# Common understanding or 'hodgepodge'? The consistency and accuracy of school-based mentors' assessment of trainee primary teachers in England

Teacher Education Advancement Network Journal

Vol 11(2) pages 72-81

Liz Zsargo and Jean Palmer
University of Huddersfield

**Abstract**
The small-scale research study reported in this paper aimed to explore the ways in which mentors in primary school settings approach the assessment of trainee teachers within the English Initial Teacher Education system. Consistency and accuracy of assessment is judged by Ofsted (Office for Standards in Education) as a key indicator of quality, and yet very little is understood about the ways in which assessment is enacted by mentors. The paper explores some of the tensions involved in mentor assessment of trainees: variable interpretations of criteria, the dichotomous role of the mentor and differing operating contexts and personal constructs of mentors. It is concluded that these all influence mentor decision making, and a 'mentor assessment identity' is theorised, based on mentors' previous and current experiences, contexts and beliefs. Consistency in terms of the application of criterion-based assessment to trainees' teaching is thus fundamentally critiqued, and a way forward is suggested based on qualitative formative feedback. It is thus suggested that ITE providers could move away from grading and focus instead on strengths and areas for development for trainee teachers.

**Keywords**
Mentor; mentoring; assessment; primary trainees; primary; criterion-based; mentor assessment identity; feedback; grading; consistency; accuracy.

This paper reports on a project undertaken at a University in Northern England during the academic year 2016/17. The aim of the project was to identify the extent to which the assessment of trainee teachers by school-based mentors could be regarded as consistent.

## Context

As has been discussed previously in this journal (e.g. Tynan and Mallaburn, 2017), initial teacher education (ITE) in England is operating in the context of a strongly performative and instrumentalist approach to measuring outcomes. This is largely driven by Ofsted (Office for Standards in Education) – within the wider policy context of performance and accountability (Ball, 2017). Ofsted's current regime for inspecting ITE includes an emphasis both on general consistency and on the accuracy of assessment (Ofsted, 2018). As Ofsted inspection outcomes have a direct impact on institutional reputation, which has the potential to affect recruitment of future students and consequently course viability, the stakes are extremely high. Besides these external drivers, internal drivers also exist. Students in England are generally assessed by their teacher mentors in schools operating in a wide range of different contexts, across the full age range of teaching. School systems, expectations and customs can display considerable diversity, besides the variability which individual understandings and experiences can bring to the assessment situation. Student teachers can be very alert to perceived differences and understandably expect ITE (Initial Teacher Education) providers to ensure that assessment criteria are applied with consistency by mentors across the full range of settings. Any

**Citation**
Zsargo, E. and Palmer, J.F. (2019) 'Common understanding or 'hodgepodge'? The consistency and accuracy of school-based mentors' assessment of trainee primary teachers in England', *TEAN journal* 11 (2), pp. 72-81.

assessment system requires its users to understand the criteria against which they are assessing, and ITE providers will often strive to ensure that this is the case as a marker of quality provision, quite apart from the Ofsted imperative.

The project was led by active teacher-educators both heavily involved in working with school partners and in preparing and supporting mentors to undertake their roles. In common with many ITE partnerships we have adopted criteria developed by the National Association for School Based Initial Teacher Training (NASBTT) and Universities' Council for the Education of Teachers (UCET) which provide descriptors for each Teacher Standard against the four Ofsted categories of 'Outstanding', 'Good', 'Requires Improvement' (RI) and 'Inadequate' (NASBTT, 2017), also often known as grades 1 (for Outstanding) through to 4 (for Inadequate). These provide our main instrument by which trainees are provided with numerical grades at regular intervals during their course. As a partnership we had already made changes to practice in response to Ofsted feedback: this included introducing online mentor materials, regular moderation meetings, improving the moderation role of university-based link tutors, and strengthening the role of senior moderator. During 2016/17 data analysis of numerical grades, as well as qualitative feedback from partners, seemed to show that we were achieving a greater degree of consistency in assessment. However, we remained uncertain whether this data gave us the full picture. Anecdotal comments heard from mentors and trainees about the application of the grading system raised continued questions about consistency. We were aware of work in the school context questioning the extent to which graded criterion-based assessment systems can offer consistency (e.g. Christodoulou, 2017). We had concerns about the way that grading can be perceived by both trainees and mentors: the eliding of 'Inadequate' and 'RI'; 'Good' being seen as disappointing; the apparently addictive effect of 'Outstanding'. Finally, we were conscious of the impact which being assigned a particular grade potentially had on trainees, and concerned about the ethical implications of this (Stobart, 2008). As a result we set up a small scale qualitative research project to understand further the actual act of assessment by mentors and the degree to which the current system of assessment can be deemed to provide a reliable measure of trainee outcomes. Clearly work also needs to be done with trainees about the ways in which they process all forms of assessment including grading: this will form the next stage of the project.

**Literature review**

Despite much relevant work on many aspects of mentoring and partnership, as well as on assessment in various contexts, there appears to have been little focus to date on how mentors actually go about the complex business of grading their student teachers. We therefore cast our net wider than ITE, seeking out examples of previous research in any aspect of assessing practice and indeed the use of graded criterion-based assessment in other educational contexts such as schools. It is immediately apparent that the assessment of practice against set criteria is a deeply problematic area, which for the purposes of this paper we have ordered into four main areas of tension: variable interpretations of criteria; the dichotomous role of the mentor; the operating context of different mentors and the personal constructs of mentors all influencing the act of grading.

*Variable interpretations of criteria*

Considerable literature exists highlighting the difficulties associated with achieving consistent interpretation of criteria between different assessors. For example, Tillema's ITE-based study focussed on the process of appraising a particular lesson between the teacher mentor, the university-based tutor, and the student themselves. He found that participants attached different weights to particular criteria based on what they deemed to be more important, which led to a high degree of variability between assessors (Tillema, 2009).

The criteria on which such assessments are based are also susceptible to multiple interpretations: Wiliam (2014), in a powerful critique of the use of national curriculum levels to support detailed criterion-based assessments, argues that 'even the most tightly specified criteria will support a range of evaluations'. He points out that in developing a shared understanding of what the criteria may mean in a particular context, what we are actually doing is creating a norm-referenced assessment, with all the inherent issues of subjectivity that may suggest. The use of qualifying words in criteria has received particular criticism. For example the use of adjectives such as 'excellent', 'good' or 'adequate' which Bloxham, den Outer, Hudson and Price (2016), writing about the use of assessment criteria in UK Higher Education institutions, show convincingly that assessors interpret very differently. Christodolou (2017) mounts a parallel critique of the use of descriptors to assign grades to children, showing that adverbs such as 'effectively', 'appropriately' 'confidently' are all subject to interpretation and yet misleadingly create the 'illusion of a common language'.

*The dichotomous role of the mentor*

Mentors in many jurisdictions including the English ITE context are expected to act both as supportive critical friend and accountable assessor of performance: Tillema, Smith and Lesham (2011:142) describe this as 'Janus-like', with mentors switching between modes in order to fulfil their different roles with all the tensions that may entail. The assessor role can be perceived as undermining the formative role (e.g. Matthews and Noyes, 2016) with students and mentors becoming overly focussed on the grade being awarded (Lofthouse and Hall, 2014). Thomson, Hall and Jones described how schooling can be seen as '*infinitely calculable and available for calibration* and permanently available for *forensic* dissection through apparently objective, scientific and transparent computational practices' (2010:652, emphasis original). We would argue that the same is true of the practice of student teachers and that the conversion of student learning and development into data-driven practice is a spurious attempt at false objectivity.

*Operating context of mentors*

Mentors do not carry out their assessment in some kind of antiseptic vacuum, rather they operate within an ecosystem comprising the national and local policy context, the particular school environment, the mentoring set-up in their school and the ITE provider partnership (possibly more than one) with its attendant processes (Kemmis et al, 2014). Lofthouse and Thomas (2014) carried out a study which suggested that the performative culture in which mentors operate had an influence on mentoring practices, with mentor-student interactions sometimes dominated by issues of evidence and target setting. Concerns about grading were also evident, with one student participant reporting that her mentor had said at the very start '…well I'll tell you now that I'll not be giving you that or that…[grade]' (Lofthouse and Thomas, 2014:212). This not only casts doubt on the equitable application of the full range of criteria by this mentor at least, but also hints at how this mentor felt the criteria for particular grades were unobtainable by a student, and that the student perceived this as unfair because other mentors had a different view.

*Personal constructs of mentors*

The example above is suggestive of the final area of tension – the influence personal beliefs and experiences appear to have on assessment decisions. Drawing on a number of studies of Swedish PE teachers' use of assessment criteria to grade, Svennberg, Meckbach and Redelius (2014:200) identify the use of 'internalised criteria' – otherwise known as gut feeling. Drawing on earlier work by Brookhart (1994), they report that teachers show awareness of the grading criteria, but use a 'hodgepodge' of factors to come to their assessment decisions including their own experiences and beliefs (2014:211). A similar phenomenon has been suggested by a number of researchers. Lofthouse and Wright (2012) note that former students who later become mentors tend to reproduce their experiences with their own students whilst Bullough (2005) posits that mentor identity is based on

teacher identity which is itself based on one's personal identity; in other words: who you are influences how you mentor and - we would argue – by extension, how you assess (Looney, Cumming, van Der Kliej and Harris, 2017). Vanassche and Kelchtermans conceptualise this as a 'personal interpretive framework', defined thus:

> the set of cognitions and beliefs that operate as a lens through which teacher educators perceive their job situations, give meaning to, and act in them (2014:118).

**Methodology and methods**

The small-scale study reported on in this paper is part of a larger piece of practitioner-based research which fits within an action research paradigm. The overall original aims of the larger project were to improve the quality and consistency of mentoring in the Partnership and to ensure greater consistency in the assessment of trainees. As the project proceeded, uncomfortable cracks appeared around questions regarding the process and nature of mentor assessment and the reliability of assessment data, and thus a key research question was developed which seemed to merit further study: How do mentors understand and carry out the assessment of trainees? It is this question which is addressed in this paper.

As part of the larger project, close to a hundred mentors from Partnership schools had already taken part in moderation discussions, worked with the university on a mentor conference and taken part in a workshop exploring their views on mentoring. The mentors represented a range of levels of experience and worked with trainees on different routes into teaching: undergraduate and postgraduate including School Direct. Volunteer mentors were sought to be interviewed about their views and experiences of both mentoring and of assessing trainees. Four mentors with varying levels of experience were interviewed using a semi-structured interview format and seven additional mentors completed an open-ended questionnaire answering the same questions as the interviewees. These participants ranged from highly experienced 'senior' mentors to those in the first year of mentoring. All signed an agreement for their data to be used in the project and shared anonymously via publications and conferences.

As interpretive researchers we see lived experience and individuals' situated and subjective accounts of that as a valid expression of epistemic truth: we were thus seeking to enable mentors to express internal and external processes they go through in order to assess trainee teachers and decide on a grade. We acknowledge that in doing so they may have, consciously or unconsciously, expressed views they felt the researchers wanted to hear. We deliberately foreground subjective interpretations and are interested in how the mentor herself interprets the assessment situation and the act of grading. In analysing the data generated, we claim no neutral stance; we have in turn sifted and interpreted the mentors' accounts. The findings are thus *contingent*: they depend upon us and our standpoints as well as those of the mentors concerned; *particular*: they apply to these mentors in this partnership on the day they were interviewed and do not claim to be generalizable, and *constructed*: we as researchers have discussed what we think was meant and built a narrative around what was said during those interviews. The findings may nonetheless illuminate mentor assessment of trainees and suggest areas worthy of further study.

**Findings**

It is perhaps important to consider first the shared understandings and interpretations of the criteria used to assess and grade student teachers.

*Variable interpretations of criteria*

The Teachers' Standards (DfE, 2011) comprise eight standards plus so-called 'Part 2' professionalism standards, all of which must be met to successfully pass the course and achieve qualified teacher status. These are standards that are constant, being the same for both trainee teachers and experienced practitioners – something that some mentors evidently found challenging when evidencing progress:

> I find that quite a difficult thing, because the standards are the standards. And the standards are the standards for if you've been teaching for three years or if you've been teaching for three weeks
>
> (Mentor 1 – M1).

The use of Ofsted grades as outlined above implies that these standards can be met at different levels: 'RI' (which paradoxically equates to meeting the standards), 'Good' and 'Outstanding' and it was an early misconception around this that led to the present research. When asked to construct a profile of an 'RI' student teacher many mentors presented instead a profile of an 'Inadequate' student – one who would not be meeting the standards. This understandable confusion was reflected in responses where practitioners struggled to see this grade as acceptable:

> The good and the outstanding is tight to get in to, but I don't know whether a requires improvement is employably ready (M2).

> I mean I've had a student who has come out as a 3 [RI] because she's not hit that extra. She's done 3 years training and I don't think she deserves to fail but neither does she deserve to go and take a job because that would be too big an ask as an NQT (M3).

The Teachers' Standards are presented as equal in that all must be met, but it became clear that some mentors gave greater weight to certain areas:

> The overall grading judgement is good. However, I think the teaching and learning could possibly hold greater, you know the old-fashioned Ofsted, you know teaching and learning is the key (M2).

It appeared that although mentors shared a language, it would seem less so a shared interpretation of the qualifying words within the criteria. When asked to identify what was meant by the grade descriptor 'outstanding' mentors construed this in different ways:

> One of the things that I have said to students is if you are wanting me to look at you as an outstanding practitioner what you need to be able to show me is something that maybe I've not seen before or maybe it's that you are taking it to another level……If I'm making a judgment of outstanding then I want it to be the best I have seen. An outstanding student is just "natural"(M3).

> What makes you outstanding – it could be just that your lessons are consistently good (M5).

> I don't think you can quantify "all outstanding teachers look like that" or "all good teachers behave like this" (M7).

*The dichotomous role of the mentor*

This leads us to the next tension identified in the literature – that of the dual role that practitioners play as ITE mentors. They act as both critical guide and assessor – a dichotomy that has been seen as a potential threat to the wellbeing and development of student teachers (Hobson & Malderez, 2013). Interestingly when asked why they had become mentors and what they saw the role to be no mentor interviewed started with the task of grading or making assessment judgements.

When asked to offer an image to represent the role of mentor the complexity of this role became more evident with one mentor depicting a triple image of a guide, a sergeant-major and a hand-holding supporter. Mentors did appear to see themselves first and foremost as supporters and guiders rather than assessors. However, the tension between the professional and personal was clearly expressed in relation to 'not just being a friend' and 'building a professional relationship:.

> You've got to have that respect between the two of you, and I think it's got to be quite clear who's doing what in that relationship……it's about building that relationship and making the contract between the two of you clear (M1).

The importance of developing personal relationships and the use of informal observation was also seen as a crucial mentoring tool in supporting student teacher progress. Alongside this there did seem to be a view of assessment as a more formal part of the overall process, to be taken seriously:

> You begin with the "getting to know me, getting to trust me" stage, and then I think we can get into the nitty gritty of "you are now *here* for this part of your assessment" (M2).

> All that informality, although it doesn't directly impact on the formal assessment side of things it brings to it a sense of knowledge. It broadens out your knowledge of that person (M4).

One mentor vocalised both the complexity of finding a balance between these personal and professional relationships in relation to their role as assessor:

> So it's almost quantifying them academically, emotionally and figuring out the way to explain how to move through the strands for them, but *that* is a very personal thing, as well as a very professional thing. So you know professionally where you need to get to, but you have to know them personally to be able to push those buttons and say the right thing (M2).

The mentors in this study therefore acknowledged this dichotomous role but seemed to see this as something which was expected and, for them, manageable.

*Operating context of the mentor*

Mentors discussed the obligation they felt to balance the needs of the student with the expectations for the children in their classrooms:

> At the end of the day we have to secure our children's education (M3).

> It's about getting it right for the student, but also getting it right for all those children that are valuable beyond measure really (M4).

They also considered the status of the mentoring role within schools, how it was a responsible role that could add pressure to what for some was already a performative environment and the importance of supporting mentors:

> I do think it can be an issue depending on what school you're at and how committed that school is to, or how able they are to be committed to the whole idea of mentoring. Having the capacity (M4).

> I think if you're in a school where you know you're going to be backed up then that's ok - you feel supported. I think if you're not that can get professionally difficult...because you don't want people to crack under the pressure (M2).

Several anxieties appeared to be evident. Firstly, tensions between teaching the children and supporting a student, in which the children were seen as the higher priority. Secondly, the school environment being supportive (or not) of mentors and their work. Interestingly, school support is framed in terms of both commitment and capacity. The reference to mentors 'cracking under the pressure' is a worrying hint that mentoring can be a very challenging role and that the context within which it is taking place is a key reason for mentors feeling under pressure.

*Personal constructs of the mentor*

When asked directly how they approached assessing their trainees against the Teachers' Standards, mentors responded thus:

> A feel for the incidental parts of teaching. A gut instinct then I use the Teachers' Standards to gauge where they are. What they need to do to improve (M6).

> Actually it's probably a lot of gut reaction. There's a general feeling – you can tell. And I guess the standards are there to unpick what it is that "you know" (M1).

This echoes Svennberg, Meckbach and Redelius' (2014) notion of internalised criteria, using the standards to confirm 'a feeling' based on one's own knowledge rather than using the criteria as a checklist. As one mentor struggled to articulate:

> You see a teacher and you think "he's got it", but then you think "what actually is it that he's got?" and sometimes you can have "it", but it's not ready yet. But it's there, and I think that's the thing with students…sometimes you can see it's there but it's not fully formed yet. And sometimes you think "you've just not got it and I don't think you'll ever have it", and sometimes as a mentor you think "can you teach people how to get that?"(M1).

So if judgements emerge from 'what we know' is it realistic to assume a shared knowledge and therefore a consistent baseline? This degree of subjectivity seems to suggest that consistency may remain elusive. As suggested earlier, if 'what we know' is also a reflection of 'who we are' it would seem that this might be reflected in how we assess, with who we are as mentors and our personal understanding of good practice leading to conflict when making judgments. One mentor acknowledged this tension, saying:

> It's helped me to think about where personal opinion tips into professional judgement. And there are times when you might not like the way that someone is doing something, but that doesn't mean that what they're doing isn't effective or working professionally. So that's quite challenging on a personal level as well (M4).

With another commenting:

> I realised that actually I'm not just projecting myself on to you, you are your own personality as a teacher….and I'm not creating just another clone of me, I've got to allow you to be your person (M1).

As can be seen, mentor responses across the study expressed personal and professional values and beliefs, assumptions and experiences. All of these appear to have an impact on how they mentor, guide and ultimately assess trainees. The mentors interviewed for this study saw themselves primarily in a supportive role, with assessment as something based mainly on prior knowledge and gut feeling. They also spoke about some of the challenges of mentoring including the tension between the mentoring and the class teacher role, and the level of support received from the school. Our findings lead us to propose that mentors are in fact drawing on multiple positions when making judgements and from this we have theorised a 'mentor assessment identity'.

**Discussion**

If we are to argue for a 'mentor assessment identity' then it is necessary to consider the idea of teacher, mentor and assessment identities more closely. Gee (2000) conceptualises identity as an 'analytical lens' through which to study an individual's beliefs and actions within a given situation. Drawing on Gee's framework (2000:100), it can be suggested that personal identity comprises many overlapping identities including a 'Nature-identity' – who we are in ourselves – and an 'Institution-identity' derived from our position within and in relation to the multiple institutions with which we have contact.

*Teacher identity*

If we posit that who someone is, their qualities and beliefs, their social and professional experiences all combine to form their personal identity then it becomes clear that this Nature-identity will influence an individual's positioning as a teacher and overlap significantly with their Institution-identity. An individual's institutional identity will be defined by their understanding of what an effective teacher is and their personal perceptions of good practice - the very act of being a teacher and the roles and responsibilities that this entails. This in turn reflects how the individual wants to be seen by peers and colleagues and is established and reframed through interaction with others (Looney et al, 2017). Ultimately this, we can argue, becomes teacher identity (Mockler 2011) Developing these ideas further, Looney et al suggest not only a complex teacher identity but a teacher *assessment* identity, defining assessment as:

> socio-cultural activity that involves social interactions among stakeholders and the nature of learning itself…it occurs in a social context, influenced by national and state policies, expected learning (curriculum), pedagogical directions, and community expectations (2017:3-4).

*Mentor identity*

Looney et al's analysis chimes with ours, albeit applied in a different context. In our research, we have the additional layer of the teacher as a mentor - so still a teacher, but also acting as a mentor to beginner teachers. As the literature in this field suggests, mentors may conceptualise their own mentor-identity in different ways, which adds additional complexity to the approach to assessment. As mentioned above, in another part of our research, mentors were asked to identify a metaphor for themselves in a mentoring role and these were as varied as 'a guiding light' 'a mamma bear' and 'sometimes a sergeant-major'. Lofthouse and Thomas (2014) discuss the influence of policy and other contexts on mentoring processes and Hobson and Maldarez (2013) identify the negative effects of 'judgementoring', where the mentor identity as an assessor has taken over from other aspects.

Mentoring is clearly a many-faceted role that in turn is subject to the individual's 'personal interpretative framework' (Vanassche & Kelchtermans 2014). This 'lens' is constructed from not only the individual's perception of the task, but also their own subjective educational theory. Looney et al (2017:4) suggest that assessment has an emotional component and that teachers' 'conceptions, beliefs, experiences and feelings are all significant in their assessment work' These also influence how mentors 'position' their practice as teacher educators and impact on how they relate to the student teacher, their expectations and judgements.

*Mentor assessment identity*

This leads us to suggest that not only as teachers but also as mentors we act out our own 'assessment identity'; thus, a mentor assessment identity constructed from 'personal experience, professional context and the external political environment' (Mockler 2011:521). The lens through which the mentor focuses will influence their final assessment as each works within their own mentor assessment identity. Looney et al proposed that:

> Who teachers are in the process of assessment is as important as what they know and are able to do. This is at the heart of teacher assessment identity (2017:15).

We argue that the same applies to teachers as mentors. This brings into question a system that relies so heavily upon the objective completion of a shared criterion-based assessment for grading purposes, or at the very least it challenges us deeply as we search for consistency based upon such a system.

**Conclusion and next steps**

Based on our findings we therefore argue in this paper for an interpretation of mentoring and assessment as – at least in part - entwined ontological acts. Put simply, that personal values, beliefs and experiences influence an individual's enactment of the role of the mentor, of which the enactment of assessment forms an integral part. It is thus never possible for a mentor to completely operate outside of their own context, experience and beliefs. Adapting Kelchtermans' words originally spoken about teaching (2009:258) we could suggest: the fact [is] that mentoring is done by somebody. Mentoring is an act, or mentoring is enacted by someone. It matters who the mentor is.

The act of assessing trainees is thus fraught with complexity both from the mentor perspective, and with professional and personal import from a trainee perspective. Although this has not been explored within this phase of the research, we are nonetheless alive to the ethical implications of assigning grades to trainees, based on the work of e.g. Stobart (2014), who argues that the assignment of a particular grade to an individual creates a new sense of identity which did not previously exist. The trainee may regard themselves as in some way defined by the grade assigned, with the implications this has for their own identity as a beginning teacher. Although, as our research has shown, all assessment will be subject to a wide range of personal, social and situational influences, assigning a grade has, we argue, a particularly deleterious effect due both to its apparent (misleading) objectivity and its impact on identity. This places a great ethical burden on to mentors, and work also needs to be done to understand the trainee teacher experience of the assessment process.

One of the great strengths in the English ITE system is surely its relational nature; the way in which personal bonds are formed between mentors and trainees. We argue that it should therefore be possible for assessment systems in teacher education not to be defined by Ofsted grades. It should be possible to provide very clear feedback to trainees, and to assess how well they meet the Teachers' Standards, without applying Ofsted grading to the trainees as individuals. We are not, of course, arguing against the importance of trainee teachers being judged and meeting a set of standards is clearly vital for entry into the profession – and with such judgements will always come debates about

subjectivity. However, the focus should surely be on the regular, deep formative discussions between mentors, trainees and tutors and on an individual's areas of strength and development – all of which are already characteristic of the ITT system in England. It is of critical importance that research continues urgently to develop and evaluate new approaches to the assessment of trainee teachers.

**References**

Ball, S. (2017) *The Education Debate* (3rd ed.). Bristol: Bristol University Press.

Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016) Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment and Evaluation in Higher Education,* 41(3), pp466-481.

Brookhart, S. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), pp279-302.

Bullough, R. (2005) Being and becoming a mentor: school-based teacher educators and teacher educator identity. *Teaching and Teacher Education,* 21, pp143-155.

Christodoulou, D (2017). *Making Good Progress? The future of assessment for learning.* Oxford: Oxford University Press.

DfE (Department for Education) (2011) Teachers' Standards: guidance for school leaders, school staff and governing bodies. Available at: https://www.gov.uk/government/publications/teachers-standards (Accessed 9 October 2018)

Gee, J. (2000) Identity as an Analytic Lens for Research in Education. *Review of Research in Education*, 25, pp99-125.

Hobson, A., & Malderez, A. (2013) Judgementoring and other threats to realizing the potential of school-based mentoring in teacher education. *International Journal of Mentoring and Coaching in Education,* 2(2), pp89-108.

Kelchtermans, G. (2009) Who I am in how I teach is the message: self-understanding, vulnerability and reflection. *Teachers and Teaching: theory and practice,* 15(2), pp257-272.

Kemmis, S., Heikkinen, H., Fransson, G., Aspfors, J., & Edwards-Groves, C. (2014) Mentoring of new teachers as a contested practice: Supervision, support and collaborative self-development. *Teaching and Teacher Education,* 43, pp154-164.

Lofthouse, R., & Hall, E. (2014) Developing practices in teachers' professional dialogue in England: using Coaching Dimensions as an epistemic tool. *Professional Development in Education,* 40(5), pp758-778.

Lofthouse, R., & Thomas, U. (2014) Mentoring student teachers; a vulnerable workplace learning practice. *International Journal of Mentoring and Coaching in Education,* 3(3), pp201-218.

Lofthouse, R., & Thomas, U. (2017) Concerning collaboration: teachers' perspectives on working in partnerships to develop teaching practices. *Professional Development in Education,* 43(1), pp36-56.

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2017) Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy and Practice*. pp1-26 doi:10.1080/0969594X.2016.1268090

Mockler, N. (2011) Beyond 'what works': Understanding teacher identity as a practical and political tool. *Teachers and Teaching*, 17, pp517–528. doi:10.1080/13540602.2011.602059

Matthews, R., & 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), pp247-261.

Ofsted. (2018) Initial Teacher Education Inspection Handbook. Available at: https://www.gov.uk/government/publications/initial-teacher-education-inspection-handbook (Accessed 9 October 2018)

Stobart, G. (2008) *Testing Times: the uses and abuses of assessment*. Abingdon: Routledge.

Svennburg, L., Meckbach, J., & Redelius, K. (2014) Exploring PE teachers' 'gut feelings': an attempt to verbalise and discuss teachers' internalised criteria. *European Physical Education Review,* 20(2), pp199-214.

Thomson, P., Hall, C., & Jones, K. (2010) Maggie's day: a small-scale analysis of English education policy. *Journal of Education Policy,* 25(5), pp639-656.

Tillema, H. (2009) Assessment for learning to teach: Appraisal of practice teaching lessons by mentors, supervisors, and students. *Journal of Teacher Education, 60(2) pp155-167.*

Tillema, H., Smith, K., & Leshem, S. (2011) Dual roles - conflicting purposes: a comparative study on perceptions on assessment in mentoring relations during practicum. *European Journal of Teacher Education,* 34(2), pp139-159.

Tynan, R., & Mallaburn, A. (2017) Consistency counts - or does it? *Teacher Education Advancement Network Journal,* 9(1), pp90-99.

Vannasche, E., & Kelchtermans, G. (2014) Teacher educators' professionalism in practice: Positioning theory and personal interpretive framework. *Teaching and Teacher Education,* 44, pp117-127.

Wiliam, D. (2014) Redesigning Schooling - 8: Principled Assessment Design. Available at: https://webcontent.ssatuk.co.uk/wp-content/uploads/2013/09/RS8-Principled-assessment-design-chapter-one.pdf (Accessed: 9 October 2018)