using the totals boxes as described in Table 1.

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

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

Table 2. Chi squared calculation for Table 1.

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

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

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

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

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

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

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

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

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

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

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

Results

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

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

Key:

OTG Overall teaching
S3 Teachers' Standard 3 grades counts
S4 Teachers' Standard 4 grades counts
← More high grades and less low grades than expected

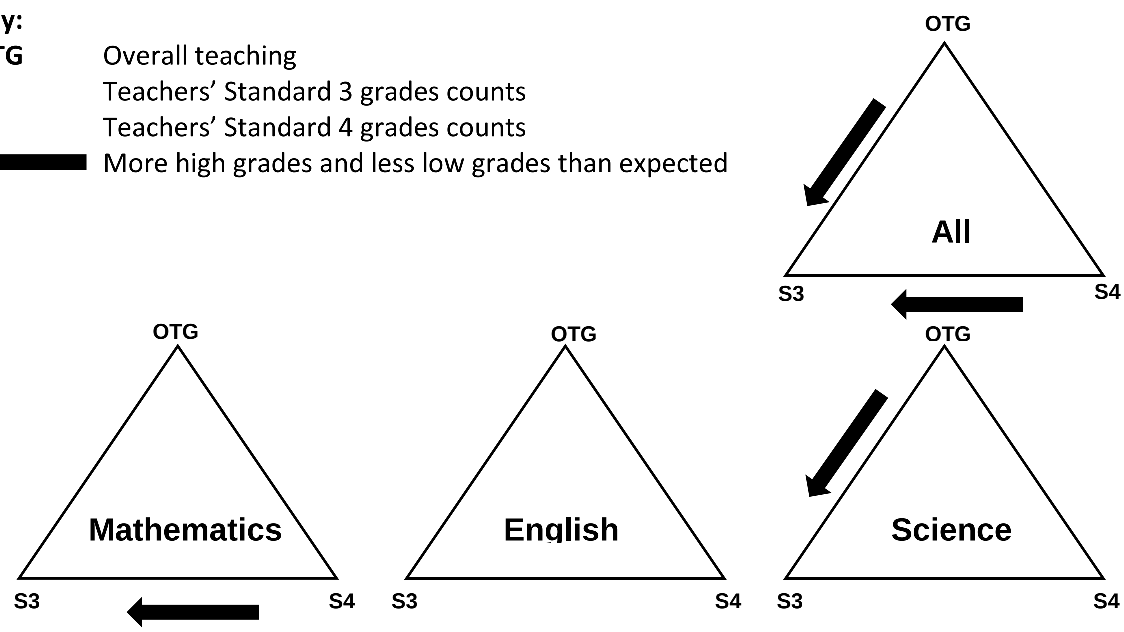


Figure 2. Visual presentation of Chi Squared analysis for the Academic Year 2014-15.

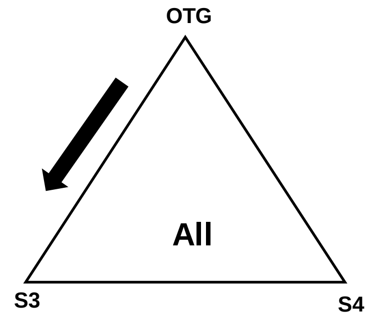
Key:

OTG Overall teaching

S3 Teachers' Standard 3 grades counts

S4 Teachers' Standard 4 grades counts

More high grades and less low grades than expected



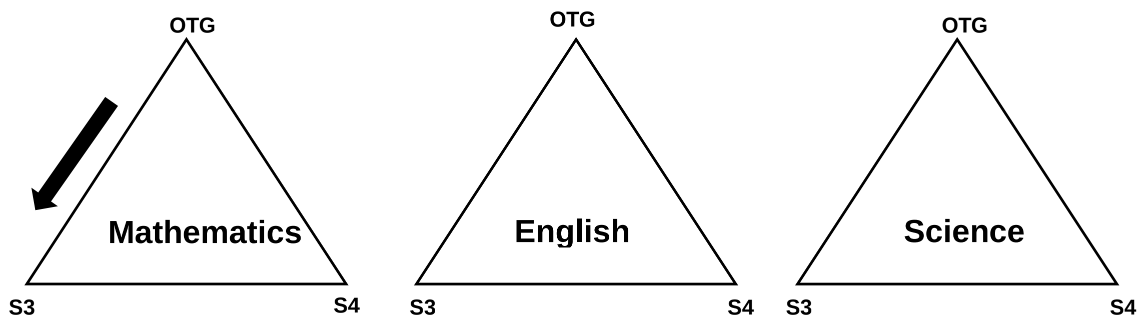


Figure 3. Visual presentation of Chi Squared analysis for the Academic Year 2015-15.

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

Conclusions and recommendations for further work

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

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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