Reconceptualising the role of teachers as assessors: teacher assessment identity

Anne Looney, Joy Cumming, Fabienne van Der Kleij & Karen Harris

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=caie20

Routledge
Taylor & Francis Group

Reconceptualising the role of teachers as assessors: teacher assessment identity

Anne Looney[a,b*], Joy Cumming[b], Fabienne van Der Kleij[b] and Karen Harris[b*]

[a]Institute of Education, Dublin City University, Dublin, Ireland; [b]Learning Sciences Institute Australia, Australian Catholic University, Brisbane, Australia

ABSTRACT

Teachers' capabilities to conduct classroom assessment and use assessment evidence are central to quality assessment practice, traditionally conceptualised as assessment literacy. In this paper we present, firstly, an expanded conceptualisation of teachers' assessment work. Drawing on research on teacher identity, we posit that teachers' identity as professionals, beliefs about assessment, disposition towards enacting assessment, and perceptions of their role as assessors are all significant for their assessment work. We term this reconceptualisation Teacher Assessment Identity (TAI). Secondly, in support of this conceptual work, we present findings from a systematic review of self-report scales on teacher assessment literacy and teacher identity related to assessment. The findings demonstrate that such scales and previous research exploring teacher assessment practices have paid limited attention to what we identify as essential and broader dimensions of TAI. We share our reconceptualisation and analyses to encourage others to consider teacher assessment work more broadly in their research.

Introduction

The significant role of assessment in student learning has been increasingly recognised over the last three decades – not only the impact of externally-conducted accountability and high stakes certification examinations but also the need for quality classroom assessment in teacher practice (Baird, Hopfenbeck, Newton, Stobart, & Steen-Utheim, 2014; Black, McCormick, James, & Pedder, 2006; Black & Wiliam, 1998; Shepard, 2000). What and how student learning is assessed identify what is valued or important for students to learn (Looney, 2014).

Assessment, simply defined as the process to establish what students know and are able to do, is generally classified into two broad categories; assessment designed to support teaching and learning in classrooms; and assessment programmes for public reporting, certification, for selection and for system accountability (Barber & Hill, 2014). While these two broad categories are sometimes referred to as formative and summative, respectively,

these titles can be misleading. For example, on occasions, summative assessment, point-in-time judgement of student achievement, can be designed with the sole purpose of improving teaching and learning. On other occasions, evidence of student learning gathered primarily through formative assessment to provide feedback to students and teachers on next steps for teaching and learning may be included for assessment for reporting or certification.

This complex and overlapping interplay of assessment purposes and types adds to the technical complexity of assessment as a process. Further complexity arises from the current focus on assessment as a lever for school and system reform and for delivering hard-to-achieve change in teaching and learning practices (Assessment Reform Group, 1999; Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016; Herman, 2004). Black (2001) suggested at the turn of the millennium that assessment was beginning to feature more and more in the dreams of educational reformers, not just as an object of reform, but as the main instrument of reform. Looney (2014) described this optimism about the potential of educational assessment as 'viral and normative in the networks of education policy makers across the globe' (p. 234). Assessment data have become a 'publically acceptable code for quality' (Broadfoot & Black, 2004, p. 9).

Assessment literacy

In this context, it is not surprising that teacher capabilities to plan and implement quality assessment tasks, to interpret evidence and outcomes appropriate to the assessment purpose and type, and to engage students themselves as active participants in assessment of their own learning have been the subject of considerable research.

These capabilities are often referred to as assessment *literacy*, a concept first introduced by Stiggins (1991) writing in the context of the United States. Assessment literacy is usually broadly defined, encompassing both assessment knowledge and skills related to teacher practice (Popham, 2009; Stiggins, 1995) as well as use and interpretation of evidence to inform instruction, generate feedback, guide student learning, and report student achievement (Stiggins & Duke, 2008; Webb, 2002).

Stiggins (1991) suggested that assessment literacy involved understanding how to produce good achievement data on both large-scale and classroom tests, and the ability to interrogate and critique the tests or assessment approaches used and the data produced. He referred to the 'built-in alarms' (p. 535) that alert those who are assessment literate, that sound 'when an assessment target is unclear, when an assessment method misses the target, when a sample of performance is inadequate, when extraneous factors are creeping into the data, and when the results are simply not meaningful to them' (p. 535). Of note, he emphasised that knowing that there is a problem is not enough – those who are literate will demand or make changes when that alarm sounds. Fullan and Watson (2000) added a collegial dimension in their definition of assessment literacy as 'the capacity of teachers – alone and together – (a) to examine and accurately understand student work and performance data, and, correspondingly, (b) to develop classroom, and school plans to alter conditions necessary to achieve better results' (p. 457).

Some years after he first introduced the concept, Stiggins noted positive developments in the field of assessment literacy including the articulation of assessment competencies for teachers developed jointly by the National Council on Measurement in Education, and American Federation of Teachers, and the National Education Association (Stiggins, 1995).

More recently, Popham (2009) suggested that assessment literacy was needed not only to inform the assessment decisions teachers need to make, but also to challenge the high stakes accorded to accountability tests that abound in education in the US.

Beyond assessment literacy

While the concept of assessment literacy continues to be widely used (see Khadijeh & Amir, 2015; Siegel & Wissehr, 2011, for example), a number of scholars have extended the concept beyond teachers' knowledge, understanding and skills. In his investigation of teachers' *conceptions* of assessment in New Zealand, Brown (2011) draws on the earlier work of Thompson (1992) in the field of the beliefs held by teachers about mathematics teaching and learning. In addition to considering these beliefs, Thompson also proposed using the idea of teacher's conceptions 'viewed as a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences and the like' (1992, p. 130). She further suggested that such conceptions might be 'the rudiments of a philosophy of mathematics' (p. 131), although warned that in general they may be 'eclectic' (p. 135), informed by classroom experience rather than any engagement with theory or research. Brown (2011) suggests that conceptualisations act as a framework through which a teacher views, interprets and interacts with the teaching environment. Such conceptions may not be consistent with the expectations of policy, nor even with classroom practice.

Brown (2011) also notes that teachers have been reported as having multiple and conflicting *conceptions* of assessment. He points out that teachers have, in the same studies, at once agreed that assessment can improve learning yet indicated that they treat assessment as irrelevant. Other work by Brown and colleagues in Queensland, Australia, found that further differences between beliefs and practice emerge when teachers are working with different assessment purposes (Brown, Lake, & Matters, 2011).

Interestingly, work by Smith, Hill, Cowie, and Gilmore (2014) specifically explored the assessment beliefs of pre-service teachers. Hypothesising that pre-service teachers bring their personal histories to their perceptions of assessment, their study found that pre-service teachers' 'experience of formal summative assessment, such as those they have experienced in gaining qualifications, dominated their thinking and emotions' (p. 313). Assessment beliefs were shaped by their past experiences of being assessed, rather than by anything they had been taught about assessment theories or the requirements of policy.

A similar case is made by James and Pedder (2006). They argue that insufficient attention is paid to teachers' values and beliefs which provide the necessary reason for teachers to act. In their view, this failure to consider this dimension will lead to assessment for learning (the focus of their study) being seen as simply another set of techniques to add to the teaching repertoire. They suggested that 'teachers' values and the moral dimension of their practice that these values express, needs to be acknowledged in a discourse that goes beyond instrumental questions of method' (p. 131).

Conceptions of assessment presume a more complex and iterative relationship between knowledge and practice, than the relatively straightforward and apparently technical and instrumental relationship enshrined in assessment literacy. Further, *conceptions* foreground teachers' beliefs about assessment (even those that may well be non- or ir-rational) as significant in shaping classroom practice and are informed by a sociological as well as psychological perspective. Assessment is seen a sociocultural activity that involves social

interactions among stakeholders and the nature of learning itself (Broadfoot, 1996; Gipps, 2002). It occurs in a social context, influenced by national and state policies, expected learning (curriculum), pedagogical directions, and community expectations. Teacher assessment knowledge is therefore a complex structure rather than a simple set of delineated skills that can be implemented in any context.

While Stiggins promotes the concept and importance of teacher assessment literacy, he also noted barriers to teacher assessment practice, including what he called 'fear of assessment and evaluation' (1995, p. 243) arising from experiences of assessment and testing used as compliance rather than learning or measurement tools:

> For most practicing educators, this fear of assessment has been cultivated over many years as a direct result of many levels of unpleasant assessment experiences. The foundation was laid in during our youth, when our own teachers often left us wondering what would be on the test, and how to prepare for it. In our youth, assessment was frequently used to gain compliance rather than to promote improvement. (1995, p. 243)

In Stiggins' analysis, referring to the US context, this negative association with assessment continues through teacher preparation courses, and on into professional practice. There, he maintained, the negative backwash from regimes of standardised testing leaves teachers 'feeling victimised by assessment once again' (p. 243).

This emotional aspect of assessment, not included in Brown's *conceptions* of assessment (2011), is not only associated with personal assessment biography or the strictures of high-stakes accountability. It can also emerge from consideration of the role of the teacher in the assessment process. The complexity of the task of judging student work in school classrooms and in higher education is emphasised by Pryor and Crossouard (2008, 2010). It arises, they suggest, from new conceptualisations of knowledge which see learning as 'an ontological as well as an epistemological accomplishment' (Pryor & Crossouard, 2010, p. 265) because it has consequences for the identity of the learner and the teacher. In proposing their socio-cultural theory of formative assessment Pryor and Crossouard describe the latter as a 'complex and tricky process' (2008, p. 6). Drawing on their research with teachers, they noted that while task criteria were generally straightforward, the quality criteria were often problematic because they drew on knowledge held only by the teacher, and because they were generally influenced by conceptions of summative assessment. They conclude that 'the full implication of the social nature of formative assessment is that it is a site where both teacher and student identities are formed' (p. 9). Each aspect of teachers' work is associated with a particular version of the teacher:

> The different identities of the educator as assessor, teacher, subject expert and learner all involve different divisions of labour and rules shaping their interaction with students. The educator therefore teaches different definitions of themselves to the students and develops different relations with the students through them. (2008, p. 10)

Writing in 2004, Rea-Dickins noted that teachers face significant dilemmas in their assessment practices; 'sometimes torn between their role as facilitator and monitor of language development and that of assessor and judge of language performance and achievement' (p. 253).

Given that teachers' conceptions, beliefs, experiences, and feelings are all significant in their assessment work and in the role of assessor, we suggest that assessment *literacy* is not sufficient to represent the range and complexity of these dimensions.

Towards teacher assessment identity

Teacher identity is as familiar a concept in education literature as assessment literacy. It has been explored in a variety of very different ways:

- from the perspective of the multiple re-inventions that teachers undergo in the course of their professional lives (Hargreaves & Goodson, 1996; Mitchell & Weber, 1999);
- in terms of the narratives that teachers create to explain themselves and their work (Connelly & Clandinin, 1999) and the metaphors that guide and inform how they see themselves in relation to their students and professional settings (Leavy, McSorley, & Boté, 2007); and
- in response to structural or political changes in education that re-frame or change how teachers are understood and understand themselves (Mockler, 2011).

Most interpretations of teacher identity adopt a socio-cultural perspective, suggesting that it is framed and re-framed over a career and mediated by the contexts in which teachers work and live. Most also agree that teacher identity is not stable, but shifts over time as a result of a range of external contextual and internal factors (Beijaard, Meijer, & Verloop, 2004; Day, Kington, Stobart, & Sammons, 2006; Mockler, 2011). Identity development is neither simple nor linear: rather it is responsive to events and circumstances.

Mockler (2011) groups these events and circumstances under three headings: *personal* experience; *professional* context; and *external political* environment (p. 521). While these are not mutually exclusive, each has its own particular focus. *Personal* experience includes biography and personal social history. In the case of teachers, over and above other professional identities, it also includes influential personal experience of teachers, teaching and schooling, including, as Stiggins noted, their historical experience of assessment. For Shulman (1986) these 'remembrances of teachings past' (p. 12) can be valuable in guiding the work of a teacher, and generally work as a source and a heuristic for teacher decision-making. In contrast, Wiggins and McTighe (2007) take a less optimistic view of such memories suggesting that they have met many teachers in the course of their research who misunderstand the role of the teacher because they imitate the teacher practice they experienced as a student in class.

The *professional* context identified by Mockler (2011) comprises those factors that shape the classroom work of teachers – curriculum, assessment system, school climate culture and organisation, collegial relationships and experience of reform. Closely related to the latter is Mockler's third set of factors, the *political* and the *public*, including media commentary and debates about education and about teachers and their work, which reflect the degree of public trust in teachers and in the teaching profession.

Beijaard et al. (2004) identified two further important dimensions of teacher identity. They suggest that identity comprises a number of sub-identities, and that the *emotional* dimensions of identity are important. They believe that in any analysis or discussion of identity, place should be given to the question of how it 'feels' to be a teacher in the school system at any point in time. Such emotion becomes particularly significant, it is suggested, at times of educational reform (Hargreaves, 2003; Hargreaves & Lo, 2000), but tends to be under-researched.

Day et al. (2006) have also noted that insufficient attention has been paid to emotional factors in the discussion of teacher identity. In their review of the theme of personal and

professional identities, they point to the important tension between the *structures* within which teachers work and which exert influence on that work, and the *agency* of teachers, which they describe as the ability to pursue valued goals. Both are significant in the construction of teacher identity, and, suggest Day and colleagues, 'emotions are the necessary link between the social structures in which teachers work and the ways they act' (p. 613).

Mockler notes the recent preference for the term *teacher role* which is associated with the *function* of teachers, what they *do*, over consideration of teacher *identity*, associated with who teachers *are*. The former is easy to see, codify and measure. This instrumentalist perspective is reflected in debates about effective teaching, and in the professional codes or standards developed and adopted across a number of jurisdictions to regulate entry into or licence the teaching profession or to evaluate teacher performance (Mockler, 2011; Wyatt-Smith & Looney, 2016). In contrast to the relatively constrained and check-list friendly term *teacher role*, *teacher identity* is a more unwieldy concept, lending itself to multiple interpretations and perspectives.

Arguably, a similar tension can be identified between the concept of teacher assessment *literacy*, which we have suggested may be overly narrow and instrumentalist, and some of the broader and more complex dimensions of teachers' assessment work discussed above. The latter, we suggest, might be more accurately represented as *teacher assessment identity.*

Conceptualising teacher assessment identity

Adie (2013), investigating teacher participation in an online moderation process, specifically referenced the concept of *teacher assessment identity*. In her study she used the term to refer to the perception teachers had of themselves as assessors and discussed how teacher participants had some concerns that they themselves might be personally judged by others (their peers) in the discussion of how they had executed professional judgement (of students' work). Concurrent with the development of our reconceptualisation of teacher assessment identity, Xu and Brown (2016)[1] had undertaken a scoping review of 100 studies on teacher assessment literacy and developed the framework of Teacher Assessment Literacy in Practice (TALiP). A specific focus was to link education assessment research with implications for teacher education. Three of the reviewed studies addressed teachers' identity as assessor in different contexts. Xu and Brown's conceptual model presented a pyramid of six components including knowledge base, teacher conceptions of assessment and emotional interactions with assessment that might be resistant to change. The pinnacle of TALiP is 'teachers' identity (re) construction as assessors' (p. 158) as teachers develop from students to professionals interacting with others and taking on the role of assessor.

Ecclestone and Pryor (2003), researching formative assessment, also addressed the question of learner and teacher identity in assessment interactions. The focus of their analysis is the learner and the experience of assessment, how 'children bring complex dispositions to the field of educational assessment' (p. 474), and how 'assessment systems have an important impact on learning identities and dispositions as children become young adults and then adult "returners" in an increasingly long life of formal learning' (p. 472). For these researchers, this socially complex interaction between assessment systems and learner identity has implications for teacher identity on two fronts. First, effective formative assessment requires teachers to be 'intellectually curious about the understandings of learners' (p. 472), and demands an openness to the learner's experience of learning. Second, the interactions

around assessment require teachers to strike a difficult balance between being supportive of learners and being critical of them. The tensions of this dual role, the demands of responding to the complex dispositions of learners in the assessment process, and a recognition of the ontological as well as the epistemological dimensions of learning all contribute to our conceptualisation of teacher assessment identity.

Also significant in the literature on teacher identity, and for our conceptualisation of teacher assessment identity, is the concept of teacher self-efficacy. This emerged from two influential psychological theories: Locus of Control (Rotter, 1966) and Social Cognitive Theory (Bandura, 1977, 1986). The latter is particularly significant in considering a concept of teacher assessment identity as it connects self-efficacy with classroom actions. Self-efficacy of teachers has been defined as 'individual beliefs in their capabilities to perform specific teaching tasks at a specified level of quality in a specified situation' (Dellinger, Bobbett, Olivier, & Ellett, 2008, p. 4). It reflects previous experiences and beliefs, Shulman's 'remembrances of teaching past', and is a predictor of future classroom practice (Smylie, 1988) including, it can be assumed, assessment practice. Dellinger et al. (2008) made specific reference to self-efficacy beliefs as a separate construct from self-efficacy, developing a scale specifically associated with these beliefs which they describe as a 'personal belief that one is able to do what it takes (plan and act) to accomplish a task at a particular level of quality' (p. 752).

Closely aligned to the literature on the emotional dimensions of teacher identity is the concept of teacher *dispositions*, a term which emerged in the education landscape in the early 1990s, thanks in no small part to the standards for teaching published in 2001 by the National Council for Accreditation of Teacher Education (NCATE) in the United States (NCATE, 2001). The standards codified the knowledge, skills and dispositions required of teaching candidates. A definition of dispositions was offered:

> The values, commitments and professional ethics that influence behaviors towards students, families, colleagues and communities, and affect student learning, motivation, and development as well as the educator's own professional growth. Dispositions are guided by beliefs and attitudes related to values such as caring, fairness, honesty, responsibility, and social justice. (p. 53)

Describing dispositions as a 'murky concept' (p. 253), and the field as lacking an agreed definition, Schussler (2006) noted that this ambiguity did not prevent test companies from constructing tests to 'measure' the dispositions of teachers and teacher candidates in the US through a proliferation of questionnaires and other instruments. She added that 'although dispositions are addressed as entities that are conceptually distinct from skills and knowledge, the existing paradigm of measuring observable behaviours is being used to assess whether a teacher candidate possesses appropriate dispositions' (p. 256). Schussler contends that it is neither useful nor accurate to separate dispositions from understanding and skills. Instead she proposes a view of dispositions as a point of convergence where the external influences such as the system demands for assessment, or curriculum requirements or school organisation for example, meet the teacher's individual internal schemata of beliefs, values, history and experience and a point of inception for teacher thinking and actions. As 'filters' between the external requirements of assessment work in classroom and in the education system, and the beliefs and values held by teachers about student learning and the quality of student work, some of which have in turn been shaped by educational histories, dispositions are a significant component of assessment identity.

This conceptualisation of teachers as assessors has demonstrated that a broader dimensionalisation of teachers' work in assessment is needed to inform future research and quality teacher assessment work. Many different aspects of teacher assessment knowledge, confidence, personal disposition and emotional engagement with assessment affect teachers' own assessment practices in classrooms with learners. The conceptual analysis evolved from an empirical large-scale project being undertaken by the authors to examine, and hence identify appropriate measures of, teacher assessment work. Having completed the conceptual analyses, we then turned to examination of existing scales on teacher assessment literacy and teacher identity related to assessment. Findings from a systematic review of these scales within the new conceptualisation of teacher assessment identity follow.

Analysis of existing scales on teacher assessment literacy and teacher identity related to assessment

As noted, the theoretical conceptualisation of teacher assessment identity emerged from an externally-funded large scale research project to examine the relationship between teachers' assessment practices and student learning and the need to identify an appropriate instrument to gauge teachers' knowledge of and engagement with assessment. For the project, such an instrument needed to be broadly encompassing of the areas identified in the conceptual analysis that might relate to teacher assessment practice.

As the starting point, we identified a number of scales developed for use in teacher assessment literacy and teacher identity research. We confined our review and analysis to self-report scales as such a format was not only necessary for our large-scale project but would also enable adaptation in recognition of the important influence of different cultural contexts on teacher assessment. The review of existing scales required two steps: firstly, appropriate scale and item identification; and secondly, development of a coding schema of items against both our conceptual framework and aspects of teacher assessment work. The method employed to review and code existing scales, and findings of the analyses are reported in the following sections.

Method: determining sources and trial development of coding schema

A recent systematic review of teacher assessment literacy measures reported in published research (Gotch & French, 2014) provided an initial source of scales. Gotch and French searched research publications in comprehensive education and social science databases from 1991, following the introduction of 'assessment literacy' as an important component of teachers' work, to 2012. Search terms included assessment literacy, assessment, teacher competence/y, and teacher understanding. They identified 36 instruments from 50 publications, with different response formats such as objective tests, self-report scales and rubrics. While Gotch and French's purpose was to analyse the scales identified through their systematic search from a psychometric perspective, the sources and scales they identified formed the initial starting point for our consideration of dimensions of existing assessment instruments.

Trial and initial coding framework
The first phase of the analysis was to examine dimensions of teacher assessment work addressed in ten of the scales identified by Gotch and French (2014, Appendix A, pp.

46–58) as teacher self-reports of assessment literacy (Arter & Busick, 2001; Bandalos, 2004; DeLuca & Klinger, 2010; Hambrick-Dixon, 1999; Kershaw, 1993; Mertler, 2000; O'Sullivan & Johnson, 1993; Vanden Berk, 2005; Zhang & Burry-Stock, 1994, 2003). An initial framework for coding categories was developed, based on the knowledge of the four members of the research team, all of whom have expertise in the field of assessment. Initial coding of items from the ten scales focused both on dimensions of teachers' work as well as the topics they addressed. Two team members independently examined the ten scales and trialled item coding for 206 items. After initial coding, discrepancies were discussed and issues/nuances in the coding scheme were identified and discussed by the four researchers. The development of the coding scheme was iterative, taking place in parallel to the identification of relevant scales, with several iterations and revisions occurring before the final coding scheme was determined. All items were then coded according to the final coding scheme.

Identification of additional sources and items

As our conceptualisation of teacher assessment identity is more comprehensive than assessment literacy, the focus of the review by Gotch and French (2014), a further database search for scales related to *teachers*, *teacher identity* and *assessment* was undertaken. Additional self-report scales were identified by searching scientific journal articles or book chapters using Google Scholar and online databases such as ERIC. Search terms used included, for example, 'teacher' in combination with 'assessment literacy', 'assessment identity' or 'classroom assessment', and 'questionnaire', 'survey', 'scale' or '(audit) instrument'. Prominent journals in the area of assessment and measurement were searched.

A requirement for final analysis was that scale items were available in the publication itself, or in supplementary materials. Items had to be related to teachers' assessment knowledge/skills/practices, confidence or disposition, not, for example, what students typically do in classrooms. Some scales included items that were not specific to assessment, but applied, for example, to instruction. In this case, the scale was included, but only items relevant to assessment were selected. Items that had a very strong focus on psychometric test development were also excluded from the item coding. Such items addressed teachers' technical skills in this area, such as the ability to conduct item and test analyses or calculate z-scores for a test. Such conceptualisations of assessment literacy dominated US measurement in the 1990s (DeLuca, LaPointe-McEwan, & Luhanga, 2016), but are not contemporary assessment skills teachers would be universally expected to develop or hold (Joint Committee on Standards for Education Evaluation [JCSEE], 2015). However, items addressing concepts such as reliability and validity were included.

The additional search process resulted in 18 new scales from 10 publications, providing a further 312 items for analysis (Brown, 2004; Dellinger et al., 2008; James & Pedder, 2006; Leighton, Gokiert, Cor, & Heffernan, 2010; Lysaght & O'Leary, 2013; Panadero, Brown, & Courtney, 2014; Pat-El, Tillema, Segers, & Vedder, 2013; Schulte, Edick, Edwards, & Mackiel, 2004; Smith et al., 2014; Young & Jackman, 2014). The results of the additional search and details on all scales used for analyses are presented in Appendix 2, available as additional online material. In total, 28 self-report scales based on 19 instruments described in 20 publications related to teacher assessment literacy and teacher identity related to assessment and 518 items were identified as suitable for coding (sources indicated with * in Reference list).

Determination of final coding scheme

Developing a coding schema went through many iterations. We considered the elements of assessment each item addressed, initially listing different aspects in an exhaustive list. As we moved from initial coding of items to coding consensus and the final coding scheme, it became evident that no singular classification of items, separating elements in a linear fashion, would be sufficient to capture teacher assessment work. The coding of each item needed to reflect multiple dimensions of the role of teachers as assessors (identity) as well as different aspects and contexts of teacher assessment work, but in a manner that provided overall cohesion. As elaborated below, the final coding scheme that emerged was a framework whereby items were coded according to four identified classifications: (1) the teacher assessment identity dimension; (2) aspect of assessment; (3) purpose of assessment; and (4) external or classroom-based contexts of assessment.

Classification 1 addressed the *dimensions* of teacher assessment identity that related teacher assessment *knowledge and skills*, their *beliefs* about assessment, their *confidence* in their assessment knowledge, skills and practices, and their overall *disposition* to assessment, including how they engage with assessment or view their assessment work or role.

Classification 2 addressed traditional *aspects* of teachers' assessment practice in seven subclasses: items that related to *conceptual/theoretical frameworks* of assessment; assessment *purpose and use* of outcomes; assessment *design and implementation*; *interpretation and use* of assessment information; *collaboration* with others in and out of school; *connections* between assessment and instruction/curriculum; and developing a *classroom assessment environment* for students.

Classification 3 addressed distinctions often seen in assessment literature regarding the *purpose* of assessment in terms of *formative*, *summative* or *diagnostic* assessment, given the focus in current assessment research on the influence of teachers' enactment of formative assessment and assessment for learning and student learning outcomes. When the purpose of an item was unclear or unspecified, the item was coded as unspecified.

The final classification, Classification 4, referred to the assessment *context*, distinguishing between teacher- and externally-developed assessment instruments or processes. When the context was not clear, items were coded as teacher-developed.

Classification 2 involved the most complex analysis. The subclass codes emerged initially through examination of scales identified from Gotch and French's (2014) review and classifying items according to focus. Additional areas of teacher assessment practice identified in research as good practice were also incorporated under Classification 2, even if not present in the reviewed scales and items. For example, work from formative assessment and assessment for learning identifies the importance of feedback, collaboration among teachers and students sharing expectations and understandings of quality, and developing student agency in assessment (Assessment Reform Group, 2002; Baird et al., 2014; Black & Wiliam, 1998; Fullan & Watson, 2000; Wyatt-Smith & Klenowski, 2013). The influence of teacher assumptions about student learning progressions also needs consideration (Popham, 2011).

An identified issue in much assessment literature is the focus on generic assessment strategies without an evidentiary basis for their applicability in different contexts (Coffey, Hammer, Levin, & Grant, 2011; Cowie & Moreland, 2015; Svanes & Skagen, 2016). Differentiating assessment by subject areas, by student level of schooling and to meet the needs of diverse students was also identified by the research team as significant for teachers' assessment work (Abedi, 2010; Cumming, 2012; European Agency for Development in

Special Needs Education, 2009; Trumbull & Lash, 2013). Teachers with subject specialisations may be responsible for students of different ages but not confident in their assessment knowledge and skills for these different students, or have different dispositions to assessment for students of different ages. Effective assessment for students must be responsive to different critical stages of learning, such as students in the middle years who are highly self-aware, needing to build confidence, develop autonomy and sense of strengths (Wyatt-Smith, Adie, Van der Kleij, & Cumming, in press). Teachers in contexts such as elementary school may have responsibility for different subjects with the same age cohorts. Are teachers necessarily knowledgeable about and confident in appropriate assessment strategies in different subjects (Hodgen & Marshall, 2005)? In current equity policy contexts and inclusive schooling, teachers express anxiety about creating fair and equitable assessments for all students including students with disability (Forlin, Keen, & Barrett, 2008). Different strategies may be needed for effective assessment with students with languages different from the language of instruction (Trumbull & Lash, 2013).

Items in Subclasses 1 and 2, conceptual/theoretical frameworks and purpose and use of outcomes (validity) were items that addressed global assessment concepts.

Items coded as Subclass 3, design and implementation, addressed activities in assessment construction, designing assessments to align with curriculum expectations, administration related to implementation of standardised assessment procedures, developing and using scoring models such as rubrics, judging student performance against rating scales or criteria (marking), and tailoring assessment processes suited for students with diverse learning needs.

Subclass 4 encompassed items that looked at how to interpret or use information for instructional decisions (planning teaching), providing feedback to students, assigning grades (determining an overall grade based on a student's portfolio of evidence or multiple pieces of assessment work), and reporting student achievement to parents or students.

Subclass 5 was developed specifically to identify items that reflected teacher work as assessor as a partnership with colleagues.

Subclass 6 emerged to encompass items that investigated links between assessment and curriculum and instruction generally, in terms of teachers' assessment-related pedagogical content knowledge and discipline area knowledge, particularly related to classroom-based activities, rather than the formal assessment practices coded as Subclass 3. Items that investigated teacher knowledge and understanding with respect to specified learning progressions were coded under learning progressions. Questioning and techniques identified as assessment for learning intended to assess students' learning informally in ongoing classroom interactions constituted the final subclass in this level. Identification and placement of these last items caused considerable discussion, as they could also be related to formative assessment purposes such as feedback under Subclass 4. Initially some items that involved teacher questioning were coded under Subclass 3, as assessment construction. However, Subclass 4 resulted to enable separation of assessment practices aligned with formal expectations and more fine-grained informal assessment practices integrated with instruction.

Items coded under Subclass 7 reflect principles related to student engagement and agency in assessment. These included sharing and/or negotiating goals and expectations (sharing direction of learning), engagement of students in self-assessment, and peer assessment.

Thus, in the final coding scheme, an individual item from a scale could be coded, for example, as related to teacher confidence in their assessment practices (Classification 1),

interpretation and use of information (Classification 2), for formative assessment purposes (Classification 3), in teacher-developed assessment (Classification 4). All items were coded using the final coding scheme. The detailed coding scheme and results are provided in Appendix 1.

Coding reliability

Initial inter-rater agreement between the two raters was 75% for the 206 items from ten scales in Gotch and French (2014). Initial discrepancies related to ambiguity for interpretation of items. Many items related to confidence interrelated with knowledge and skills, or affective aspects. For example, a teacher's response to '*I am confident using assessments for diagnostic evaluation*', by itself, could not determine whether a lack of confidence was linked to their level of knowledge/skills/practices of diagnostic evaluation. This interrelationship was not usually able to be investigated in existing scales.

Discussions of discrepancies and nuances clarified defining aspects of the final coding scheme which improved inter-rater agreement. The 18 scales from the second search stage were then coded, with inter-rater agreement for these scales of 88%. Disagreement in coding was again due principally to double-barrelled questions or ambiguities presented by possibilities for multiple coding of items. For example, an item coded as Disposition, '*My assessment practices have little impact on student achievement*' (self-rated from Strongly Disagree to Strongly Agree), was ambiguous as to whether a teacher was rating their belief of the role of assessment in their practice (not valuing or a strong objection to assessment) or their confidence with their assessment practices. Each item for which there were different initial rater codings was assigned a final consensus code following discussion.

Results and discussion

Results of the analysis of items from existing scales are provided in Appendix 1. As might be expected, given the focus of much research on assessment literacy and despite our search to identify more broadly conceptualised scales that identified any aspect of teacher identity and assessment work or role, most scales and items focused on those areas strongly associated with assessment literacy. For the coded items, 87% addressed teacher-developed assessment, possibly reflecting exclusion of items focused on psychometric skills, but also reflecting the overall focus of research on teacher assessment practice.

Over three-quarters of the items (78%) focused on the dimension of teacher assessment knowledge, skills and practices (Classification 1). Nearly all of these were in the context of teacher-developed assessment practices. Twelve per cent of items related to teacher confidence in assessment, again predominantly for teacher-developed assessment activities. Less than 10% of items were identified as relating to teacher dispositions in assessment, with disposition compounded in only a small number of confidence items. These included items such as 'My assessment practices have little impact on student achievement' (Vanden Berk, 2005, p. 144) or 'I consider the most worthwhile assessment to be assessment that is undertaken by the teacher' (James & Pedder, 2006, p. 118). More than half of the last group of items were related to conceptual or theoretical assessment frameworks with the remainder spread lightly across other assessment aspects.

The assessment content focus of items (Classification 2) was more distributed: 19% of items investigated teachers' engagement with conceptual and theoretical frameworks;

13% were associated with assessment purpose and use of outcomes; 21% addressed assessment design and implementation, including administration of assessment, and item and assessment construction; 22% examined teacher interpretation and use of information; 6% addressed connections between assessment and instruction and curriculum; and, 18% related to development of the classroom assessment environment for students.

The purpose of more than half the items (Classification 3) across all dimensions and aspects was coded as unspecified. Twenty-two per cent and 24% of items overall addressed formative or summative assessments, respectively. For items coded as related to knowledge, skills and practices under Classification 1, 26% were identified as related to formative assessment, 20% were related to summative assessment (of which 63% were for teacher-developed assessment), and only 4 items, that specifically identified a diagnostic purpose, for example, 'Diagnostic information from standardised tests is used to identify strengths and needs in teaching and learning' (Lysaght & O'Leary, 2013, p. 11), were coded as related to diagnostic assessment. Nearly half of these items (48%) did not have a specific focus.

It is evident that the existing scales encompass a range of aspects of teacher work in implementing assessment activities, especially in their classroom assessment. However, even when viewed from the established perspective of assessment literacy, there are gaps. The tendency of most scales was to have a generic focus; teacher assessment knowledge, skills and practices were not distinguished according to the age of the learner or system stages such as primary or secondary schooling. Most items that asked teachers about their knowledge did not necessarily seek information on their confidence in applying their knowledge, nor the extent to which they do so. From our perspective, the issue became the gaps and omissions in the items coded – missing areas that could be considered important within existing definitions of assessment literacy, and the additional conceptual dimensions we identify for teacher assessment identity.

While some scales were contextualised within a subject area, generic scales did not seek information about assessment within specific pedagogical content knowledge frameworks. We only identified three items relating to such knowledge, for example, 'Identify a curriculum area that lends itself to performance assessment' (O'Sullivan & Johnson, 1993, p. 22). Teachers generally teach more than one subject area, especially in primary schools. Questions arise as to whether teachers consider their knowledge, and confidence as equally strong or weak across different subject contexts, or whether their beliefs about assessment differ between mathematics for example, and areas such as art and music.

Diversity, even in its broadest definitions, was also a gap. Only seven items were coded as addressing flexibility or adaptability of assessment for students with diverse needs, whether due to language background, disability, or range of achievement levels within classrooms, such as 'Establishing student expectations for determining grades for special education students' (Zhang & Burry-Stock, 2003, p. 342) and 'Accommodating assessment for ESL students' (DeLuca & Klinger, 2010, p. 432).

While items were identified that related to emphases in current research on formative assessment and assessment for learning, including use of evidence from a range of sources to direct teaching and improve student learning, there were still omissions. Major omissions identified in the items coded, were items about teacher assessment practice in collaboration with other teachers, valuable in developing teachers' professional assessment knowledge and deeper understanding of demonstrated quality of student work. Aspects of formative assessment and assessment for learning addressing sharing learning goals, and peer and

self-assessment, were relatively limited. Very few items addressed teacher and student agency in assessment.

Most importantly, from our conceptual framework for teacher assessment identity, while a proportion (10%) of items were coded as related to teachers' assessment disposition, very few of these related to engagement with specific assessment knowledge and skills or confidence, such as 'I want to learn new formative assessment strategies' (Young & Jackman, 2014, p. 406). This is an area that needs to be addressed more deeply.

Conclusion

The principal focus and significant contribution of this paper is the reconceptualision of teachers' assessment work and the proposal of a new concept, beyond assessment literacy, of teachers' assessment identity. We have also identified that while existing scales, and related research which focus on strategic and technical assessment skills have at least face validity on many aspects of teachers' work as understood in assessment literacy, their overall representation of teachers' assessment work is limited. They ignore the more complex dimensions of this assessment work. Recent work by Xu and Brown (2016) addresses these complex dimensions and points to the significance of identity in teachers' assessment work, and in particular, in taking on the role of assessor. Our conceptualisation goes further; we propose a dynamic and interactive teacher assessment identity constituted by beliefs, feelings, knowledge and skills.

While we are undertaking further research to develop a Teacher Assessment Identity Instrument to address the gap in current scales, we present these conceptual and analytical findings to promote a broader perspective in research by assessment and, more generally, education researchers on teacher assessment practices and engagement.

In summary, we propose that the role of the teacher as an assessment practitioner goes beyond what has been previously identified through conceptualisations of assessment literacy and what has been established through existing scales. This previous work has focused on teacher knowledge and capabilities as distinct bases of practice. It is our contention that when teachers assess more is in play than simply knowledge and skills. They may have knowledge of what is deemed effective practice, but not be confident in their enactment of such practice. They may have knowledge, and have confidence, but not believe that assessment processes are effective. Most importantly, based on their prior experiences and their context, they may consider that some assessment processes should not be a part of their role as teachers and in interactions with students. Teachers can, quite literally, have mixed feelings about assessment. These interlinked dimensions of teacher assessment identity are represented in Figure 1 below.

We include the various dimensions of assessment literacy as described by Stiggins (1991, 1995) and Popham (2009) under the heading 'I know'. How a teacher feels about assessment, the emotional dimensions of assessment identity as discussed by Beijaard et al. (2004), is included under the 'I feel' category, but the categories are intentionally linked to highlight the interconnectedness of the disparate dimensions. The very particular and complex 'role' dimensions of teacher assessment identity, reflected in the work of Pryor and Croussouard (2010) and Ecclestone and Pryor (2003), are identified as a further dimension, although this dimension, together with assessment 'feelings' and 'knowledge' is informed and shaped by teachers' assessment dispositions, those points of inception to use Schussler's terminology

Figure 1. Reconceptualising teacher assessment identity.

(2006). Drawing on Shulman (1986), Broadfoot (1996) and Mockler (2011), we propose that teachers' beliefs about assessment, some of which are informed by their personal assessment histories, also shape assessment identity. Finally, we consider that teachers' sense of self-efficacy in assessment work, and the degree to which they feel in control of their practice, drawing on the work of Bandura (1977, 1986), are also significant for assessment identity. These cannot be considered in isolation from teachers' understanding and experience of their role as assessor, particularly in the light of the work of Day et al. (2006) on teacher agency, and the importance of teachers' sense that they can pursue valued goals. We have presented these dimensions as associated with a teacher subject, rather than as abstractions, to emphasise our contention 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.

Following this work to date, we are now in the process of developing a new instrument, the Teacher Assessment Identity Instrument (TAII) that will be cognisant of the multiple dimensions and contexts of teachers' assessment enactment that we have conceptualised. The work is being undertaken within the Learning Sciences Institute Australia, Australian Catholic University, supported by the National Council for Curriculum and Assessment, Ireland and the National Institute of Education, Singapore. The intention is that the Instrument will have international applicability, albeit with the need for local adaptations in recognition of cultures and contexts of different education systems. Challenges to date have been designing a new framework, based on Figure 1, for piloting and validation, in combination with user-friendly self-report response forms. Once the TAII has been validated in three international educational systems, it will be used in our current project to examine the import of teacher assessment identity in conjunction with classroom practices and student learning, and for initial teacher education.

We present the findings in this paper to encourage international colleagues to engage with the reconceptualised representation of teacher as assessor and to broaden research on teacher assessment capability and engagement. This reconceptualisation of teacher assessment identity encompasses not only a range of assessment strategies and skills, and even confidence and self-efficacy in undertaking assessment, but also the beliefs and feelings

about assessment that will inform how teachers engage in assessment work with students, and focuses not simply on what teachers do, but on who they are.

Note

1. Submitted January 2016, while this paper was under review.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Notes on contributors

Anne Looney commences as founding Executive Dean, Institute of Education, Dublin City University, Dublin, in 2017, following a role as Interim Chair of the Higher Education Authority, Ireland. Previously, Looney was professorial research fellow at the Learning Sciences Institute Australia, Australian Catholic University (ACU), Australia, in 2014–2015. She served as Chief Executive of the National Council for Curriculum and Assessment, Ireland, for fifteen years. Her research publications include curriculum and assessment, school ethos, civic and political education and the work of teachers.

Joy Cumming is Research Director of the Assessment, Evaluation and Student Learning research area in the Learning Sciences Institute Australia (ACU). Her core research focus is educational assessment, in particular, equity and social justice in assessment, including work in education law that examines the impact of educational policy and legislation in assessment and accountability on students.

Fabienne van der Kleij is a research fellow in the Assessment, Evaluation and Student Learning research area in the Learning Sciences Institute Australia (ACU). Her main research interests are formative assessment and feedback. Her work has involved several systematic reviews of formative assessment and feedback. Current research is focused on student perceptions and engagement with feedback to improve their learning.

Karen Harris worked in the Assessment, Evaluation and Student Learning research area in the Learning Sciences Institute Australia (ACU) during the development of this research. She has had an extensive career in educational assessment and measurement, including roles at National Institute of Education, Nanyang Technological University, Singapore, Educational Testing Services, Princeton, US, and the Australian Council for Education Research.

References

References marked with an asterisk (*) indicate scales relating to teacher assessment.

Abedi, J. (2010). Research and recommendations for formative assessment with English language learners. In H. Andrade & C. Cizek (Eds.), *Handbook of formative assessment* (pp. 181–197). New York, NY: Routledge.

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice, 20*, 91–106. doi:10.1080/0969594X.2011.650150

*Arter, J. A., & Busick, K. U. (2001). *Practice with student-involved classroom assessment: A workbook and learning team guide*. Portland, OR: Assessment Training Institute.

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. England: University of Cambridge School of Education. Retrieved from http://www.nuffieldfoundation.org/sites/default/files/files/beyond_blackbox.pdf

Assessment Reform Group. (2002). *Assessment is for learning: 10 principles. Research-based principles to guide classroom practice*. Retrieved from http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf

Baird, J., Hopfenbeck, T., Newton, P., Stobart, G., & Steen-Utheim, A. (2014). *State of the field review: Assessment and learning*. Report for the Norwegian Knowledge Centre for Education. Retrieved from http://taloe.up.pt/wp-content/uploads/2013/11/FINALMASTER2July14Bairdetal2014AssessmentandLearning.pdf

*Bandalos, D. L. (2004). Can a teacher-led state assessment system work? *Educational Measurement: Issues and Practice, 23*, 33–40. doi:10.1111/j.1745-3992.2004.tb00157.x

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review, 84*, 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Upper Saddle River, NJ: Prentice-Hall Inc.

Barber, P., & Hill, M. (2014). *Preparing for a renaissance in assessment*. London: Pearson. Retrieved from https://research.pearson.com/content/plc/prkc/uk/open-ideas/en/articles/preparing-for-a-renaissanceinassessment/_jcr_content/par/articledownloadcompo/file.res/Preparing_for_a_Renaissance_in_assessment.pdf

Beijaard, D., Meijer, P. C., & Verloop, N. (2004). Reconsidering research on teachers' professional identity. *Teaching and Teacher Education, 20*, 107–128. doi:10.1016/j.tate.2003.07.001

Black, P. (2001). Dreams, strategies and systems: Portraits of assessment past, present and future. *Assessment in Education: Principles, Policy & Practice, 8*, 65–85. doi:10.1080/09695940120033261

Black, P., McCormick, R., James, M., & Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education, 21*, 119–132. doi:10.1080/02671520600615612

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*, 7–74. doi:10.1080/0969595980050102

Broadfoot, P. (1996). *Education, assessment and society: A sociological analysis*. Bristol: Open University Press.

Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles Policy and Practice, 11*, 7–26. doi:10.1080/0969594042000208976

*Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*, 301–318. doi:10.1080/0969594042000304609

Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters, 3*, 45–70.

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27*, 210–220.

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching, 48*, 1109–1136. doi:10.1002/tea.20440

Connelly, F. M., & Clandinin, D. J. (1999). *Shaping a professional identity: Stories of education practice*. New York, NY: Teachers College Press.

Cowie, B., & Moreland, J. (2015). Leveraging disciplinary practices to support students' active participation in formative assessment. *Assessment in Education: Principles, Policy & Practice, 22*, 247–264, doi:10.1080/0969594X.2015.1015960

Cumming, J. (2012). *Valuing students with impairments: International comparisons of practice in educational accountability*. Dordrecht: Springer.

Day, C., Kington, A., Stobart, G., & Sammons, P. (2006). The personal and professional selves of teachers: Stable and unstable identities. *British Educational Research Journal, 32*, 601–616. doi:10.1080/01411920600775316

*Dellinger, A. B., Bobbett, J. J., Olivier, D. F., & Ellett, C. D. (2008). Measuring teachers' self-efficacy beliefs: Development and use of the TEBS-Self. *Teaching and Teacher Education, 24*, 751–766. doi:10.1016/j.tate.2007.02.010

*DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*, 419–438. doi:10.1080/0969594X.2010.516643

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability, 28*, 251–272

Ecclestone, K., & Pryor, J. (2003). 'Learning careers' or 'assessment careers'? The impact of assessment systems on learning. *British Educational Research Journal, 29*, 471–488. doi:10.1080/0141192032000099324

European Agency for Development in Special Needs Education. (2009). *Assessment for learning and pupils with special educational needs*. Retrieved from https://www.european-agency.org/sites/default/files/assessment-for-learning-and-pupils-with-special-educational-needs_assessment_for_learning_en.pdf

Forlin, C., Keen, M., & Barrett, E. (2008). The concerns of mainstream teachers: Coping with inclusivity in an Australian context. *International Journal of Disability, Development and Education, 55*, 251–264.

Fullan, M., & Watson, N. (2000). School-based management: Reconceptualizing to improve learning outcomes. *School Effectiveness and School Improvement, 11*, 453–473. doi:10.1076/sesi.11.4.453.3561

Gipps, C. (2002). Sociocultural perspectives on assessment. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century* (pp. 73–83). Oxford: Blackwell Publishing Ltd.

Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice, 33*, 14–18. doi:10.1111/emip.12030

*Hambrick-Dixon, P. J. (1999, April). *Meeting the challenges to urban school reform: Assessment portfolios for teachers' professional development*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, QC.

Hargreaves, A. (2003). *Teaching in the knowledge society. Education in the age of insecurity*. New York, NY: Teachers College Press.

Hargreaves, A., & Goodson, I. (1996). Teachers' professional lives: Aspirations and actualities. In I. Goodson & A. Hargreaves (Eds.), *Teachers' professional lives* (pp. 1–27). London: Falmer.

Hargreaves, A., & Lo, L. N. (2000). The paradoxical profession: Teaching at the turn of the century. *Prospects, 30*, 167–180.

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review, 17*, 50–62. doi:10.1016/j.edurev.2015.12.002

Herman, J. (2004). The effects of testing on instruction. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 141–167). New York, NY: Teachers College Press.

Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. *The Curriculum Journal, 16*, 153–176. doi:10.1080/09585170500135954

*James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *Curriculum Journal, 17*, 109–138. doi:10.1080/09585170600792712

Joint Committee on Standards for Education Evaluation (JCSEE). (2015). *Classroom assessment standards for PreK-12 teachers* [Kindle version]. Retrieved from https://www.amazon.com.au/dp/B00V6C9RVO/ref=docs-os-doi_0

*Kershaw, I. (1993). *Ohio vocational education teachers' perceived use of student assessment information in educational decision-making* (Order No. 9411987). Available from ProQuest Dissertations & Theses Global. (304060852). Retrieved from http://search.proquest.com/docview/304060852?accountid=8194

Khadijeh, B., & Amir, R. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education, 3*, 139–146. doi:10.5296/ijele.v3i1.6887

Leavy, A. M., McSorley, F. A., & Boté, L. A. (2007). An examination of what metaphor construction reveals about the evolution of pre-service teachers beliefs about teaching and learning. *Teaching and Teacher Education, 23*, 1217–1233. doi:10.1016/j.tate.2006.07.016

*Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom–versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice, 17*, 7–21. doi:10.1080/09695940903565362

Looney, A. (2014). Assessment and the reform of education systems. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 233–247). Heidelberg: Springer.

*Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies, 32*, 217–232. doi:10.1080/03323315.2013.784636

*Mertler, C. A. (2000). Teacher-centered fallacies of classroom assessment. *Mid-Western Educational Researcher, 13*, 29–35.

Mitchell, C., & Weber, S. (1999). *Reinventing ourselves as teachers: Beyond nostalgia.* London: Falmer.

Mockler, N. (2011). Beyond 'what works': Understanding teacher identity as a practical and political tool. *Teachers and Teaching, 17*, 517–528. doi:10.1080/13540602.2011.602059

National Council for Accreditation of Teacher Education (NCATE). (2001). *Standards for professional development schools.* Washington, DC: NCATE. Retrieved from http://www.ncate.org/documents/pdsstandards.pdf

*O'Sullivan, R. G., & Johnson, R. L. (1993, April). *Using performance assessments to measure teachers' competence in classroom assessment.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

*Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice, 21*, 365–383. doi:10.1080/0969594X.2014.919247

*Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of assessment for learning questionnaires for teachers and students. *The British Journal of Educational Psychology, 83*, 98–113. doi:10.1111/j.2044-8279.2011.02057.x

Popham, J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice, 48*, 4–11. doi:10.1080/00405840802577536

Popham, J. (2011). *Transformative assessment in action: An inside look at applying the process.* Alexandria, VA: ASCD.

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education, 34*, 1–20. doi:10.1080/03054980701476386

Pryor, J., & Crossouard, B. (2010). Challenging formative assessment: Disciplinary spaces and identities. *Assessment & Evaluation in Higher Education, 35*, 37–41. doi:10.1080/02602930903512891

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*, 249–258. doi:10.1191/0265532204lt283ed

Rotter, J. B. (1966). Internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1–28.

*Schulte, L., Edick, N., Edwards, S., Mackiel, D. (2004). The development and validation of the teacher dispositions index. *Essays in Education, 12*(4). Retrieved from http://www.usca.edu/essays/vol122004/schulte.pdf

Schussler, D. (2006). Defining dispositions: Wading through murky waters. *The Teacher Educator, 41*, 251–268. doi:10.1080/08878730609555387

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*, 4–14. doi:10.3102/0013189X029007004

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*, 4–14. doi:10.3102/0013189X015002004

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education, 22*, 371–391. doi:10.1007/s10972-011-9231-6

*Smith, L. F., Hill, M. F., Cowie, B., & Gilmore, A. (2014). Preparing teachers to use the enabling power of assessment. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 303–323). Dordrecht: Springer.

Smylie, M. (1988). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal, 25*(1), 1–30. doi:10.3102/00028312025001001

Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan, 72*, 534–539.

Stiggins, R. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*, 238–245.

Stiggins, R., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan, 90*, 285–291. Retrieved from http://www.pdkmembers.org/members_online/publications/Archive/pdf/k0812sti.pdf

Svanes, I. K., & Skagen, K. (2016). Connecting feedback, classroom research and Didaktik perspectives. *Journal of Curriculum Studies*. Advance online publication. doi:10.1080/00220272.2016.1140810

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York, NY: MacMillan.

Trumbull, E., & Lash, A. (2013). *Understanding formative assessment: Insights from learning theory and measurement theory*. San Francisco, CA: WestEd. Retrieved from http://www.wested.org/online_pubs/resource1307.pdf

*Vanden Berk, E. (2005). *Improving the evaluation of students through teacher training: An investigation of the utility of the student evaluation standards* (Order No. 3202970). Available from ProQuest Dissertations & Theses Global. (304987564). Retrieved from http://search.proquest.com/docview/304987564?accountid=8194

Webb, N. L. (2002, April). *Assessment literacy in a standards-based urban education setting*. Paper presented at American Educational Research Association Annual Conference, New Orleans, LA. Retrieved from http://www.wcer.wisc.edu/archive/mps/AERA2002/AssessmentliteracyNLWFinal32602.pdf

Wiggins, G., & McTighe, J. (2007). *Schooling by design: Mission, action and achievement*. Alexandria, VA: ASCD.

Wyatt-Smith, C., Adie, L., Van der Kleij, F., & Cumming, J. (in press). Assessment. In D. Pendergast, K. Main, & N. Bahr, *Teaching middle years: Rethinking curriculum, pedagogy and assessment* (3rd ed.). Sydney: Allen & Unwin. (Accepted 4 July 2016)

Wyatt-Smith, C., & Klenowski, V. (2013). Explicit, latent and meta-criteria: Types of criteria at play in professional judgement practice. *Assessment in Education: Principles, Policy & Practice, 20*, 35–52. doi:10.1080/0969594X.2012.725030

Wyatt-Smith, C. M., & Looney, A. (2016). Professional standards and the assessment work of teachers. In D. Wyse, L. Hayward, & J. Pandaya (Eds.), *The Sage handbook of curriculum, pedagogy and assessment* (pp. 805–820). London: Sage.

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162. doi:10.1016/j.tate.2016.05.010

*Young, J. E. J., & Jackman, M. G. (2014). Formative assessment in the Grenadian lower secondary school: Teachers' perceptions, attitudes and practices. *Assessment in Education: Principles, Policy & Practice, 21*, 398–411. doi:10.1080/0969594X.2014.919248

*Zhang, Z., & Burry-Stock, J. A. (1994). *Assessment practices inventory*. Tuscaloosa, AL: The University of Alabama.

*Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*, 323–342. doi:10.1207/S15324818AME1604_4

Appendix 1: Coding schema and outcomes for analysis of 28 self-report scales related to teacher assessment practices aligned to conceptual framework

Classification 3: Purpose	Classification 1: Dimension												
	Knowledge, skills, practices				Disposition								
					Confidence								
	Formative	Summative	Diagnostic	Unspecified	Formative	Summative	Diagnostic	Unspecified	Formative	Summative	Diagnostic	Unspecified	Total
Classification 2: Aspect													
1. Conceptual/Theoretical framework	4	9(14)		33(1)		2		6(2)	2			22(3)	78(20)
2. Purpose & use of outcomes (validity)	1(2)	12(4)		31(5)		1(3)		4		1		2(3)	52(17)
3. Design & implementation													
a) Item and assessment construction	2	3(1)		22				6		1		1	35(1)
b) Alignment		2(5)		11(2)									13(7)
c) Administration	3	6(1)		10		2		1				1	23(1)
d) Scoring models		2		3				2					7
e) Marking	(3)	7		1		5							13(3)
f) Students' diverse needs	1	1			1			4					7
4. Interpretation & use of information		(2)		1									1(2)
a) Instructional decisions	6	1	1(2)	8(1)	3		1	4				5	29(3)
b) Feedback	14	6(2)		17				2	1				40(2)
c) Grading		3(5)		2		1		1					7(5)
d) Reporting		6(2)		7		7		1					21(2)
5. Collaboration				2				1					
6. Connections between assessment and instruction/curriculum												1	4

Continued

Appendix 1: (*Continued*).

Classification 3: Purpose	Classification 1: Dimension												
	Knowledge, skills, practices				Confidence				Disposition				Total
	Formative	Summative	Diagnostic	Unspecified	Formative	Summative	Diagnostic	Unspecified	Formative	Summative	Diagnostic	Unspecified	
a) Pedagogical content knowledge for assessment	1	1		1									3
b) Curriculum area knowledge				3									3
c) Learning progressions				1									1
d) Questioning & techniques (AFL) used in classrooms (to gather evidence of learning on the spot)	20			3	1								24
7. Developing classroom assessment environment for students	6												6
a) Sharing and negotiating goals and expectations	17	1		13				1		1		1	34
b) Student self-assessment	17	5(3)		15					1			3	41(3)
c) Peer assessment	8			2									10
Subtotal–dimension & purpose	100(5)	65(39)	1(2)	186(9)	5	18(3)	1	33(2)	4	3	0	36(6)	452(66)
Totals by dimension	407				62				49				518

Note. Classification 4 is represented by the coding format X(Y), where X refers to the number of items relating to classroom assessment and Y refers to the number of items relating to externally-controlled assessment.

Appendix 2: Overview of scales

Reference	Study purpose	Dimension	Focus of instrument	Intended use	Respondents
Arter and Busick (2001). (extracted information about research from Arter, 2001–conference paper)	Evaluate the results of an assessment professional development programme	Confidence in knowledge and skills	(1) Clear achievement targets, (2) assessing student achievement, (3) student-involved classroom assessment, (4) communicating effectively and accurately about student achievement. Addressing specific standards of classroom assessment quality from the Assessment Training Institute (ATI).	Evaluate impact of a professional development programme and inform future professional development programs	• Principals • Teachers
Bandalos (2004)	Investigate the benefits and drawbacks of a teacher led assessment system. Investigate whether teachers had learned more about assessment as a result of the implementation of the system.	Knowledge and practices	Aspects of classroom assessment.	Contribute to an evaluation of the effects of Nebraska's Standards-Based Teacher-Led Assessment and Reporting System (STARS)	• Primary school teachers • Secondary school teachers
Brown (2004)	Investigate teachers' conceptions of assessment	Conceptions (beliefs)	Four conceptions of assessment: (1) Assessment improves teacher instruction and student learning (2) Assessment makes students accountable for their learning (3) Teachers or schools are made accountable through assessment (4) Assessment is irrelevant to the work of teachers and life of students.	Inform policy and professional development	• Primary school teachers • Managers
Dellinger et al. (2008)	Report on the development and validation of a teacher self-efficacy beliefs instrument	Self-efficacy (beliefs in capabilities; knowledge, skills and confidence)	A variety of aspects considered essential to effective teaching and learning in a classroom context.	Inform grouping of teachers, assessing needs of teachers and informing professional development	Primary school teachers
DeLuca and Klinger (2010)	Investigate the effects of a local teacher education programme in developing assessment literacy.	Confidence in knowledge	Practice, theory and philosophy related to assessment of and for learning.	Inform pre-service teacher education programmes and inform future research and practices	Pre-service teachers (primary and secondary)
Hambrick-Dixon (1999)	Investigate teachers' self-reported knowledge of assessment and assessment literacy before and after a course evaluation and assessment of multiethnic and racially diverse students.	Knowledge and skills	Principles of assessment in Assessment of and for learning. Approaches and methods for assessment. Rigour in Assessment. Based on various US standards.	Evaluation of a course	Teachers enrolled in an education course

Appendix 2: (*Continued*).

Reference	Study purpose	Dimension	Focus of instrument	Intended use	Respondents
James and Pedder (2006)	Report on the development and implementation results of a questionnaire to measure teachers' values of classroom assessment and contrast levels of assessment practice with their values	Practices and values (importance of practices)	Teacher classroom assessment: Purposes, goals and functions, classroom practices and learning processes, context, roles and responsibilities, and approaches to classroom assessment.	Inform professional development research project and future research	• Primary school teachers • High school teachers
Kershaw (1993)	Investigate teachers' use of student assessment information in making educational decisions	Perception of competence (knowledge and skills) (we did not consider the section of the questionnaire that focused on practices and attitudes)	A wide range of formal assessment design and evaluation activities, partly based on 'Standards for Teacher Competence in Educational Assessment of Students' (American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990).	Inform research and teacher training programmes and future research	Secondary vocational education teachers
Leighton et al. (2010)	Investigate teachers' beliefs about cognitive diagnostic information from classroom assessment versus large-scale tests	Beliefs	Focus on three categories of cognitive diagnosis: (1) providing information about student learning process (2) influence meaningful student learning and (3) eliciting learning or test taking skills.	Inform theory and professional development policies for teachers and pre-service teachers	Secondary school teachers
Lysaght and O'Leary (2013)	Report on the development and validation of an Assessment for Learning audit instrument	Knowledge and practices	Assessment for Learning with four scales: (1) sharing learning intentions and success criteria, (2) questioning and classroom discussions, (3) feedback, and (4) peer- and self-assessment.	Self-reflection tool for teachers and inform professional development	Primary school teachers

Study	Focus	Description	Purpose	Population
Mertler (2000)	Confidence	Examine the current assessment practices of teachers and specifically the use of methods to ensure the validity and reliability of their classroom assessments	Formal classroom assessments; assigning grades. (only self-report items were considered)	Inform teacher education programmes and professional development	• Primary school teachers • Secondary school teachers
O'Sullivan and Johnson (1993)	Competency in skills and knowledge	(1) Describe the development and pilot testing of an instrument to measure teachers' assessment competencies (2) evaluate the effectiveness of a teacher education course	(1) Formal classroom assessment instruments: development, analysis, validity and reliability. (2) Developing performance tasks. (3) Using standardised tests. (4) Ethical issues. In line with 'Standards for Teacher Competence in Educational Assessment of Students' (1990). Strong focus on psychometrics.	Inform teacher education course	Teacher education students (Masters level)
Panadero et al. (2014)	Conceptions (beliefs about purposes) and practices	Explore teachers' beliefs about student self-assessment in relation to their use of self-assessment	Using self-assessment in classrooms including advantages and disadvantages.	Inform theory and policy about the impact of a reform to include self-assessment in classroom practice	• Primary school teachers • Secondary school teachers • University/adult education teachers
Pat-El et al. (2013)	Practices	Report on the development and validation of an instrument to measure teacher and student perceptions to Assessment for Learning practices in classrooms and identify differences in their perceptions.	Assessment for learning, focusing on (1) Monitoring student progress and (2) Scaffolding to assist students' self-assessment.	Use by teachers (and students) to identify and reflect on practices	Secondary school teachers (and students, this scale was not considered).
Schulte et al. (2004)	Dispositions	Report on the development and validation of an instrument for measuring teacher disposition.	Numerous aspects of instruction and classroom management, some of which are relevant to formative assessment [used a subset of items] that align with the dispositions of effective teachers as specified under INTASC's (1991) Model Standards for Beginning Teacher Licensing and Development.	Teacher candidate using self-assessment of their dispositions over the course of their pre-service education to help them reflect on their future profession.	Pre-service teacher education students
Smith et al. (2014)	Beliefs	Examine the impact of a 3-year teacher education programme on beliefs about assessment	Assessment of and for learning: Principles, approaches and methods for assessment and rigour.	Inform pre-service teacher education programme	Pre-service primary teachers

Continued

Appendix 2: (*Continued*).

Reference	Study purpose	Dimension	Focus of instrument	Intended use	Respondents
Vanden Berk (2005)	Evaluate and report on the development of a training programme for teachers in assessment	Knowledge, attitudes and confidence	Aspects of the Student Evaluation Standards ([SES] JCSEE, 2003); propriety, utility, feasibility, and accuracy.	Evaluate the effectiveness of the SES, inform research on teacher training	High school teachers
Young and Jackman (2014)	Specifically investigates untrained and trained teachers' perceptions of, attitudes to and use of strategies in formative assessment	Perceptions, attitudes and practices	Formative assessment: (1) perceptions of formative assessment (2) attitude to practices (3) frequency of practices.	Inform policy on teacher education and professional development	Secondary school teachers
Zhang and Burry-Stock (2003)	Investigate teachers' assessment practices and self-perceived assessment skills, relative to their teaching experience and measurement training	Practices:• Use of • Skilled in	A broad range of classroom assessment activities such as test construction, interpreting test results, grading and using evidence from assessment in decision-making (partly based on Standards for Teacher Competence in Educational Assessment of Students (AFT, NCME, and NEA, 1990)).	Inform teacher education, professional development and future research.	• Primary school teachers • Middle school teachers • High school teachers