



Research paper

## Evidence of teacher assessment work and its relationship to their assessment identity

Lenore Adie<sup>a,\*</sup>, Claire Wyatt-Smith<sup>a</sup>, Mary Finch<sup>a</sup>, Christopher DeLuca<sup>b</sup>

<sup>a</sup> Institute for Learning Sciences & Teacher Education, Australian Catholic University, Brisbane, Queensland, Australia

<sup>b</sup> Queen's University, Kingston, Canada



### ARTICLE INFO

#### Keywords:

Assessment identity  
Teachers' work  
Teacher judgment  
Assessment literacy  
Extended performance

### ABSTRACT

This article presents evidence of teachers' assessment work to further understandings of the notion of teacher assessment identity. Data are drawn from transcripts of fourteen teacher meetings involving forty Australian middle-school teachers. Using discourse analysis, we examine teacher talk of their assessment practices to distil underlying influences on collective and personal decisions and actions. Results reveal three main influences on assessment identity: the policy context; teacher collaborative networks that build shared understandings and promote self-confidence in grading decisions; and inclusion of targeted resources. The findings can be used in the continuous development of assessment practices across pre- and in-service teaching.

### 1. Introduction

The construction of teacher assessment identity is a continuous process across pre- and in-service teaching. In a 2016 issue of this journal, Xu and Brown presented the teacher assessment literacy in practice (TALiP) framework, based on a scoping review of assessment literacy studies. This included "teachers' identity (re)construction as assessors" as "the ultimate goal of TALiP" (p. 158). At a similar time, Looney et al. (online version 2016; published 2018) published their reconceptualization of teachers' assessment practices rooted in a view of teacher assessment identity, which included interlinked dimensions of assessment knowledge, skills, confidence, beliefs, feelings, and role. Taking into consideration these two accounts of assessment identity, we understand teacher assessment identity to be an interweaving of systemic and local assessment contexts with disciplinary and assessment knowledges and conceptions (including beliefs and values). In this way, we understand that teachers working within one education system may hold some shared practices but will also have quite individual assessment identities. These different influences need to be considered in the support of in-service and pre-service teachers as they develop their assessment practices.

There have been almost 200 articles that have referred to one or both articles (i.e., Looney et al., 2018; Xu & Brown, 2016) with some recent empirical work beginning to explore assessment identity in diverse contexts (e.g. Estaji & Ghiasvand, 2022, 2023; Feola et al., 2023; Gan &

Lam, 2023; Leonardsen et al., 2022). In this article, we contribute to this exploration by drawing on teacher talk to empirically see school and classroom assessment practices and investigate the influences echoed in this talk. The assessment practices are situated within a policy context of data and evidence-informed decision-making, and national professional standards that include expected assessment knowledge and skills. While statements of professional standards are not uniform across all education systems, the findings of this study are relevant to an international audience in that all teachers work within frameworks (e.g., policy, principles, curriculum) that regulate content, pedagogy, and assessment practices to differing degrees (Bourke et al., 2018; Skourdoumbis, 2019; Talbot, 2016). How these regulating mechanisms are taken up and appear as practice is of interest in this paper.

Teacher talk of their school and classroom assessment practices are analyzed through discourse analysis methods proposed by Gee (2014), in particular his theoretical tool of *Intertextuality*. These methods enabled enquiry into how teacher talk of their local assessment practices merged with systemic policy and was recognized or accepted by other teachers as shared understandings of how assessment is conducted. The findings are discussed in terms of the situated nature of assessment practice and how targeted resources can support in-service and pre-service teachers to continually grow in their knowledge and skills as assessors. The article addresses three key questions:

What does teachers' talk reveal about their assessment identities?

\* Corresponding author.

E-mail address: [Lenore.adie@acu.edu.au](mailto:Lenore.adie@acu.edu.au) (L. Adie).

How does systemic and local school policy and practices shape teachers' enacted assessment identities?

How can targeted resources stimulate thinking about assessment?

This investigation is located within a larger Australian Research Council (ARC) Linkage project (LP180100046)<sup>1</sup> that examined the use of scaled exemplars with cognitive commentaries<sup>1</sup> of judgment decisions to support consistency of teacher judgments. The project was conducted from 2019 to 2022 across two Australian states. It involved over 200 teachers in Years 4, 6, and 8 teaching in the disciplines of English, mathematics, science, and religious education. Teachers submitted assessment work samples, across the range of achievement levels or grades (A-E), for inclusion in pairwise comparison and standard setting processes (Humphry et al., 2023). Selected scaled exemplars were chosen to represent A to E grades. Teachers wrote cognitive commentaries of these exemplars describing how the strengths and weaknesses of a performance came together in their overall judgment decision, as well as identifying next teaching steps to progress student learning (Wyatt-Smith et al., 2024). Teachers then met online in year level and discipline groups to finalize the A to E cognitive commentaries for the selected exemplars. The data for this paper are drawn from the online meetings that occurred in the state of Queensland as teachers finalized the cognitive commentaries and discussed their assessment judgments and assessment practices more generally.

The following section provides an overview of the theorizing of teacher assessment identity. Next, the research and assessment policy contexts and the research methods are presented, followed by the findings and discussion, which focus on three main influences on assessment identity. We provide evidence of how assessment identity is apparent in teacher talk.

## 2. Assessment identity and practice

Identity is a means of explaining and making sense of oneself in relation to the world at large; identity is socially constructed and always shifting and dynamic, rather than fixed (Van Landveld, Schoolneboom, Volman, Croiset, & Beishuizen, 2017). Professional identity is composed of multiple elements which continuously emerge within a given context (Garner & Kaplan, 2019). These elements include one's ontological and epistemological beliefs, self-perceptions, purposes, and goals. Teacher professional identity may be considered "the personal version of the teacher role that a person construes within his or her lived context" (Garner & Kaplan, 2019, p. 9).

Teacher assessment identity is typically considered an aspect of teacher professional identity (Estaji & Ghiasvand, 2022). In education systems where teachers are responsible for assessment, their assessment competence (knowledge and skills) and their beliefs about their role as assessors are implicit in every aspect of their work (Looney et al., 2018; Xu & Brown, 2016). In this context, assessment is interlinked with teachers' decisions about planning, teaching, and their professional learning. How a teacher views their role as an assessor may affect the assessment knowledges and skills they seek to develop and their professional discussions with colleagues. Assessment identity is, therefore, more than assessment knowledge, skills, and beliefs; it is the interaction of these aspects as well as a teacher's own history with assessment that forms their assessment identity.

Frameworks of assessment identity acknowledge that assessment operates within systemic boundaries and that teachers can exercise

agency within these structures to differing degrees. Xu and Brown's (2016) TALiP framework conceptualizes teacher assessment identity as the embodiment of teacher as assessor. In their view, assessment identity is constructed and reconstructed through reflection on assessment processes. Such reflective practice leads to self-awareness as an assessor who can insightfully accommodate and translate assessment policies and principles to classroom practices. Foundational to assessment identity formation, in the TALiP framework, is the development of assessment literacy as a set of inter-related competencies based on mastery of assessment knowledge, including content knowledge (CK) and pedagogical content knowledge (PCK); knowledge of assessment purposes, content, and methods; grading; feedback; peer and self-assessment; assessment interpretation and communication; and assessment ethics. Above the knowledge base sit conceptions of assessment, cognitive dimensions, views of learning, and epistemological beliefs and feelings about assessment. As teachers assess, these act as interpretative frameworks between the demands of the macro-sociocultural and micro-institutional contexts and teachers' knowledge.

Looney et al. (2018) described teacher assessment identity as "who teachers are in the process of assessment" (p. 456). Their model describes teachers' assessment identities as dynamic and interactive, shaped by their knowledge, skills, dispositions, beliefs about assessment, self-efficacy, and sense of agency in their roles. Furthermore, the socio-historical contexts of teachers and the political contexts in which they work shape their assessment identity. Similar to Xu and Brown (2016), Looney et al. acknowledge the affective and ethical dimensions of teachers' assessment practices within situated contexts. More recently, Pastore and Andrade (2019), in presenting an expanded model of assessment literacy containing conceptual, praxeological, and socio-emotional dimensions, have also highlighted the importance of context in understanding how teachers develop capacity to be assessors of student work.

How teachers' knowledge and skills interact with their values, attitudes, and beliefs about assessment in practice has been identified as a gap in research (DeLuca et al., 2016; Kippers et al., 2018). Recent empirical work across different contexts has started to take up this challenge. Studies have identified that assessment literacy and identity can be developed through reflection on assessment outcomes and practices (see DeLuca et al., 2023; Leonardsen et al., 2022). In particular, reflective practice with peers has been demonstrated as supportive for developing assessment literacy and identity for (1) Israeli mathematics pre-service secondary teachers (Ayalon & Wilkie, 2020); (2) Irish student teachers' online learning (Doyle et al., 2021); (3) United States university instructors (Feola et al., 2023); (4) Chinese English as a Foreign Language [EFL] teachers (Gan & Lam, 2023); and (5) Canadian on-line practica (Ge, 2024). In addition, the use of e-portfolios to promote reflection on assessment practices and beliefs was found to be supportive for novice and experienced EFL teachers in Iran (Estaji & Ghiasvand, 2022, 2023). However, the stresses of workload, particularly time for administrative tasks, continue to limit opportunities for meaningful discussions of student work (Bedford & Barnes, 2022).

Another concern raised in the literature is how teachers "are prepared to source and use evidence for improving learning and teaching" (Wyatt-Smith et al., 2017, p. 250); that is, how the range of data types, the complexities of judgment making and the translation of evidence to next-step teaching and learning and reporting is taught in real and meaningful ways. In this paper, by investigating how teachers' assessment identities are revealed in their talk, we aim to contribute to understandings of how teachers develop an identity as an assessor to support pre- and in-service learning.

## 3. Research context

In Australia, the assessment knowledge and skills expected of teachers are broadly described in systemic assessment policies and

<sup>1</sup> A cognitive commentary is a teacher's explanation of their judgment of the quality of a student's performance. The commentary describes what the teacher considered to be the strengths and weaknesses of the performance, and how these were combined to give an overall judgment. Commentaries can also describe next steps for teaching based on this analysis of the performance (see Wyatt-Smith & Adie, 2021; Wyatt-Smith & Bridges, 2008).

articulated in professional standards. Australian teachers of students in Years 1–10 are required by legislation to report to parents using A to E grades (or equivalent 5-point scale) (Australian Government, 2022). To support national consistency in reporting, the Australian Curriculum sets out the expectations for what students should be taught with an overall achievement standard for each year level and discipline (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2022b). State and territory governments are responsible for school education within the framework of national legislation and curriculum.

Standard 5 of The Australian Professional Standards for Teachers (APST) lays out the expected assessment knowledge and skills across five focus areas: (1) Assess student learning; (2) Provide feedback to students on their learning; (3) Make consistent and comparable judgments; (4) Interpret student data; and (5) Report on student achievement (Australian Institute for Teaching and School Leadership [AITSL], 2017). Of specific interest to this paper are focus area 5.3, which refers to teachers' involvement in assessment activities such as participation in moderation meetings to support consistent and comparable judgments, and focus area 5.4, which identifies the need for teachers to interpret and use classroom and standardized test data to inform their teaching.

This professional standard is taken up in the Queensland Department of Education policy, *P-12 Curriculum, assessment, and reporting framework* (P-12 CARF), which requires schools to:

- use assessment and reporting data to determine the focus of moderation processes
- use a whole school approach to moderation processes to align curriculum, pedagogy, assessment and reporting; and to ensure consistent judgments and accurate reporting against the achievement standards (Queensland Department of Education, 2022, p. 5, p. 5)

Moderation processes are described as requiring “professional conversations and a series of calibration and confirmation activities to reach consensus about the evidence in student work against the relevant standard” (Queensland Department of Education, 2022, p. 23). Advice to Queensland teachers describes moderation occurring across four junctures or phases within the teaching/learning cycle. These processes are labelled ‘*Before After End Moderation*’:

Moderation processes at multiple junctures ideally occur:

- before assessment takes place, at the planning stage, and at appropriate times throughout teaching and learning
- after assessment takes place, but before it is graded
- after assessment is graded; and
- at the end of the reporting period using student assessment folios. (Queensland Department of Education, n.d, n.p.)

The CARF guidelines further direct schools to use summative assessment “to gather evidence and to report on student learning and academic achievement against the relevant standard” (p. 5) and formative assessment to “monitor student progress against the achievement standards using formal and/or informal monitoring tasks, to inform ongoing teaching and learning” (Queensland Department of Education, 2022, p. 6).

As previously stated, the ARC Linkage project aimed to investigate whether the use of scaled exemplars with cognitive commentaries, that detail how qualities of performance were combined to reach a final grade, can improve teacher judgment consistency. The project consisted of five stages: (1) collection of classroom assessment tasks focusing on extended performances that include complex thinking; (2) pairwise comparison of tasks; (3) standard-setting; (4) development of cognitive commentaries; and (5) trial of scaled exemplars with cognitive commentaries (see Humphry et al., 2023).

Using scaled and/or common assessment tasks and consensus moderation for the purpose of comparability of grading has been utilized in Australian states for many decades. For example, Hill et al. (1997)

discussed the use of reference tests to check for “the reasonableness of schools’ assessments” (p. 32) related to final year certification in the state of Victoria. Leigh-Lancaster and Rowe (1999) illustrated how extended common assessment tasks as used in Victoria can be part of a valid and reliable assessment system. Furthermore, in Victoria and in Queensland, consensus moderation has been extensively used to maximize the comparability of assessments (Hill et al., 1997; Stanley et al., 2009; Wyatt-Smith & Colbert, 2014). For example, the system of consensus moderation in Queensland has been in operation for over 50 years, with claims that participation in moderation discussions supports teacher learning about assessment (Allen, 2012; Maxwell & Cumming, 2011), as well as recommendations for the need to strengthen the moderation system and ongoing research and development to ensure the reliability of such a system (Matters & Masters, 2014; Wyatt-Smith & Colbert, 2014). The current project differs from this group of work in that (1) the assessment tasks were not common, rather they were tasks selected or designed by individual schools or teachers; (2) the assessment tasks were related to middle schooling rather than a high-stakes leaving certificate; and (3) the purpose of the scaled exemplars was for teacher use as a form of self-moderation to support judgment consistency, rather than for the purpose of systemic verification of scoring comparability.

This paper focuses on Stage 4 of the ARC Linkage study in which teachers wrote and submitted cognitive commentaries to the research team for two to three exemplars for their grade and discipline (e.g., Year 4 English A grade). Commentaries written for the same exemplar were merged into one document by the research team:

1. Each commentary was analyzed for points of confluence and disagreement.
2. Points of confluence were merged into one composite statement.
3. Points of difference were left in the commentary and highlighted.

The merged commentaries were the focus of online meetings (120–150 min) in which teachers negotiated the wording of the set of A-E commentaries for their discipline and year level to progress to the final stage. There was a maximum of five teachers per meeting to allow for discussion and views from all participants to be heard. The discussions focused on how the expected standard was represented in the exemplar, as well as next-step teaching strategies to progress student learning. In these discussions, teachers also shared their classroom assessment practices and how consistency of judgment was ensured within their schools (see Adie et al., 2023).

#### 4. Methods

This paper draws on transcripts of 14 Stage 4 teacher meetings that were conducted with 40 Queensland teachers (Table 1). Ethics approval was received from the Human Research Ethics committee of the Australian Catholic University (Ethics approval number: 2019–11H) and partner organisations. All participants provided signed consent for the meeting transcripts and their feedback to be used in publications. Video

**Table 1**  
Number of teacher participants and online meetings (year level and discipline).

Discipline	Year level	Teachers	No. of meetings
English	4	3	1
	6	1	1
	8	4	1
Maths	4	6	2
	6	9	3
	8	2	1
Science	4	4	1
	6	10	3
	8	1	1
<b>Total</b>		<b>40</b>	<b>14</b>

recordings of the online meetings were deleted after transcription as per ethical requirements.

#### 4.1. Analysis

Analysis of meeting transcripts focused on what teacher discussions revealed about their assessment work and their assessment identity. Examination of the data corpus occurred in two stages. First, we identified major themes using an analytic method described by Braun and Clarke (2006). This is a flexible interpretative method whereby patterns of meaning are distilled and explored iteratively across the data set. After surveying the corpus, initial codes were generated that reflected key aspects of teacher assessment identity as conceptualized by Xu and Brown (2016) and Looney et al. (2018). These two models were used to direct attention rather than as pre-determined codes. A further analysis of each script was undertaken by writing reflections (one per script) highlighting how the data instantiated aspects of teacher assessment identity.

Two researchers then independently constructed tables of our coded references to identify patterns of sub-themes, noting key ideas. These tables were compared to reach a shared interpretation of the data among the research team. Three overarching themes were identified: (1) influence of policy context in participant talk about their assessment practices; (2) importance of developing shared understandings of expected standards of performance through collaboration with colleagues, and impact on self-confidence; and (3) function of targeted resources to stimulate linking of teaching, learning, and assessing as continuous practice.

Second, we used a discourse analysis method (Gee, 2014) to examine how participants' language usage revealed latent or implicit social perspectives and features of identity. Gee's discourse analysis is based on a view of language as both shaping and shaped by its social context; meaning can be understood only with reference to context. Gee describes his method of discourse analysis as a research tool to better understand the myriad ways people use language to create identities and to perform actions in the world. This analytic method met our goal of examining how teachers understood their roles as assessors within the education and school policy contexts in which they worked.

Gee's (2014) theoretical tool of *Intertextuality* offered a key lens to inquire into links between participant descriptions of their assessment practices and systemic texts such as the Australian Curriculum and state assessment policy. Direct or indirect quotation of texts in conversations implies a link of some kind. Gee provided several analytic tools that guide different questions to ask of a text, for example, why participants had worded utterances in particular ways and not others; why certain topics are chosen or omitted; what assumptions underlie language use; and how language is used to enact recognizable social identities. Examples of Gee's analytic tools that were of particular use in our investigation are included in Table 2 with illustrative examples of participant talk and our analysis of this talk.

#### 4.2. Limitations

Results of the project should be interpreted noting the following limitations. Data for this article were gathered from a single large Australian state where the employing authority had a longstanding declared policy commitment to teacher judgment using achievement standards and social moderation. A relatively small number of judges participated in the research. With a small sample, results depend more on the specific schools and contexts of participants than would be the case with a larger sample of teachers. Results are indicative but cannot necessarily be extrapolated to all teachers.

### 5. Findings

Our analysis identified three major influences in the formation of

**Table 2**  
Examples of Gee's (2014) discourse analysis tools in use.

Tool	Example in participant talk	Analysis
<i>Big "D" Discourse</i>	When I started teaching about 1000 years ago, marking was picking up your red pen and putting ticks and crosses even on English.	Socially identifiable identity: historical notion of assessor as judge of student work. Socially identifiable activity: marking as a series of red pen ticks and crosses.
<i>Figured World</i>	"Can you just mark this one here and just give me two or three of yours?" I feel like that's just a natural thing you just do.	Type of world: grading of student work or judgment making is a collaborative activity that results in fairness for the student.
<i>Why This Way and Not That Way</i>	... that sentence in the C descriptor it says, "Predictions and explanations of inferences and results are based on scientific understanding and communicated in a formal way". It doesn't ... you can see it doesn't say that you have to justify.	Rewording the final sentence to begin "you can see" adds emphasis to the unstated assumption that participants agree that the descriptors are authoritative in defining grades.
<i>Subject</i>	... coming back to having moderation at the start of a unit's writing and teaching is that we know what we're aiming for, and then we can check in on that student's progress towards those goals as we go along. So, it's certainly got value.	The benefits of the <i>before</i> stage of moderation (that is, during planning) are promoted. A significant omission: participants never discussed explicitly or implicitly any resistance to moderation processes in schools.
<i>Making Strange</i>	... the simple statement like interpret a map, does that mean read a map? Does that mean they [students] can make a map? What does that mean? And if we all have that similar understanding or have a shared understanding of the word 'interpret', that makes it a lot easier to then go, well here's my success criteria for this.	Why is the participant concerned with the meaning of "interpret"? The participant's concern is linked to developing shared understanding of the term. The assumption is that shared understandings of terms will lead to easier (connected) teaching practices.
<i>Identities Building</i>	I think she did 3000 trials or something ridiculous. She's then come up with 800 odd successes from that. She's got the relative frequency of 27.9%. She's said that is the theoretical probability, instead of referring to it as the relative frequency.	Using language related to mathematics (e.g., "trials", "relative frequency", "theoretical probability") to enact a socially identifiable role as mathematics teacher, and so the authority to assess and comment on student work.

teacher assessment identity: (1) the policy context; (2) teacher collaborative networks that build shared understandings and promote self-confidence; and (3) the inclusion of targeted resources.

Central to participants' assessment practices were national and state policy and curriculum documents, including the Australian Curriculum and related artefacts provided by the Australian Curriculum Assessment and Reporting Authority (ACARA), the Queensland Curriculum and Assessment Authority (QCAA), and Queensland Department of Education, which oversees education in Queensland and is the dominant employer of teachers. Local implementation of employer expectations – the enactment of the overarching syllabus documents – was evident in participants' descriptions of their school assessment and moderation processes.

Our analysis also revealed that processes built around collegial decision-making and developing shared understandings of expected achievement standards related to beliefs of fair assessment processes, improved teaching and learning, and increased confidence in oneself as an assessor. Furthermore, targeted artefacts that directed teachers' gaze to connect their analysis of student work to specific next-step teaching strategies stimulated the perception of teaching, learning, and

assessment as a continuous process in contrast to teaching isolated units of work.

### 5.1. The policy context: intertextuality of systemic text and teacher talk

National and state syllabus and curriculum documents provided an overarching context within which participants' assessment practices were enacted. The words and phrases used by teachers aligned closely with official documents such as the Australian Curriculum (ACARA, 2022a), and resources provided by the QCAA and the Queensland Department of Education (e.g., Curriculum into the Classroom [C2C],<sup>2</sup> the Guide to Making Judgments [GTMJ]<sup>3</sup>). The intertextuality (Gee, 2014) between the teachers' talk of their local assessment practices and national and state assessment policy and related documents suggests that their assessment practices and beliefs were directly or indirectly shaped by these documents.

The authority of the Australian Curriculum was stressed by participants, for example, "if it's [the assessment task] not matching the Australian Curriculum, then obviously that needs to be addressed". The participant's inclusion of "obviously" emphasizes that matching the task requirements and the Australian Curriculum is a taken-for-granted assessment practice in Queensland schools which did not require further justification. This assumption was also visible when participants described how the Curriculum was used to settle grading differences.

My Assistant Principal always just goes back to the Curriculum. That's literally, "Okay, we've got a disagreement. Let's go to ACARA. Let's look at the achievement standard and let's have a look at a satisfactory portfolio, an above satisfactory" and it goes from there. So, it's not opinions of teachers ...

The participant contrasts the act of "always" consulting the Curriculum ("ACARA", "achievement standard", "satisfactory portfolio") with the intuitive evaluations or "opinions of teachers". The Curriculum was the authoritative source of the expected standard rather than what teachers themselves may think. Understanding context is important here – the teachers are required to teach the Australian Curriculum and state assessment policy emphasized the need for "consistent judgments and accurate reporting against the achievement standards" (Queensland Department of Education, 2022, p. 5). As illustrated in the following findings, teachers stated a belief in the fairness and impartiality of assessment judgments by being able to track decisions back to the curriculum documents and related standards statements.

The authority of the achievement standard was illustrated when participants used the wording of a descriptor to defend a decision:

When you go back to that sentence in the C descriptor it says, "Predictions and explanations of inferences and results are based on scientific understanding and communicated in a formal way". It doesn't ... you can see it doesn't say that you have to justify.

Having quoted the standard, the participant identified that justification was not a requirement. The participant amends the final sentence to add "you can see" before "it [the descriptor] doesn't" which implies an expectation that others will view this information in the same way.

The close textual alignment between participant descriptions of their personal and school assessment processes and the state policy and supporting documents was mirrored in every teacher's discussion in the

<sup>2</sup> C2C are units of work with accompanying assessment tasks, marking guides, and resources developed by the Queensland Department of Education to support teachers to deliver the Australian Curriculum. These materials are optional to use or adapt to school contexts (Queensland Department of Education, 2020).

<sup>3</sup> The GTMJ refers to the criteria sheet or rubric. This may be in a traditional grid format or as a continuum (for an explanation of these types of rubrics see Matters, 2006; Wyatt-Smith & Klenowski, 2013).

meetings. For example, one participant contrasted historic marking practices associated with accuracy and correcting errors – "ticks and crosses" with "your red pen" – with current practices of "finding evidence" of achievement "to show me you understand or can do something in particular". Another participant described their local assessment practices for summative and formative purposes:

We annotate to collect evidence and assign a grade ... that's not the end of the student's learning. That's just a point in time thing ... [which] says "These are the strengths. These were the things we need to work on. Here's the implications for me as a teacher going forward to the next piece of work or all the things that are coming" because even though I might not assess exactly the same thing again, the student still needs to develop these skills, and that needs to inform my teaching.

The participant described their summative assessment process of "collect[ing] evidence and assign[ing] a grade" and how this was "not the end" point. Rather, "going forward" used this evidence to 'inform my teaching'. These examples illustrate a strong intertextual relationship between teachers' descriptions of their assessment practices and state assessment policy. This relationship may be forged in initial teacher education as preparatory teachers learn the language of assessment. It is then further strengthened in school meetings, moderation discussions and ongoing professional learning events. The interweaving of these policy and curriculum texts are further illustrated in the following sections as teachers describe their work with colleagues and their developing confidence as an assessor.

### 5.2. Teacher collaborative networks to build shared understandings and promote self-confidence

Reflecting the Queensland CARF guidelines, participants described moderation as "not just an end point where we ... have student work that we're looking to grade from A to E – it's about coming back, ... developing that shared understanding of the Australian Curriculum first". In this statement, moderation is a collegial process that connects curriculum and assessment of student work to develop "shared understandings".

Participants described their involvement in a range of formal moderation processes in year level, whole-school, and district contexts: "Education Queensland ... their directive is that we're using the four phases of moderation"; "our focus at the moment is doing what the Ed. Queensland wants in regard to the four phases of moderation". In these statements, teachers use the collective pronouns 'we' and 'our' to describe their responses to Departmental advice. In addition, "doing what the Ed. Queensland wants" and naming this as a "directive" illustrates how systemic advice is seen as authoritative. It is notable that resistance to moderation was not presented as a topic by any of the teachers; rather their discussions showed a focus on improving their moderation processes to improve their assessment skills and judgment making. Assessment practice and learning about assessment were portrayed as a collaborative process, informed by official documents, but in which teachers took fully active roles in developing shared understandings of standards.

Teachers spoke of embedding each phase of moderation into their teaching practices. Their involvement in the first phase of moderation (*Before*) was described as occurring during collaborative planning meetings:

... when you fully unpack the *Guide to Making Judgments* [rubric], making sure there's alignment between the task and the *Guide to Making Judgments*, making sure, seeing if there's any changes that we need to make, and making sure that we've got alignment and consistent understandings across our cluster of schools as well.

Using the rubric, teachers ("we") decide collectively how to align their task expectations either within their school or within the group of

schools (“our cluster”). Repeating “making sure” three times suggests the strength of the belief in the goal of comparability (“alignment”, “consistent understandings”) alongside the use of “we” and “our” to reinforce the imperative of shared understanding and agreement of assessment standards.

Developing alignment and consistent understandings of the required standard of work was described by another participant as reducing “the potential for conflict” about a future grading decision:

If you invest the time in the *Before*, [then] *After* [is], dead easy, whereas *After* used to have ... the potential for conflict or robust conversation. Whereas putting in that *Before* phase, you iron out all of those issues then and you do have a lot more clarity around what we’re all looking for and what we’re teaching. So, we all teach better too.

In the talk segment above we hear the teacher declaring that spending time during planning made assessment “dead easy”. The teacher identified “clarity” about standards (“what we’re looking for”) and how these will be taught as important considerations within the teaching team. In this statement, the teacher shifts the “robust conversation” from the moderation stage to the planning stage so that these shared understandings of the required standard may be used to inform their teaching and subsequent assessment work. While the moderation phase is described as “dead easy”, the teacher couches their statement in phrases such as “potential for conflict” and “a lot more clarity” suggesting that there is no guarantee of absolute alignment between the phases. The teachers’ approach is one of using assessment standards for the benefit of teaching, learning, and assessing rather than just assessing. This belief – that shared understandings of criteria and standards led to better teaching – was expressed by participants across all meetings and never questioned or raised as a point of discussion.

The importance of connecting planning and teaching through collaborative discussion to achieve consensus on the meaning of standard descriptors is further illustrated in another teacher’s description:

What do we understand by that statement and what do we know about that cognitive verb? So, what does ‘interpret’ mean to you, you, and you, and you? And if we all see that differently, it makes it really difficult. So that’s why that *Before* phase is really important. So, when we’re having the simple statement like interpret a map, does that mean read a map? Does that mean they [students] can make a map? What does that mean? And if we all have that similar understanding or have a shared understanding of the word ‘interpret’, that makes it a lot easier to then go, well here’s my success criteria for this. I need to teach this, this, this, and this, because I know what that word means.

The participant described a process of seeking consensus around cognitive verbs such as “interpret” stressing that it has multiple possible meanings for different teachers (“you, and you, and you”). Different interpretations could create difficulties for teaching and assessing. Having canvassed some possible meanings (“interpret”, “read”, “make”) the participant explained that a “shared understanding of the word” makes it possible to define “success criteria” and, therefore, what to teach. There is an assumption that “a shared understanding” of these terms among colleagues is a central assessment goal and part of a teacher’s role.

The second moderation phase focuses on calibrating markers before grading takes place. A participant related a process used in a large secondary school:

There will often be sort of a meeting where the teachers gather to talk about a sample ... What the Head of Department or the coordinator ... will usually do is they’ll pick a paper out of the cohort and they’ll de-identify it and ... copy that paper ... and get every teacher to mark it ... then checking that we are all marking in the same way and discussing any variations ... and coming to a consensus and then

making sure that consensus is written down and then communicated to all the staff.

The authority of the Head of Department, the staff consensus reached, and the written communication of the consensus create an expectation of achieving comparable grading decisions. Teachers proudly shared the rigor of their multistep school processes to illustrate how “we are all marking in the same way”.

In addition, teachers identified their individual processes to ensure comparable grading decisions. Asking a colleague for a second opinion was presented as typical and assumed shared professional practice: “Just like every other teacher at every other school, we’ll all go up to each other and go, ‘I don’t know what mark to give this one; can you help me decide?’” Another participant characterized this co-operation as “natural”:

Just the lady I always work with ... I go, “Oh, I’ve given this student this. Can you just have a little look? What would you give them?”, or sometimes I don’t even mark it down and go, “I’m really confused with this one” or “I feel like I really want to reward this kid. Can you just mark this one here and just give me two or three of yours?”. I feel like that’s just a natural thing you just do.

Describing the practice of seeking a second opinion as “just like every other teacher” and “natural” identifies an embedded belief in grading as a process of collaboration. The quotes distinguished four possible reasons why these teachers place importance on a second opinion rather than trust their own decision: (1) support with decision-making (“can you help me decide?”), (2) verification of a grading decision (“I’ve given this ... What would you give them?”), (3) professional learning about scoring student work (to reduce confusion), and (4) a recognition of the possible impact of latent criteria on their grading (not to unduly “reward this kid”). There is a central belief here about grading as a collaborative and fair process.

The third moderation phase involves teachers bringing a selection of marked assessments to confirm grades. One secondary teacher described these meetings: “We all bring our samples with us, and we all pass them around to each other ... we just open straight up to the rubric, and we all argue why this one’s that one and why that one’s that one”. “The rubric” is presented as the linchpin to a rigorous process of negotiation (“we all argue”) and was apparent in each teacher’s description of their moderation meetings. These meetings were also described as a:

very respectful environment ... we take all the emotion out of it and we’re: “Hang on, I’m wondering if this has actually hit that standard”, or “I think this might be above that standard because” and we’re doing the justification ... we come back to the achievement standard.

The shared decision was reached by teachers putting forward possibilities (“hang on, I’m wondering if this has actually hit that standard”), with the phrase “I’m wondering if” softening the suggestion of disagreement. Teachers referred to the achievement standard as the reference point for their judgment (“hit that standard”), using evidence in the student work in their “justification” of a grade (“above that standard because ...”). As another participant stated, “We’re not putting down, we’re actually building and supporting one another ... Everyone has a voice equally, everyone’s involved equally, and no one has really a bad answer; we’re supporting one another through that discussion”. This participant suggested that moderation procedures rested on collegiality (“building and supporting one another”) in which views (“voice”) are exchanged (“everyone’s involved equally”) in the service of consensus and comparability of grading.

Receiving this collegial endorsement of grading decisions was particularly important for those teachers in regional areas, often several hours away from the next nearest school:

As teacher in charge of Year 8 for English, I’m just finding this [project] really handy because ... it means that the judgments that

the existing English teachers and I have been making, I feel like we are on track ... because it is really hard for us to identify that.

The chance to identify comparability with other schools across the state was valued by this teacher and implies a context where comparability of grading is sought. For teachers in regional and remote areas, accessing how other colleagues are judging student work was implicit in building confidence in themselves as assessors.

Similarly, another participant “from a school that sometimes doesn’t always have the best reputation” found that “our As and Bs are quite comparable” and concluded that “we must be doing something right”. For this teacher, finding that their top performances (“our As and Bs”) were comparable to those in other schools validated the awarded grades and the pedagogy in the school. In these words, you can almost hear the relief and see the strengthened confidence going forward in their teaching and assessment practices. You can also see how identity as part of a school merges with their identity as an assessor.

The *End* phase assesses overall achievement and is the final stage before reporting: “You’re actually mapping out every achievement standard for a student folio and then coming up with an on-balance judgment based on where the student’s at within each element, each achievement standard”. An “on-balance judgment” is described in Queensland assessment policy as “based on evidence of student performance in the assessment folio”, an “overall level of achievement against the aspects of the achievement standard assessed” and for “each assessable element” (Queensland Department of Education, 2022, pp. 11, 23). The teacher’s description of the *End* phase echoes the state assessment policy and guidelines.

Comparability was also described as occurring among groups of schools in the same region or geographic cluster: “we take our samples and work with other cluster schools around us to compare as well”. Cluster moderation was discussed during the planning stage and after assessment:

We have a validation process, but it’s not just with our school, it’s with the cluster. And sometimes ... we moderate ... and sometimes they’ll bring a different assessment that the school has created or like a C2C, and we go over the rubrics that we’ve got and make sure it connects with the questions and make sure that the students are able to get an A and is there things that we need to change before we give it to the students.

The “validation process” involves neighboring schools checking alignment of “the rubrics” and the assessment task to ensure there are opportunities for students to demonstrate their mastery of the curriculum, including at the A level. Cluster moderation focused on collaboration with colleagues to establish comparable standards of assessment across schools and fine-tune assessment planning.

Overall, moderation processes were described as focused on cycles of collegial practices seeking consensus in interpreting the Australian Curriculum and achievement standards and applying this to teaching and assessing. In doing so, teachers are performing the “consistent and comparable judgments” required by the Professional Standards (AITSL, 2017). Receiving confirmation of grading decisions through these processes increased teachers’ self-confidence as assessors.

### 5.3. The inclusion of targeted resources for teaching, learning, and assessment

Writing a cognitive commentary of a judgment involves deep thinking about the qualities evident in an assessment response. Teachers told us that this process helped them to develop their knowledge and understanding of student work, and the application of criteria and standards. The commentaries were used in two broad ways by the teachers. First, the scaled exemplars with associated commentaries developed as part of the project were used to sharpen understanding of a standard and inform A to E judgments. Teachers also used these to gain

ideas for differentiated next-step teaching strategies. Second, teachers used the cognitive commentary template to develop their own commentaries based on their student work samples. While this was not a project activity, the teachers could see the advantages of using the template in their practice. These two uses can be seen in the following illustrative examples of teacher talk.

The cognitive commentary connected assessment with teaching and learning. This connection was described by one teacher as: “we were talking about some of the points [strengths and limitations] that people had made ... that leads nicely into what, as a teacher, we could do, because there are activities [next-step teaching strategies] explained about these points”. Teachers found the range of suggested next-step teaching strategies provided useful options:

I found it very valuable ... hearing other people’s ideas on how to go back and improve it, particularly even with the high-level kids. Someone suggested modelling, teaching the same story but from different perspectives and perhaps making that a group activity where you give groups of kids, “Okay, you’re going to write it from this perspective. You’re going to try writing in third person. You guys are going to try this” ... I think that’s really valuable to hear other people’s ideas.

The link from areas for improvement to teaching strategies was continually noted by teachers as “valuable” for them. Having suggestions for supporting students working at high levels of achievement was specifically noted. In the example above, the teacher emphasizes their point by listing the strategies that have been suggested, including for differentiated teaching. Another teacher contrasted the cognitive commentaries with the systemic ACARA samples highlighting that the latter are “more annotated and don’t actually have the next step of teaching in areas of improvement”.

Being new to the year level was another reason some teachers provided for finding the commentaries useful:

I probably wouldn’t know, because I haven’t been on the grade level, where to go next with them [students]. But I found it really eye-opening that teachers could still find areas for improvement in this [A] sample.

Participants identified that extending high-achieving students can be difficult, especially for those new to a year level. As this teacher states, it was “eye-opening” to see how others found areas for improvement in the A sample.

Teachers also considered that the exemplars and cognitive commentaries could be shared with students.

I think having these samples and having it actually highlighted what they’re doing well, and I like it because it is student work. You can actually show students, this is what they’ve done, here’s the things that they’ve included, this is what you need to do.

In this example, the teacher lists the appealing features of the scaled exemplars and commentaries to support students in analyzing and improving their work: “it is student work”, it highlights the strengths (“what they’re doing well”), and the work sample can be shared with students as illustrations of “what they’ve [other students] done”. Another teacher considered that they could create their own commentaries and use their analysis of strengths and weaknesses with students to track progress: “This is what your observations were in your assessment from last term. Let’s put that up on the board and let’s look at how can you interpret this table differently or how could we go about improving this”.

Teachers also saw value in using the cognitive commentary template to conduct their own analyses to inform their planning and teaching.

[I’m] starting to see an evaluation of student work as informing our next steps in teaching. I think the cognitive commentaries make that particularly clear for teachers and while I wouldn’t do it on every

sample, to actually sit with a set of samples like that and say, “Alright, our next unit of work that we’re about to start, what does it mean for us? Is there a different starting point from what we thought the starting point was, either better or worse?”. I think it does help to focus teacher planning.

The template provided an avenue to enhance planning processes in this school. The teacher articulated an understanding of how analysis of student work is necessary to establish the “starting point” for teaching, which may be different from what was expected or written into a curriculum document. Through their analysis of “a set of samples” using the template, the teaching team could “focus teacher planning” and directly respond to their students’ learning needs.

Similarly, another teacher proposed that the cognitive commentaries enabled them to identify trends in students’ areas for improvement and how this could support their individualized or differentiated group teaching.

What the cognitive commentaries allow us to do ... is say, “Well, hang on, if there was something consistent to all the students at a particular level, that could be about my teaching and something I didn’t do in my teaching”, but if the reasons for the students being at the various levels are different, if I can’t find a pattern or a trend, then that might indicate that it’s about the individual next steps for those students. So, the cognitive commentaries could provide two types of feedback if we then looked at it through that lens.

The value of analyzing student responses and the explicit connection to pedagogic decisions is evident in this teacher’s talk. The trends or lack of trends could provide information about unsuccessful teaching strategies that need to be adjusted or identify a gap in a student’s learning for individual attention. Other teachers also linked points for improvement with next steps for teaching, moving beyond teaching the next unit of work to attending to learning needs.

To have that cognitive commentary that says, these are the strengths, these are the things we need to work on, here’s the implications for me as a teacher going forward to the next piece of work ... because even though I might not assess exactly the same thing again, the student still needs to develop these skills, and that needs to inform my teaching.

This teacher has connected evidence of learning to the starting point for teaching in the next unit of work. The teacher takes this thinking forward by identifying that even though this is a different unit of work (“I might not assess exactly the same thing again”), there are core or basic skills that will need to be taught or revised for learning progress. Thinking about teaching in this way is distinctly different from moving onto the next unit of work. In this case, evidence of learning is guiding pedagogic decisions. In another example, a teacher is thinking aloud their analysis of a student’s response:

Students need to understand the difference between physical and chemical changes ... [this] needs to be clear in the next teaching, making sure students understand and can differentiate between them and just being able to identify that this one here was this change because of this.

We can hear the teacher moving closer to the specifics of what needs to be taught. The teacher starts with a generic strategy, “understand the difference between physical and chemical changes” and concludes by specifying this as “identify[ing] that this one here was this change because of this”. This is now a tangible teaching point incorporating a framework for reasoning.

In another example, primary school teachers stated how they could address weaknesses in performances within the teaching of related disciplines.

Creating tables and graphs, they weren’t getting that very well ... Sometimes you’re, “Well, we’ve finished that unit”, but we’re

actually still going to do some maths, but with a science lens on. So, you can bring your science into your maths, your next steps.

This talk illustrates how the next-step teaching section in the template caused teachers to think about progressing student learning to include, across disciplines.

Teachers identified how the commentaries could enhance their current assessment processes by reminding them of decisions made in planning meetings when marking.

Sometimes we bring up past students’ work and be like, “This was the A, and this was the D”, and we just quickly all talk about it among us teachers and then just leave it at that. I do think having that commentary would then be something that we can give teachers, so then in eight weeks’ time between that meeting and when they’re actually marking, they can have that with them to then actually remember what we spoke about.

The teacher contrasts historic practices of using “past students’ work” with using the commentary as a memory aid. The commentary captures thinking about a judgment decision which is often forgotten after the discussion. The teacher notes the value of having this artefact to connect planning and marking or grading student work to support consistency in judgments.

The scaled exemplars and commentaries also provided an overview of the quality of a particular standard of performance and supported teachers to make difficult judgment decisions: “Even those ones you’re not too sure on, you can actually compare and go, ‘Okay, well these sort of are like the figurative language that the A standard’s using’“. Here the teacher has recognized a similar quality of performance in the exemplar and their own students’ work. Unstated but inferred from this comment is that while the assessment tasks and student responses in the exemplars may differ from the work samples that the teacher is grading, they are still able to discern a quality or standard of performance and use this to inform their judgment.

Finally, the conceptual features of the commentaries were considered by some teachers as useful for their reporting processes.

We’re redoing our parent-teacher interviews ... and we’re looking to that saying, “What are the strengths, mathematics, literacy, numeracy, areas of improvement, and our next steps for teaching?” So, conveying to the parents what we actually see next steps curriculum wise, not just social, emotional discussions that can take up your whole time. And then convey it like, “This is what your child has shown, this is what we recognise their strengths and areas for improvement are, and this is what we’re doing as our next steps” ... hence I’m excited to talk to you about it because we’re inspired by it.

The template has been the inspiration for a focus on learning, providing a framework for this focus with parents that concludes with pedagogic actions. The teacher stresses their point by contrasting historic parent-teacher meeting practices with the renewed focus on talking about student learning and direct connections to teaching.

## 6. Discussion

Previous models of teacher assessment identity, which have been largely built on research reviews and theoretical perspectives, have proposed that (1) assessment identity and teachers’ assessment decisions are based on various kinds of knowledges, and (2) teachers’ beliefs about the cognitive and affective dimensions of assessment act as an interpretative framework within a socio-political context (Looney et al., 2018; Xu & Brown, 2016). Our empirical study provides evidence that teacher assessment identity is deeply embedded within and constructed in relationship to employer policy expectations and local school contexts. Teachers’ identity within their school and year level teaching group contributed to their assessment identity as they collaborated to develop their assessment knowledge and skills, and worked towards



what they viewed as fair assessment processes. Professional learning was a consequence of teachers' deep analysis of student work using the cognitive commentary. Our findings identified that producing their own cognitive commentaries caused teachers to purposefully connect assessment, teaching, and learning.

The analysis of teacher talk showed the pervasiveness of assessment policy on teachers' accounts of their assessment practices with key terms and phrases seamlessly interwoven in teachers' descriptions of their local assessment practices. Explicit and implied beliefs about comparability and assessment were evident in these descriptions. For example, several aspects of fairness featured in the data, including teacher awareness of the influence of latent criteria, and 'bias suppression' (Rasooli et al., 2019, p. 15) in the creation of comparable assessment conditions for all students. Such an interplay is indicative of the balancing between horizontal and vertical assessment knowledges described by DeLuca et al. (2019). Our evidence shows teacher assessment identities as a balancing act as they worked to achieve comparability of their judgment decisions, and align curriculum, teaching, and assessment responsive to student learning needs; these decisions were always contextualized within school curricula and policy contexts, broadly shaping teachers' assessment identities.

The valuing of collaboration in assessment practices was clear in the data. Teachers described school cultures of formal and informal co-operation with colleagues as helping to shape and finalize their grading decisions. Teachers connected their shared understanding of achievement standards to improved teaching and assessment processes. Pride in their school assessment processes was evident in their descriptions and as they looked for ways to improve their assessment practices. Teachers' pride was also evident when they found that their school gradings were similar to others across the state, especially for those teachers from less advantaged and regional areas. Confidence in themselves as assessors increased when they discerned their judgments as consistent with others. The collaborative efforts of teachers to sustain an assessment practice and consistency in teacher judgments focused not only on an individual assessment identity but also on an identity as part of a school community. Evident here is a relationship between collaboration and Looney et al.'s (2018) self-efficacy dimension of assessment identity, where collaborative contexts of assessment appear to support more efficacious practice in teachers.

Two crucial implications of this research for teaching and teacher education are (1) the value of focused analysis of student work for building assessment skills and confidence; and (2) the prioritizing of collaborative activities, such as the development and use of resources, to support judgement making. The project has demonstrated the power of focused analysis of student work for professional learning as teachers identified a range of assessment and teaching practices to improve in their local contexts. In addition, targeting development on teaching strategies for progressing students achieving at an A standard was highlighted by teachers. Within our study, it was the cognitive commentary template that focused teachers' attention on how strengths and weaknesses of a response combined in an overall grade and led to consideration of next steps for teaching and learning, as well as reflection on improving related assessment practices. We propose that resources that support teacher analysis of student work in this way can be used in teacher education and across a career to support teachers in cultivating their assessment identities and engaging in professional learning about assessment. Activities that build assessment knowledge and skills can promote teachers' confidence in their role as an assessor. When teachers confidently use assessment to inform and improve their teaching and student learning, their identity as teacher-assessor is strengthened.

## 7. Conclusion

This study contributes to theory, research, and practice in teaching and teacher education by adding empirical evidence to the theory of

assessment literacy and establishing a linkage between teachers' assessment work and their identity. Our findings illustrate how the translation of assessment policy to practice is dynamic, embedded in context, and involves professional collaboration, targeted artefacts, and a willingness to reflect and enquire into practice. We have also identified how collaboration among teachers to develop shared understandings of achievement standards builds confidence in oneself as an assessor and promotes an assessment identity within a policy context.

This empirical study advances the broader, international field of teacher education and professional learning by providing evidence from teachers of how they work to achieve and learn about fairness in assessment, in particular consistency in their grading decisions. Findings illuminate how deep interrogation of student work against standard statements, through guiding resources, offers promise for developing assessment literacies and ultimately a professional identity as a confident assessor of student work for pre-service and in-service teachers.

## Funding

This research was partially funded by an Australian Research Council Linkage Project (ARC LP180100046). We acknowledge the funding received for this project from the Australian Research Council, Australian Catholic University, University of Western Australia and Partner Organisations: the Queensland Department of Education, Western Australian School Curriculum and Standards Authority, and Catholic Education Western Australia Limited.

## Financial interests

The authors have no relevant financial or non-financial interests to disclose.

## Ethics approval

This study was approved by the Human Research Ethics committee of the Australian Catholic University (Ethics approval number: 2019-11H) and partner organisations.

## CRedit authorship contribution statement

**Lenore Adie:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Data curation. **Claire Wyatt-Smith:** Writing – review & editing, Validation, Resources, Methodology, Funding acquisition, Conceptualization. **Mary Finch:** Writing – review & editing, Writing – original draft, Data curation. **Christopher DeLuca:** Writing – review & editing.

## Data availability

Data will be made available on request.

## Acknowledgments

Our sincere gratitude goes to the teachers in Queensland who contributed work samples, their time and expertise in this research project. We also acknowledge the contribution to the outcomes of this research by project team members including Stephen Humphry, David Andrich, and Terry Gallagher.

## References

- Adie, L., Gallagher, J., Wyatt-Smith, C., Spina, N., & DeLuca, C. (2023). Mediating teachers' assessment work. *Australian Educational Researcher*. <https://doi.org/10.1007/s13384-023-00675-z>
- Allen, R. (2012). Developing the enabling context for school-based assessment in Queensland, Australia. *Systems Approach for Better Education Results (SABER)*. <http://>

- [//documents.worldbank.org/curated/en/139961468206057475/Developin-g-the-enabling-context-for-school-based-assessment-in-Queensland-Australia](https://documents.worldbank.org/curated/en/139961468206057475/Developin-g-the-enabling-context-for-school-based-assessment-in-Queensland-Australia).
- Australian Curriculum, Assessment and Reporting Authority (ACARA). (2022a). Australian Curriculum. <https://acara.edu.au/curriculum>.
- Australian Curriculum, Assessment and Reporting Authority (ACARA). (2022b). *Planning, teaching, assessing and reporting*. <https://v9.australiancurriculum.edu.au/f-10-curriculum/f-10-curriculum-overview/planning-teaching-assessing-reporting>.
- Australian Government. (2022). *Australian education regulation 2013*. <https://www.legislation.gov.au/Details/F2022C00111>.
- Australian Institute for Teaching and School Leadership (AITSL). (2017). *Australian professional standards for teachers*. <https://www.aitsl.edu.au/standards>.
- Ayalon, M., & Wilkie, K. J. (2020). Developing assessment literacy through approximations of practice: Exploring secondary mathematics pre-service teachers developing criteria for a rich quadratics task. *Teaching and Teacher Education*, 89, 1–14. <https://doi.org/10.1016/j.tate.2019.103011>
- Bedford, A., & Barnes, N. (2022). The education minister's terrible, horrible, no good, very bad idea. *EduResearch Matters*. August 4, 2022 <https://www.aare.edu.au/blog/?p=13912>.
- Bourke, T., Ryan, M., & Ould, P. (2018). How do teacher educators use professional standards in their practice? *Teaching and Teacher Education*, 75, 83–92. <https://doi.org/10.1016/j.tate.2018.06.005>
- Braun, & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- DeLuca, C., Willis, J., Cowie, B., Harrison, C., & Coombs, A. (2023). *Learning to assess: Cultivating assessment capacity in initial teacher education*. Springer.
- DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., Gibson, A., & Trask, S. (2019). Policies, programs, and practices: Exploring the complex dynamics of assessment education in teacher education across four countries. *Frontiers in Education*, 4. <https://doi.org/10.3389/educ.2019.00132>
- Doyle, A., Johnson, M. C., Donlon, E., McDonald, E., & Sexton, P. J. (2021). The role of the teacher as assessor: Developing student teacher's assessment identity. *Australian Journal of Teacher Education*, 46(12), 52–68. <https://doi.org/10.14221/ajte.2021v46n12.4>
- Estaji, M., & Ghiasvand, F. (2022). Teacher assessment identity in motion: The representations in e-portfolios of novice and experienced EFL teachers. *Issues in Language Teaching*, 11(2), 33–66. <https://doi.org/10.22054/ilt.2022.70302.741>
- Estaji, M., & Ghiasvand, F. (2023). Tracing the dynamics of teacher assessment identity (TAI) through web-based audio diaries. *Qualitative Report*, 28(9), 2681–2715. <https://doi.org/10.46743/2160-3715/2023.5903>
- Feola, S., Lemons, P. P., Loertscher, J. A., Minderhout, V., & Lewis, J. E. (2023). Assessor in action: Assessment literacy development in a biochemistry context. *Chemistry Education: Research and Practice*, 24, 914–936. <https://doi.org/10.1039/d2rp00334a>
- Gan, L., & Lam, R. (2023). A tale of two teachers: Construction of novice university English teachers' assessment identity. *Chinese Journal of Applied Linguistics*, 46(2), 219–236. <https://doi.org/10.1515/CJAL-2023-0205>
- Garner, J. K., & Kaplan, A. (2019). A complex dynamic systems perspective on teacher learning and identity formation: An instrumental case. *Teachers and Teaching*, 25(1), 7–33. <https://doi.org/10.1080/13540602.2018.1533811>
- Ge, J. (2024). Supporting student teacher assessment identity development on online practice: A narrative inquiry. In P. Seitz, & L. Hill (Eds.), *Assessment of online learners: Foundations and applications for teacher education* (pp. 34–52). Routledge.
- Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*, 4<sup>th</sup> ed. Routledge. ISBN 9780415725569
- Hill, P. W., Brown, T., Rowe, K. J., & Turner, R. (1997). Establishing comparability of Year 12 school-based assessments. *Australian Journal of Education*, 41(1), 27–47. <https://doi.org/10.1177/000494419704100103>
- Humphry, S., Adie, L., Maxwell, C., & Sappl, S. (2023). Using pairwise comparison and ordered exemplars as a basis for a novel method of standard setting in narrative writing. *Frontiers in Education*, 8, 1–13. <https://doi.org/10.3389/educ.2023.1183695>
- Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, 75, 199–213. <https://doi.org/10.1016/j.tate.2018.06.015>
- Leigh-Lancaster, D., & Rowe, K. (1999). Extended common assessment tasks in VCE mathematics: Validity, reliability and other issues. In J. M. Truran, & K. M. Truran (Eds.), *Making the difference (proceedings of the 22nd annual conference of the mathematics education research group of australia)* (pp. 320–328). MERGA.
- Leonardsen, J. K., Utvær, B. K. S., & Fjørtoft, H. (2022). The five faces of an assessor: Conceptualizing the enactment of teacher assessment identity in vocational education and training. *Educational Assessment*, 27(4), 339–355. <https://doi.org/10.1080/10627197.2022.2106967>
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467. <https://doi.org/10.1080/0969594X.2016.1268090>
- Matters, G. (2006). *A report to the Queensland Studies Authority. Assessment approaches in Queensland senior science syllabuses*. [https://www.qcaa.qld.edu.au/downloads/publications/research\\_qsa\\_science\\_assess.pdf](https://www.qcaa.qld.edu.au/downloads/publications/research_qsa_science_assess.pdf).
- Matters, G., & Masters, G. (2014). *Redesigning the secondary–tertiary interface: Queensland review of senior assessment and tertiary entrance*. Australian Council for Educational Research (ACER). [https://research.acer.edu.au/qld\\_review/1/](https://research.acer.edu.au/qld_review/1/).
- Maxwell, G., & Cumming, J. (2011). Managing without public examinations: Successful and sustained curriculum and assessment reform. In L. Yates, C. Collins, & K. O'Connor (Eds.), *Australia's curriculum dilemmas: State cultures and the big issues* (pp. 202–222). Melbourne University Press. [https://www.acer.org/files/11\\_YATES.pdf](https://www.acer.org/files/11_YATES.pdf).
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>.
- Queensland Department of Education. (n.d.). Moderation processes at multiple junctures in the teaching and learning sequence. <https://www.youtube.com/watch?v=LOBugY5AJ.0>.
- Queensland Department of Education. (2020). *Curriculum into the classroom*. <http://education.qld.gov.au/curriculum/stages-of-schooling/C2C>.
- Queensland Department of Education. (2022). *P–12 Curriculum, assessment and reporting framework*. <https://education.qld.gov.au/curriculums/Documents/p-12-curriculum-assessment-reporting-framework.pdf>.
- Rasooli, A., Zandi, H., & DeLuca, C. (2019). Conceptualising fairness in classroom assessment: Exploring the value of organisational justice theory. *Assessment in Education: Principles, Policy & Practice*, 26(5), 584–611. <https://doi.org/10.1080/0969594X.2019.1593105>
- Skourdumbis, A. (2019). Theorising teacher performance dispositions in an age of audit. *British Educational Research Journal*, 45(1), 5–20. <https://doi.org/10.1002/berj.3492>
- Stanley, G., MacCann, R., Gardner, J., Reynolds, L., & Wild, I. (2009). *Review of teacher assessment: Evidence of what works best and issues for development*. Oxford University Centre for Educational Assessment. Qualifications and Curriculum Authority (QCA) [http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2011/01/2009\\_03-Review-of-teacher-assessment-QCA.pdf](http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2011/01/2009_03-Review-of-teacher-assessment-QCA.pdf).
- Talbot, D. (2016). Evidence for no-one: Standards, accreditation, and transformed teaching work. *Teaching and Teacher Education*, 58, 80–89. <https://doi.org/10.1016/j.tate.2016.05.006>
- Van Landveld, T., Schoolneboom, J., Volman, M., Croiset, G., & Beishuizen, J. (2017). Developing a teacher identity in the university context: A systematic review of the literature. *Higher Education Research and Development*, 36(2), 325–342. <https://doi.org/10.1080/07294360.2016.1208154>
- Wyatt-Smith, C., & Adie, L. (2021). The development of students' evaluative expertise: Enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, 53(4), 399–419. <https://doi.org/10.1080/00220272.2019.1624831>
- Wyatt-Smith, C., Adie, L., & Harris, L. (2024). Supporting teacher judgement and decision-making: Using focused analysis to help teachers see students, learning, and quality in assessment data. *British Educational Research Journal*, 00, 1–29. <https://doi.org/10.1002/berj.3984>
- Wyatt-Smith, C., Alexander, C., Fishburn, D., & McMahon, P. (2017). Standards of practice to standards of evidence: Developing assessment capable teachers. *Assessment in Education: Principles, Policy & Practice*, 24(2), 250–270. <https://doi.org/10.1080/0969594X.2016.1228603>
- Wyatt-Smith, C., & Bridges, S. (2008). *Meeting in the middle: Assessment, pedagogy, learning and educational disadvantage*. <https://education.qld.gov.au/literacy/docs/deewr-myp-final-report.pdf>.
- Wyatt-Smith, C., & Colbert, P. (2014). *An account of the inner workings of standards, judgement and moderation: A previously untold evidence-based narrative*. Australian Council for Educational Research (ACER). [https://research.acer.edu.au/qld\\_review/7/](https://research.acer.edu.au/qld_review/7/).
- 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(1), 35–52. <https://doi.org/10.1080/0969594X.2012.725030>
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>