Received: 6 October 2023 | Accepted: 15 January 2024

DOI: 10.1002/berj.3984

ORIGINAL ARTICLE

BERJ BERA

Supporting teacher judgement and decision-making: Using focused analysis to help teachers see students, learning, and quality in assessment data

Claire Wyatt-Smith | **Lenore Adie** | **Lois Harris**

Australian Catholic University, Brisbane City, Queensland, Australia

Correspondence
Lenore Adie, Institute for Learning Sciences and Teacher Education, Australian Catholic University, 229 Elizabeth St, Brisbane City, Qld 4000, Australia.
Email: lenore.adie@acu.edu.au

Funding information
Department of Education and Training, Queensland Government; Western Australia School Curriculum and Standards Authority; Australian Catholic University; University of Western Australia; Australian Research Council; Catholic Education Western Australia

Abstract

This paper reports results from an Australian study into how teachers see features of quality in student work and connect these to next-step teaching. Data were drawn from a national 3 year project investigating teacher judgement using A–E standards. The project developed scaled exemplars of authentic student written performance assessments to support teacher judgement and inform next-step teaching and learning. Fifty-seven participants created evaluative explanations of grading decisions (i.e. cognitive commentaries), wrote reflective responses and participated in online meeting discussions of their cognitive commentaries. These data were examined using qualitative content and thematic analyses. Findings highlighted how stated standards influence teachers' judgements of student work. Variation was evident in teachers' focus on (1) content (i.e. core skills vs. extended thinking), (2) specificity of improvement points and suggested teaching strategies and (3) the connection between identified areas for improvement and teaching strategies. Most teachers were able to identify specific points for improvement. However, in the main, they selected next-step teaching strategies that were general rather than targeted and specific. Teachers reflected that structured analysis helped direct their thinking and judgements, targeting attention on next-step teaching. The results suggested

that teacher education and professional development should focus on supporting teachers to link specific teaching strategies to identified student weaknesses. This study found that artefacts, such as cognitive commentaries that connect assessment, teaching and learning, can help build professional knowledge and expertise, which remain key components of teacher assessment literacy.

KEYWORDS

assessment literacy, next-step teaching, performance assessment, teacher judgement

Key insights

What is the main issue that the paper addresses?

The paper examines how teachers see features of quality in student work and connect these to next-step teaching. It explores how structured analysis of student performance assessment data via a cognitive commentary allowed teachers to connect evidence of learning to standards and use these to devise next steps teaching strategies.

What are the main insights that the paper provides?

The study found that teachers used the cognitive commentary process to direct their thinking and judgements, identifying possibilities for improving teaching, student learning and assessment tasks. Data showed that teachers were better able to identify specific areas for improvement than specific strategies to support learning in these areas.

INTRODUCTION

This paper presents new insights from an Australian study into how teachers see features of quality in student work and connect these to next-step teaching. The empirical study focused on tasks requiring complex thinking in English, mathematics and science. The paper contributes to understandings about (1) teachers' interpretation of assessment evidence of student learning and (2) their use of this evidence to formulate clear, actionable next-step teaching strategies. While grading usually signals the end of a task and often a unit of work, this study looked at how the process and thinking of judgement making can be used to inform future teaching and identify goals to progress learning. The paper also presents scaffolded approaches for teachers to examine and reflect on their assessment practice using standards, exemplars and cognitive commentaries.

See[ing] quality is understood in the paper through Kress' (2000, 2009) notion of how learning may be recognised and expressed. Teachers' knowledge of the curriculum and their expectations of how learning may be represented influence their assessment of that

learning. Their openness to considering different perspectives of how others see features of quality in a given work sample will influence how shared understandings of quality may be developed with colleagues and students. Thus, developing shared understandings is understood as dialogic, in which teachers share how they see quality, and transformative in which teachers and students come to recognise quality and develop as assessors.

A cognitive commentary is an explanation of how an overall judgement of student work is reached (Smith, 1995; Wyatt-Smith & Bridges, 2008). The commentary captures teachers' thinking and decision-making about qualities in student work. Preparing it involves a structured analysis of student performance, identifying elements that come to the fore in appraising performance, considering both strengths and weaknesses evident in the work. The commentary is attached to the work sample and teachers use it to *anchor* their enacted judgement to a stated, official standard of quality. Rather than remaining invisible, these aspects of teachers' work can become tangible referents to develop shared understandings of quality, and for ongoing teaching and assessing.

Other approaches that have been trialled to document teachers making judgements include (1) think-aloud protocols (e.g. Phung & Mitchell, 2022; Suto & Greatorex, 2008) and (2) processes to explain and justify judgements within moderation discussions (e.g. Black et al., 2011; Wyatt-Smith et al., 2010). Both approaches rely on recorded talk. A distinguishing feature of cognitive commentaries as presented in this paper is the pedagogical focus: connecting identified areas for improvement in student work with next-step teaching decisions for the whole class and individual students. This type of structured analysis has potential to enhance teacher reflection on their assessment practices and result in pedagogic actions.

Four related research questions, linking teachers' suggested areas for improvement and their proposed teaching strategies, are addressed in this paper:

1. What aspects of learning do teachers identify as areas for improvement when they engage in focused analysis of student work?
2. What are the characteristics of the next-step teaching strategies teachers propose as a result of their focused analysis of student work?
3. How do teacher-identified areas for improvement in student work align with their proposed next steps for teaching and learning?
4. How does a structured process of analysis assist teaching and assessment decision-making?

HUMAN JUDGEMENT AS A KEY ELEMENT OF TEACHER PROFESSIONALISM

Human judgement is a core element in a wide range of professions, underpinning the translation of evidence into appropriate action (e.g. sport—Collins & Collins, 2021; medicine—Ludolph & Schultz, 2018; nursing—Nibbelink & Brewer, 2018; and social work—Taylor & Whittaker, 2018). Sadler (1989) described this type of human judgement as qualitative and defined it as 'made directly by a person, the person's brain being both the source and the instrument for the appraisal. Such a judgment is not reducible to a formula which can be applied by a non-expert' (p. 124). Broadly speaking, sound judgement relies on discernment and the use of evidence, alongside the ability to minimise the impact of both bias (i.e. systematic deviation) and noise (i.e. random scatter) on decisions (Kahneman et al., 2022). Sound judgements require thoughtful use of evidence and are based on objective opinions, without trace of bias or whimsey. They are generated wisely, authoritatively and with discretion and discrimination, especially in matters affecting action, choices or decisions.

However, in the twenty-first century, the rise of machine scoring, artificial intelligence (AI) and intelligent tutoring programs has led to questions about the place of humans in judgement making and how working with these tools impacts notions of professionalism. Reliance on algorithms is claimed to save time and bring objectivity to the process, reducing human error and subjectivity. For example, Pearsons (2023) argued that their *Intelligent Essay Assessor* brings accuracy, consistency, agility and efficiency to the assessment process.

One compelling reason for retaining and strengthening human judgement is the need for results to be made defensible to assessment users. Wyatt-Smith et al. (2021) have noted that 'machine learning, AI and algorithms have the potential to black box decision making' (p. 7), meaning that people may no longer understand how decisions have been made. In education, it is generally accepted that teachers' professional judgement is central to valid and reliable decisions of complex performances (Sadler, 2009; Smith, 1989). For example, Valentine et al. (2021) have argued that 'Changing focus to look at what is "fair" human judgement in assessment, rather than what is "objective" human judgement in assessment allows for the embracing of many different perspectives and allows for the legitimising of human judgement in assessment' (p. 2). In their model, they proposed that fair judgement decisions are transparent, credible, fit for purpose, defensible and supported by individual (e.g. evidence, boundaries, agility, expertise, narrative) and system factors (e.g. procedural fairness, documentation, multiple opportunities, multiple assessors, validity evidence).

Models of teacher assessment literacy foreground the importance of teachers being able to 'interpret evidence of student learning [and] use data on learning to adjust instruction and adapt curriculum' (Pastore & Andrade, 2019, p. 135) and engage in 'assessment decision-making and action-taking' (Xu & Brown, 2016, p. 155). These understandings highlight the importance of both judgement as an evaluation of performance quality, and decision-making as the ability to base actions on assessment evidence and reflect on their effectiveness. Research has also highlighted the importance of teachers developing their students' understanding of what quality looks like and their skills to reflect on the efficacy of their own learning approaches (e.g. Harris & Brown, 2018; Wyatt-Smith & Adie, 2021).

Within this paper, judgement is understood from a sociocultural perspective and considered to be a socially situated, cognitive act (Wyatt-Smith et al., 2010). Teacher judgement is understood as a core professional practice in which teachers demonstrate and share expertise in assessment and make decisions to monitor student learning and support progression (Allal, 2013; Biesta, 2015). The study is situated in an educational context in which teacher judgements are informed by standards, written in qualitative terms, as quality indicators. How teachers read, interpret and understand standards is dependent on the various social and cultural contexts in which they work. Teacher expertise in the application of standards is needed to judge the variety of responses that assessment, in particular complex performance assessments, can bring forth.

CLASSROOM ASSESSMENT DATA

Schools are recognised to be awash with data (Farrell & Marsh, 2016), collected for diverse purposes (e.g. student and school accountability and improvement; Brown, 2008). Often these data take the form of numerical results from standardised tests or school surveys. Barnes et al. (2022) have called for 'the definition of data to include actual artefacts that schools and classrooms collect on a regular basis and make these data the focus of data team meetings or other instructional decision-making initiatives' (p. 282). Such data are often referred to as classroom assessment, defined as 'a process in which teachers and students gather evidence of student performance to make decisions about further instruction and grading' (McMillan, 2023, p. 519). This definition includes both

teachers and students, underscoring the need for students to develop their evaluative expertise (Wyatt-Smith & Adie, 2021), consistent with self-regulated learning perspectives (Panadero, 2017).

Classroom assessment data can be used for both formative (i.e. improving student competence, teaching and learning; Sadler, 1989) and summative purposes (i.e. summarising achievement for reporting; Sadler, 1989). For example, the formative assessment movement has spent over two decades championing the use of classroom assessment data to inform next steps in teaching (e.g. Black et al., 2003; Black & Wiliam, 1998, 2018). Educational accountability approaches argue that assessment data can help teachers, school leaders and members of the public to identify needed areas of improvement and motivate progress in these areas (Cizek, 2001). Data-based decision-making (DBDM) also provides a mechanism for assessment data to be used for school-wide improvement (e.g. Lai & Schildkamp, 2016). However, many studies have focused on the judgement of individual tasks (e.g. Jansen & Möller, 2022; Phung & Michell, 2022) or how results combine to determine overall grades (e.g. Allal, 2013; Jönsson et al., 2021), without extending to how such data translate to improvement. Assessment-based actions are seldom documented in research and may be inadvertently omitted from the data use cycle (Adie et al., 2020). Furthermore, research has tended to focus more on students performing below grade level even though assessment evidence is valuable for directing all students' progress (Datnow & Hubbard, 2015).

FROM CLASSROOM DATA TO IMPROVEMENT ACTIONS

Translating assessment data to appropriate improvement actions is complex. For assessment data to drive improvement, teachers and students need to be able to identify then action appropriate classroom strategies to address gaps in learning. As Maxwell (2021) noted, 'Simply knowing that there is need for improvement is not sufficient to produce improvement unless a strategy can be devised for doing so' (p. 19). This involves having a clear goal, an understanding of the learner and his or her current strengths and weaknesses, and specific and pedagogically appropriate strategies that can be employed to support progress. To make quality judgements and use these to support learning (Suto & Greatorex, 2008) teachers need content knowledge (CK) and pedagogical content knowledge (PCK), along with understandings of standards, the learners and the context.

Content knowledge and pedagogical content knowledge

Shulman (1986) defined CK as 'the amount and organization of knowledge per se in the mind of the teacher' (p. 9). Content knowledge helps teachers understand how students develop expertise within the discipline and identify common misconceptions. Shulman (1986) described PCK as 'subject matter knowledge for teaching' (p. 9). Teachers draw on PCK to develop realistic learner goals, requiring understanding of discipline progression and contextually appropriate next steps for learning.

Since Shulman's (1986) coining of CK and PCK, others have drawn on these ideas using a variety of terms (e.g. content knowledge for teaching; Ball et al., 2008). Within assessment, Shulman's (1986) CK and PCK are explicitly drawn on within models of assessment or data literacy (e.g. Mandinach & Gummer, 2016; Xu & Brown, 2016), assessment competence (e.g. Herppich et al., 2018), and required teacher assessment knowledges and skills (e.g. Brookhart, 2011). Additionally, Haug and Ødegaard (2015) identified that gaps in CK and PCK limited teachers' abilities to effectively act on formative assessment data.

Knowledge of standards, curriculum and learners

Besides CK and PCK, within the classroom context, teachers also draw on understandings of standards, curriculum requirements, and knowledge of learners and their histories (Wyatt-Smith et al., 2010). Within Australia, the national curriculum provides stated standards that teachers use for grading and reporting. Standard descriptors provide a basis for teachers to develop shared understandings of the quality of work being assessed. Establishing shared understandings of quality is a dialogic process in which teachers jointly review grades against established standards and criteria (Sadler, 2009; Willis & Adie, 2013). Such processes often occur as the focus of professional learning groups, year level meetings or during social moderation (Adie et al., 2023). Other countries like England also have a tradition of using standards, with standardisation training and moderation both used to improve judgement consistency for General Certificate of Secondary Education (GCSE) tasks (Black et al., 2011).

The benefits of social moderation and teacher professional discussions have been well documented. Research has found that involvement in focussed discussions of student work samples, such as in moderation meetings, can deepen teachers' assessment knowledge and skills (Matre & Solheim, 2016; Smaill, 2020; Wyatt-Smith & Bridges, 2008), lead to improved judgement consistency and feedback (Wyatt-Smith et al., 2010) and extend content and pedagogical knowledge (Adie et al., 2023; DeLuca et al., 2017; Farrell & Marsh, 2016; Wyatt-Smith & Gunn, 2009).

Alongside professional discussions, other practices identified to support consistency and comparability of grading performances are the use of exemplars that illustrate a standard of work, commentaries of how judgements were made, and opportunities to develop assessment expertise (Sadler, 2009; Smith, 1989). The use of exemplars within professional learning groups has been shown to support teacher and student understanding of expected quality features (Chong, 2021; To et al., 2022). However, the use of calibrated exemplars, with only limited training and moderation opportunities for teachers, still 'produced high levels of inter-rater reliability and concurrent validity' (Heldsinger & Humphry, 2013, p. 233).

Teachers' epistemological frames, expertise and experience also impact on what they attend to or how they see quality in student work (Kress, 2000, 2009). This includes aspects such as their personal beliefs about subject priorities, and the impact of school priorities, expectations and grading policies. Research has shown that grading decisions may also be informed by implicit standards and latent criteria which have varying degrees of educational relevance (e.g. neatness, teacher perceptions of intelligence or effort, student attitude; Castleton et al., 2003; Randall & Engelhard, 2010; Wyatt-Smith & Klenowski, 2013). Teachers need to understand their students when making pedagogical decisions (McKnight & Morgan, 2023), while not allowing latent criteria to influence their standards-based judgements. Professional discussions and exemplars can support teachers to bring together relevant information while guarding against extraneous influences.

Determining appropriate strategies

Once teachers consider key learner characteristics and histories and identify how the performance relates to quality expectations (e.g. standards) and expected disciplinary progression, they then draw on PCK to determine appropriate strategies for progressing learning. This step has been reported as challenging for teachers, although working collaboratively on these aspects of practice has been shown to be effective (DeLuca et al., 2017; Farrell & Marsh, 2016; Matre & Solheim, 2016). Research has also suggested that identifying

student weaknesses is easier for teachers than devising strategies to address them (Gupta et al., 2018; Heritage et al., 2009; Schneider & Gowan, 2013).

Many studies have documented teachers drawing on PCK when using data for formative purposes to identify next teaching and learning steps, both in the moment and via planned intervention (e.g. Andrade & Heritage, 2017; Cisterna & Gotwals, 2018; Furtak et al., 2016; Sun et al., 2016). However, fewer studies have explored teachers using data from assessments for summative purposes to improve learning (e.g. Choi et al., 2022; Hoover & Abrams, 2013; Sun et al., 2016). Overall, studies that have examined teachers' proposed or enacted strategies based on assessment data have identified (1) broad categories of next-step strategies (e.g. changing instructional groupings, Choi et al., 2022; Hoover & Abrams, 2013), (2) variation in the quality and appropriateness of teachers' proposed strategies (e.g. Cisterna & Gotwals, 2018; Hoover & Abrams, 2013; Sun et al., 2016), (3) teachers' strategies as underspecified (e.g. Horn et al., 2015) and (4) limited changes made to daily instruction (Farrell & Marsh, 2016; Hoover & Abrams, 2013). For example, in their study of formative assessment in science, Cisterna and Gotwals (2018) reported that teacher focus and subsequent strategy use generally centred around improving core skills via correction (i.e. scientific knowledge and facts) rather than progressing the deeper understandings of science focal to the curriculum. Timing was a key issue, with teachers perceiving the data came too late to inform teaching (Sun et al., 2016).

TAKEN AND MISSED OPPORTUNITIES IN CLASSROOM ASSESSMENT

Despite evidence that teachers often devalue or fail to use data from tasks designed for summative purposes, other research has suggested that these can be valuable sources of information about student learning to inform next-step teaching decisions (Black et al., 2003; Fives & Barnes, 2020). Not using these data is a missed opportunity in classroom assessment. Students may put more effort into graded tasks (Wise & Smith, 2016) and these often allow for more comprehensive demonstration of knowledge and skills, including both core (unistructural and multistructural; Biggs & Collis, 1982) and extended (relational and extended abstract; Biggs & Collis, 1982) aspects of learning.

Data may also provide opportunities for individual and collaborative teacher reflection (Xu & Brown, 2016), and teacher and student evaluation of the effectiveness of teaching and learning approaches (Wyatt-Smith & Adie, 2021). While practices like DBDM have formal evaluation processes within the data use cycle, these are seldom systematically applied to classroom assessment contexts. There is a need to further explore how teachers and students may be supported to use data from graded assessments of extended performance (e.g. assignments, reports, speeches, live performances) to improve teaching and learning. The study considers how the use of a process of structured analysis of student performances may help teachers take advantage of such assessment data to progress next-step learning and reflection on practice.

RESEARCH CONTEXT

This paper is based on a large nationally funded project that utilised psychometric scaling of student work samples and online moderation to improve the consistency of teacher standards-based judgements (Humphry et al., 2023). The project was conducted across two Australian states (Queensland and Western Australia). Australian education is a standards-based system, with a national curriculum and a national assessment program of literacy

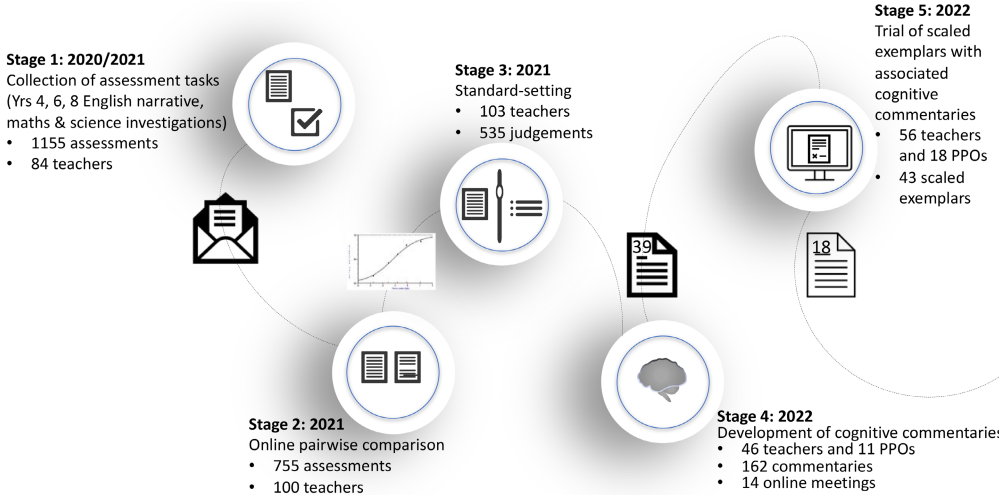

FIGURE 1 Project stages. [Colour figure can be viewed at wileyonlinelibrary.com]

and numeracy (NAPLAN). Learning progressions are available for informing instruction in literacy and numeracy. While the Australian Curriculum Achievement Standards identify year level expectations and have accompanying annotated work sample portfolios (satisfactory, above and below the Standard), a framework linking standards, exemplars and related commentaries on judgement decisions is not available. Data for this paper involved teachers from the state of Queensland. Queensland has strong traditions relating to standards-based assessment and moderation, as per state policy (Queensland Department of Education, 2022; Queensland Government, 2014), dating back more than five decades (Maxwell & Cumming, 2011; Sadler, 1987; Smith, 1995).

The project systematically investigated the use of scaled exemplars to make comparable judgements against achievement standards, with a focus on the explication of judgement decisions. The project involved five stages (Figure 1) and employed a mixed-method research design. It focused on the disciplines of English, science and mathematics and year levels 4, 6 and 8 (students aged 9, 11 and 13 years). This paper draws on data from Stage 4 of the project involving Queensland teachers.

METHODS

To address this paper's research questions, data were drawn from Stage 4 of the larger ARC Linkage Project (Figure 1).

Sample and data sources

This paper drew on data from 46 teachers and 11 Department of Education Principal Project Officers (PPOs); PPOs are registered teachers, seconded from the classroom to support school curriculum implementation. All PPOs were Brisbane-based; teachers came from schools with diverse locations ($n = 31$ major cities, $n = 10$ inner regional, $n = 5$ outer regional) and socio-economic statuses (Index of Community Socio-Educational Advantage range 900–1151, mean = 1032; Australian Curriculum, Assessment and Reporting Authority [ACARA], 2014).

TABLE 1 Response lengths (number of words) for improvement points and next-step strategies per discipline and year level.

		Improvement points		Next-step strategies	
Discipline	**Year level**	**Range**	**Mean**	**Range**	**Mean**
English	Year 4	13–131	70	26–177	70
	Year 6	31–337	109	45–181	101
	Year 8	51–315	149	38–161	96
Mathematics	Year 4	14–234	83	12–152	53
	Year 6	12–180	63	13–161	56
	Year 8	18–289	113	23–177	78
Science	Year 4	32–345	148	22–280	94
	Year 6	10–241	88	7–109	41
	Year 8	53–214	133	23–179	65

The study drew on teacher-designed and implemented assessments rather than standardised tests. All tasks were designed as open ended and required core and extended thinking. All participants were experienced in marking and providing student feedback, but none had previously written cognitive commentaries of their judgement decisions (Supplementary Material, Figure S1 provides the template with expandable boxes). In writing cognitive commentaries, teachers were asked to identify the strengths and areas for improvement in each performance, and how they combined these features to reach an overall grade. They were also prompted to identify next-step teaching strategies to progress the student's learning.

Project Stage 4 included three types of data. First, teachers and PPOs produced 162 cognitive commentaries of grading decisions ($n=45$ English, $n=63$ mathematics, $n=54$ science) in response to scaled student work samples (identified through Stages 2 and 3; Humphry et al., 2023). The level of detail varied substantially within submitted cognitive commentaries (Table 1). Improvement points were longer, on average, in all instances except Year 4 English. Second, all participants completed a reflection form about their experience of writing cognitive commentaries. Third, teachers participated in audio-recorded online meetings (year level and discipline specific; duration of 120–150 minutes) where the submitted cognitive commentaries were discussed and the wording of the final commentary attached to a scaled exemplar was agreed upon (e.g. Year 4 English, B exemplar). Meeting recordings were transcribed verbatim.

Data analysis

We analysed the content of the teacher-submitted cognitive commentaries (RQs1, 2, 3) and the teachers' perspectives on writing their cognitive commentaries via their reflections and online meeting transcripts (RQ4). Across datasets, we drew on Miles et al.'s (2014) work to guide our approach to the coding process, simultaneously drawing on data and on our knowledge of existing research to generate codes. To examine aspects of content within cognitive commentary artefacts, we conducted qualitative content analysis (Mayring, 2015). Our aim in providing frequencies was to show the broader data patterns we were observing; these are illustrated qualitatively for transparency via tables, figures (e.g. Figure 4, Tables 4 and 5) and our codebook (Supplementary Material, Table S1).

Cognitive commentaries were first read in their entirety to identify how teachers documented areas for improvement and next-step teaching strategies. One author initially

TABLE 2 Coding for learning (core/extended) identified in improvement points and teaching strategies.

Codes	Solo taxonomy levels	Curriculum focus
Core learning	Unistructural; multistructural	*Science report*: use of scientific language, grammar/spelling, report structure, knowledge of scientific principles, description of methods *Mathematical investigation*: mathematical computation, use of mathematical language and conventions, report structure, knowledge of principles of mathematics *English narrative*: grammar, mechanics, spelling, basic descriptive and narrative writing (e.g. having an orientation, describing characters)
Extended learning	Relational; extended; abstract	*Science report*: explain, analyse, justify and draw conclusions, reflect on process, formulate improvements. *Mathematical investigation*: explain mathematical thinking, analyse data, draw conclusions, justify results, formulate improvements *English narrative*: audience directly considered, language used for specific purposes (e.g. creating a hook for the reader), more sophisticated narrative elements (e.g. climax, resolution)

organised distinct text segments into separate rows within Excel spreadsheets, with a second checking these organisations. Authors worked together to generate initial descriptive codes and form a preliminary codebook. Two authors used this codebook to assign codes to each text segment relating to content (i.e. core/extended, drawing on Biggs & Collis, 1982, SOLO taxonomy; Table 2), specificity (Table 3; Figures 2 and 3) and alignment of improvement points with proposed teaching strategies. As interrater reliability is argued as important for content analysis reliability (Mayring, 2015), we used Cohen's Kappa (1960) to calculate agreement between two researchers using the preliminary codebook. These initial interrater reliability calculations showed moderate to substantial agreement (Supplementary Material, Table S2) and motivated further refinement of codes and their definitions, with coding disagreements resolved via discussion to reach consensus and form the final codebook.

The content focus of improvement points (RQ 1) and proposed teaching strategies (RQ 2) was first classified by the complexity of the student thinking required (Table 2). Tasks had been designed to evaluate both core and extended curriculum aspects, as specified in the Australian Curriculum (ACARA, 2023). Codes, aligned with the SOLO taxonomy (Biggs & Collis, 1982), were created to differentiate between core learning (correlating to SOLO taxonomy unistructural and multistructural levels) and extended learning (correlating to relational and extended abstract aspects of the curriculum).

Table 3 shows how we distinguished between general and specific improvement points (RQ 1). While the example coded as 'general' identified a broad improvement focus (i.e. paragraphing), the specific structural issues were not identified in the work. The comment coded as specific provided detail of the areas within the work pinpointed as an improvement focus.

Figure 2 shows how we distinguished between general and specific teaching strategies (RQ 2), with examples from English provided. This distinction is represented in a quadrant showing general or specific 'what' (i.e. content focus for the next-step teaching) and general or specific 'how' (i.e. the strategy for addressing this 'what'). Figure 3 illustrates the possible configurations of improvement points and teaching strategies that could arise from our analysis.

We then analysed connections between identified areas for improvement and recommended strategies (RQ 3). These were classified as aligned (i.e. improvement area corresponding to a strategy addressing the weakness), improvement area only (i.e. no aligned strategy) or strategy only (i.e. no aligned weakness). Multiple strategies could relate to a particular improvement area and vice versa.

To further understand teachers' experiences of judgement- and decision-making when writing their cognitive commentaries, we conducted qualitative analyses of survey responses and meeting data (RQ 4). This involved first and second cycle coding to generate themes relating to participant experiences (Miles et al., 2014). Four researchers initially identified preliminary codes withing these data, using NVivo to assign data to codes. Two authors examined the codes and, via second cycle coding, grouped data into seven themes: (1) engaging with standards; (2) improving consistency, fairness and justification; (3) identifying paths for student improvement; (4) capturing assessment thinking for future use; (5) building teacher confidence; (6) reflecting on assessment practice; and (7) considering workload.

RESULTS

The process of writing a cognitive commentary focused the teachers' gaze on features within students' work that contributed to the grade. All teachers also formulated some strategies for next-step teaching. However, there was variation in the cognitive commentary data,

TABLE 3 Example of coding for specificity of improvement points.

Codes	Explanation of code	Example
General	Broad statements identifying area for improvement. Readers have to infer the problem areas within the work	Some structural issues with paragraphing (Year 8, English, C sample)
Specific	Enough detail provided for the reader to pinpoint where a specific type of improvement was needed (i.e. through an example, a named specific problem or aspect of work)	There are also elements which the reader cannot understand because they are not explained. For example, what is Planet Watermelon and why is it named so/what's its special purpose? Who are these people to each other? This would explain and fill in some of the major plot holes in the story (Year 8, English, B sample)

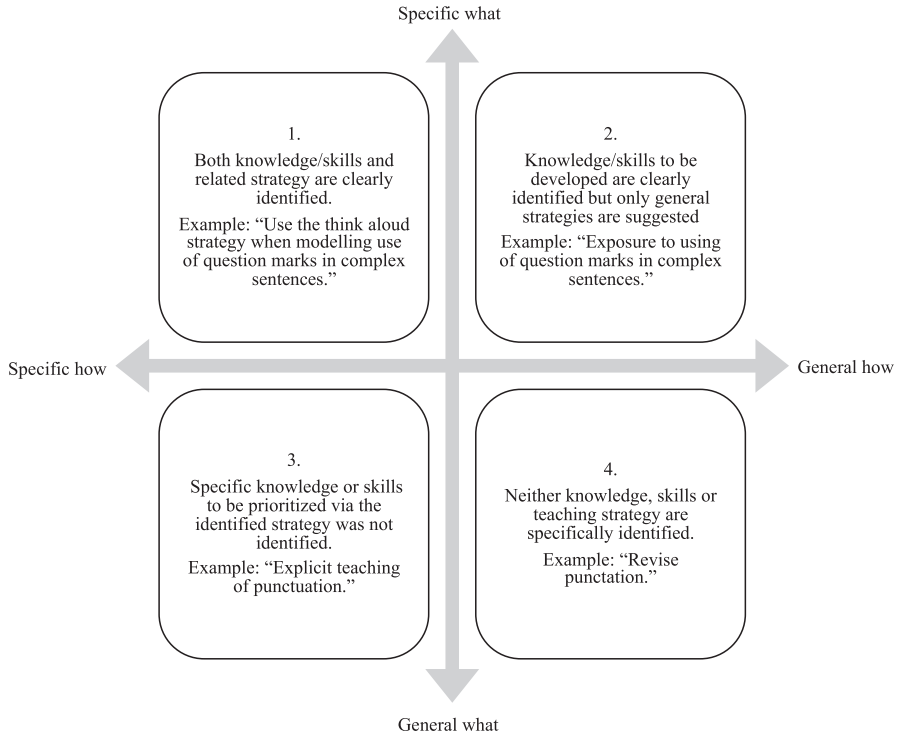

FIGURE 2 Explanation and examples of coding for specificity of teaching strategy (what/how).

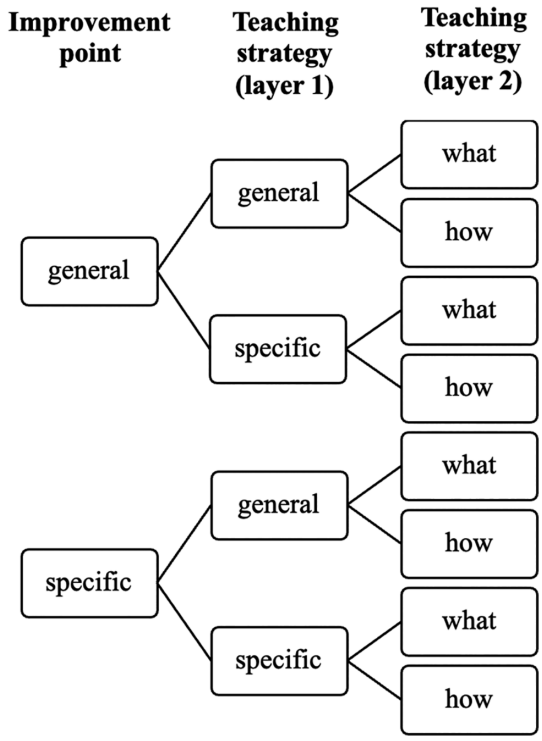

FIGURE 3 Possible configurations of improvement points and suggested teaching strategies.

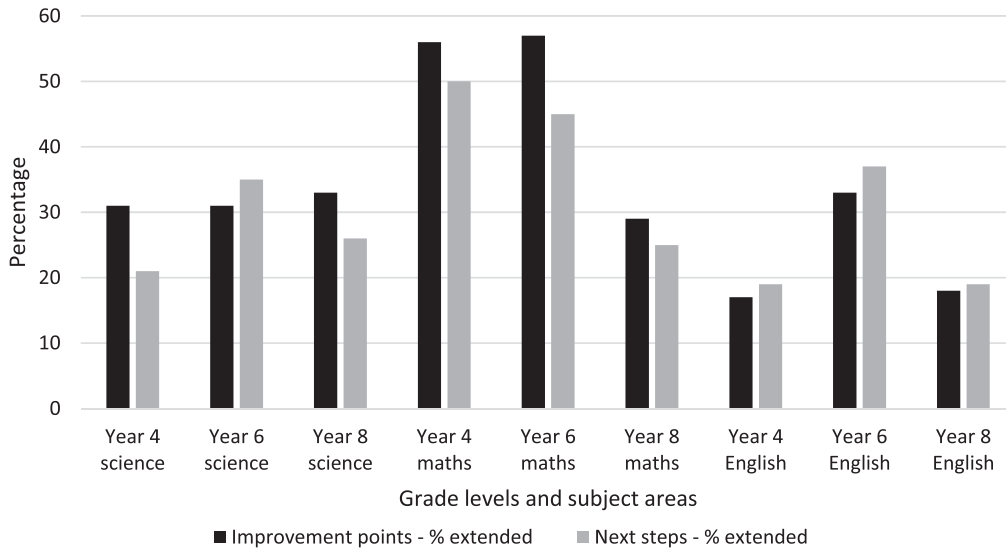

FIGURE 4 Percentage of comments focused on extended subject content.

with identified improvement areas and proposed strategies ranging substantially in their content focus (core and extended skills), specificity (general and specific points) and alignment (between the improvement points and the suggested teaching strategies).

Content of improvement points and teaching strategies (core, extended)

All teacher-provided assessment tasks were designed to elicit both core and extended skills. Figure 4 shows the percentage of comments where teachers' improvement points and strategies focused on extended skills, with core skills being the focus of all other comments. As extended skills differ across discipline areas, comparing results across subjects is not appropriate. However, the data do show that apart from Year 4 and 6 mathematics, around 30% or less of comments focused on extended skills, making core skills the focus of the majority of comments across subject areas.

Specificity of improvement points and teaching strategies (general, specific)

Within each discipline, the submitted cognitive commentaries contained a range of general to specific improvement points, followed by either general or specific suggestions for next-step teaching strategies (categorised further into 'what' knowledge and skills, and 'how' these were to be taught; Figure 3). Across disciplines, it was evident that teacher comments were classified as specific contained knowledge, skills and next-step teaching suggestions that were embedded in the discipline and relevant to the student work sample. Strategies ('how') were identified as general when teachers drew on educational jargon without providing examples (e.g. 'high-yield strategies') or named a broad approach (e.g. 'provide opportunities for', 'encourage', 'scaffold', 'focus on') without specifying how this would occur. The qualitative differences between comments highlighted possible variation in teacher CK and PCK.

TABLE 4 Illustrative examples of general and specific teacher comments about improvement points (Year 4, Science, B standard, extended).

Science: Improvement point	
General	**Specific**
Adequate understanding is shown, but they require more support to provide explanations based on science	The student was able to compare results but was unable to justify why the results presented this way. They have only justified it by saying 'wool is usually warm' or 'if you look at the results' but haven't explained WHY that is. Why is wool usually warm?

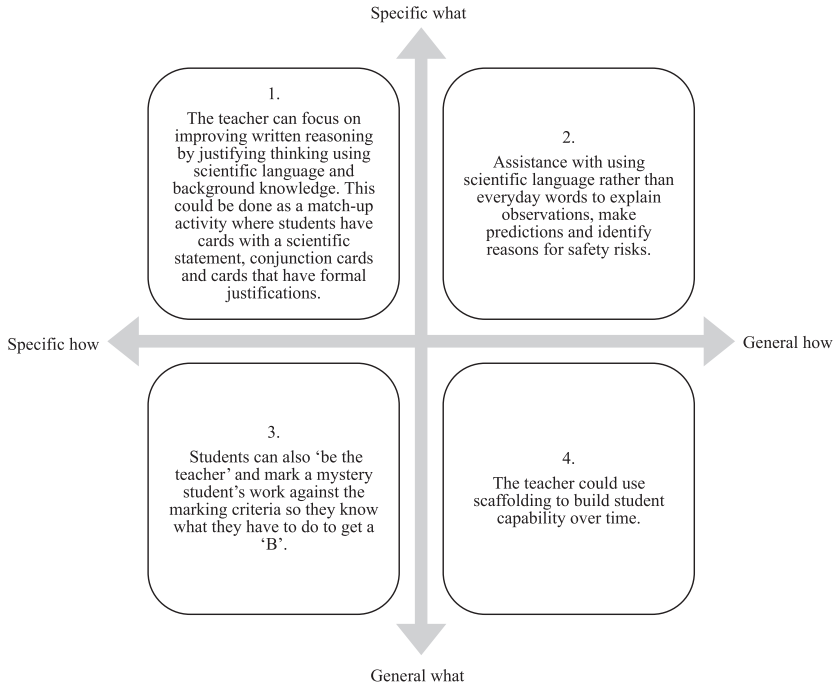

FIGURE 5 Illustrative examples of general and specific teacher comments (what/how) for next-step teaching (Year 4, Science, B standard, extended).

Table 4 shows how two teachers' improvement points differed when considering the same Year 4, B standard sample of a scientific explanation (Supplementary Data Tables S3, S4 and Figures S2, S3 provide English and mathematics examples). The general comment identified that the student had an 'adequate understanding' but not 'based on science'. Readers would need to share an understanding of the knowledge and skills necessary for an 'adequate understanding', and the year-level-appropriate scientific knowledge and skills required in an explanation. In contrast, the specific statement identified the skill that required further development (i.e. stating but not explaining reasons), providing a particular example from the text to illustrate the problem.

Figure 5 provides examples of next-step teaching strategies which four teachers suggested to support this student to write an improved scientific explanation in response to their analyses of the Year 4, B standard sample. The specific 'what' quadrants (1 and 2) identified the actual knowledge or skill to be taught, that is, using scientific language and knowledge in explanations. The general 'what' quadrants (3 and 4) only identified that the knowledge and skills were related to a 'B level' or 'student capability'. The specific 'how' quadrants (1 and 3) provided strategies that taught the knowledge or skill: for example, participating in a

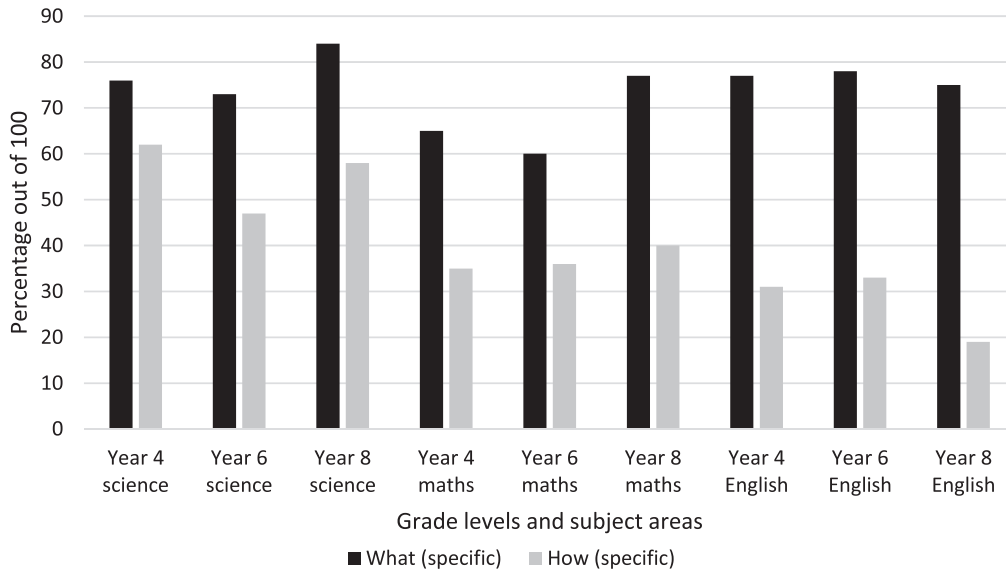

FIGURE 6 Percentage of next-step teaching strategies (what, how) coded as specific.

match-up card activity and marking other students' work against criteria ('be the teacher'). In contrast, the general 'how' quadrants (2 & 4) identified only 'assistance' or 'scaffolding' as next-step strategies.

Science had the highest percentage of specific improvement points; however, this difference may be due to differing tasks and work samples, the coding scheme or the teacher sample. When examining the next-step teaching strategies, a more distinct pattern was found (Figure 6). Across grade levels and subject areas, teachers were more specific about the focus for improvement (what) than the corresponding strategies (how). This finding was similar to those of other studies that have also reported that teachers were better able to identify areas for improvement than specific strategies to address these (e.g. Gupta et al., 2018; Heritage et al., 2009; Schneider & Gowan, 2013). While teachers were able to list a range of potentially applicable generic strategies, the identification of clear, actionable strategies was less frequent.

Alignment of identified improvement areas and next steps for teaching and learning

Alignment between the identified improvement areas and next-step teaching strategies was frequently found (Table 5). Examples of aligned comments are provided in Table 6. While a generic improvement point and generic strategy could be aligned, lack of alignment sometimes occurred because the generic strategies named could not be paired to an improvement point (e.g. 'Giving students feedback on their answers and showing them what they need to get the next level of marking'; Year 4 Science). Instances where alignment was not achieved could reflect unintentional omission of an improvement point or strategy, temporary inability to think of a strategy when writing the cognitive commentary, or more substantive gaps in CK (e.g. not knowing what the next learning should be based on discipline progression) or PCK (e.g. inability to match particular weaknesses with appropriate strategies).

TABLE 5 Alignment between improvement areas and strategies.

	Aligned improvement area and strategy	Improvement area only	Strategy only
English	149	29	55
Mathematics	148	41	93
Science	176	55	86

Teacher evaluative feedback of structured analysis to support their assessment knowledge and skills

Data from teacher reflections and meetings were analysed to understand how the process of creating a cognitive commentary may have informed assessment practice. Seven themes were identified: (1) engaging with standards; (2) improving consistency, fairness and justification, (3) identifying paths for student improvement; (4) capturing assessment thinking for future use; (5) building teacher confidence; (6) reflecting on assessment practice; and (7) considering workload. These themes are described and illustrated below. Further examples are provided in Table 7, with key phrases related to themes bolded.

Across the data set, teachers identified that writing a cognitive commentary assisted in *engaging with standards*. As one Year 6 English teacher explained in a reflection, 'I have now a clearer expectation of achievement levels and I feel I can communicate this better with my team'. Statements like this highlight how understanding standards in educational contexts is not privatised or individual work; teachers are socialised around assessment knowledges and practices. Discussions during the project's online meetings illustrated how teachers reasoned and justified their interpretations of standards and work samples. For example:

> I agree that the B sample demonstrated an overall B level, but I would have had issues with the student not giving any explanations for their understandings. It would have been a low B … as they did not demonstrate their ability to justify or explain. (Year 4 science teacher, B standard)

In this example, the teacher identified 'explanation' as a critical feature to be awarded a B standard.

Teachers argued that using the scaled exemplars could *improve consistency, fairness and justification* processes. For example, a Year 4 mathematics teacher's reflection stated, 'If exemplars of samples were given to teachers this could help determine levels much clearer than just gut instinct. I think having reasons as to why one sample is an A and not a B could be very helpful to have grades more consistent'. In this way, while not becoming standards themselves, exemplars and corresponding commentaries were argued to help make visible what standards looked like within student work samples. Teachers also noted that the focus on standards and the possibility of blind marking made judgements fairer as the work became the basis of the judgement rather than knowledge and assumptions about the student (e.g. perceptions of student ability or effort).

Teachers reflected on how the commentaries could also be useful in teacher meetings to *identify paths for student improvement*. For example,

> I really liked the inclusion of 'Areas for improvement' and 'Next steps for teaching'. We often spend so much time marking (all learning areas) and moderating (mainly English) that not much time is given to actually using the task to identify

TABLE 6 Illustrative examples of alignment between improvement areas and strategies.

Year level, discipline, standard	Improvement comment	Next steps strategy
Year 6 English E standard	Consideration needs to be given to the correct spelling of common words such as 'everything' (everythink) (core, specific)	The student will use a common word checker to self-assess for the accurate spelling of common words which will include words from the student's writing that are commonly misspelt (core, specific what, specific how)
Year 8 mathematics B standard	Strengths and limitations are not discussed (extended, general)	Opportunity needs to be provided for students to identify strengths and limitations of models and solutions. Emphasis needs to be on these strengths and limitations being about the process as well as the final model/solution. This teaching should also involve the identification of relevant observations and assumptions, and how to appropriately document these considerations rather than merely stating them (extended, specific what, specific how)
Year 4 science B standard	The explanation of results needs to give more detail by justifying results with knowledge of the materials' properties—why didn't the paper 'go well'? (extended, specific)	The teacher can focus on improving written reasoning by justifying thinking using scientific language and background knowledge. This could be done as a match-up activity where students have cards with a scientific statement, conjunction cards and cards that have formal justifications (extended, specific what, specific how)

TABLE 7 Further examples of data illustrating RQ4 themes.

Theme	Examples
Engaging with standards	It really makes you analyse exactly why you're giving someone that grade and by continuously having to refer to the evidence, you have to justify to yourself. … You really have to look at the data. **I kept going back to the standard descriptor**, and yeah, I learned a lot (Year 4 mathematics, meeting) This process has **solidified the importance of the achievement standard**. Without a deep understanding of that and its relationship to the other curriculum elements, it is difficult for teachers to **explain how the evidence in student work aligns to the standard**. This has allowed me to reflect on why I utilise the language of **'looking for evidence of the standard'** rather than 'marking' when working with teachers to undertake moderation processes (Year 4 English, meeting)
Improving consistency, fairness and justification	Sometimes teachers feel like this kid is a B, I want to give them a B because they are a B. But if you **give it to different eyes** that don't know them, you can get those decisions made a bit quicker: 'they haven't got this maths, they're not really explaining it'. Then you can go yeah, **fair point** (Year 4 mathematics, meeting) Teachers completing the cognitive commentary on a couple of their own students to share could be really powerful not just in **ensuring consistency of teacher judgement** but also in **allowing teachers time to discuss students' learning and teaching practice** (Year 6 science, reflection)
Identifying paths for student improvement	I've also realised **I need to invest more in finding ways for top level students to improve more**—there are less obvious improvements for these students to make … (Year 8 English, reflection) I have learned that often the next step for teaching isn't necessarily harder questions but **teaching how to justify their results** (Year 6 math, reflection)
Capturing assessment thinking and appraisal process for future use	I do think having that commentary would then be something that we can give the teachers, so then **in the 8 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** (Year 8 science, meeting) Cognitive commentaries that support samples of work being presented at different levels of moderation would **provide the thinking behind assigned grades** without the need for the teacher to verbally recall the thought process (Year 6 mathematics, reflection) These samples will provide great **exemplars of different levels of work for students to observe** as well, giving far greater **clarification to them of mark-level expectations** than any written document ever could (Year 8 English, reflection)
Building teacher confidence	Being so early in my career … **I want to be able to have the confidence** … **to stand by a judgement I've made**. I noticed in my first year of teaching … a teacher would come along and go … 'That doesn't look like a B to me'. … and I'd go 'Okay, yep, yep, sure, it's a C, yep, okay, no worries' and I wouldn't stand by my own call. But now **through this process I feel like I have authority behind me** … I also, **believe in myself more**. … I feel like I have the **confidence** to go 'No, this here is a strength, or this here is where a student hasn't actually justified, or this is a personal opinion. That is not B standard' (Year 6 mathematics, meeting) It has also **boosted my confidence** and understanding of using marking and moderation as a tool to understand **where students are at** and what I can do as a teacher to **support their next steps** (Year 6 mathematics, reflection)
Reflecting on assessment practice	**That the quality of the assessment task plays a major role in determining the level of student achievement.** Some tasks were unclear, and the questions did not provide an opportunity for the student to demonstrate the criteria—hence **this process helped them come to realise more re assessment tasks** (Year 4 science, reflection) It has been a valuable process and I feel like I have learnt a lot from being part of it. It has **influenced my own practice**, which is exactly what I look for in quality professional development (Year 6 mathematics, reflection)

TABLE 7 (Continued)

Theme	Examples
Considering workload	The process would be useful during moderation at a school; however, there could be **concerns about workload for teachers**. But the ability to use the commentaries as a conversation prompt would be highly valuable (Year 4 mathematics, reflection) Cognitive commentaries would be useful when moderating samples of student work, however **the time they take to develop would only add to teacher's workload**. If they were **developed by Head of Curriculum** staff, they would be a welcome addition to the teaching and assessing cycle (Year 4 mathematics, reflection)

> areas for individual student improvement or collating next steps for teaching … It really is a missed opportunity. (Year 6 science teacher, reflection)

Teachers also noted that documenting their decisions and potential next steps in the cognitive commentary provided opportunities for shared practice and *captured their assessment thinking for future uses*:

> While teachers will often do this mentally, completing the cognitive commentary on live student samples across a cohort of teachers would definitely provide opportunities to gain a shared understanding of the achievement standard, identifying student needs and sharing practice around next steps for learning. (Year 6 science teacher, reflection)

Other teachers identified that the cognitive commentary data could be used (1) to inform their own professional reflections and classroom planning, (2) as examples which could support new and out-of-field teachers build professional knowledge and (3) as the basis for feedback to students and parents.

The process was discussed as *building teacher confidence* by developing knowledge of curriculum and assessment and improving their trust in their judgements. It also supported *reflecting on assessment practice*, including (1) their own marking processes, (2) their abilities to identify next learning steps for students and (3) the quality of their assessment tasks. For example, teachers noted that the structured analysis of student responses helped them identify when tasks did not provide opportunities for students to demonstrate particular levels of achievement.

Finally, teachers commented that while this process was beneficial, *considering workload* was necessary when implementing 'these types of activities'. For example,

> Providing opportunities for this type of activity has great benefit for ensuring the consistent interpretation of achievement standards, understanding of student learning and the sharing of practice. These types of activities need to be supported by giving teachers time to undertake these tasks. I am conscious of how crowded the curriculum currently is and how overwhelmed teachers are increasingly feeling across systems and sectors. (Year 6 science, reflection)

Teachers noted the professional development value of creating and/or having access to exemplars and corresponding cognitive commentaries but were mindful that it was only feasible to examine a small sample rather than the whole class. Some felt that Heads of Department might be best placed to create cognitive commentaries to use as Department examples, while others suggested the learning gained from writing them was valuable for all teachers.

DISCUSSION

This study has shown that structured analysis of student work has the potential to support teachers' judgements of complex performances both during grading and when reflecting on assessment and teaching practices. Participating teachers said that the process facilitated deep analysis of student work, helped them devise potential next learning steps for the student and prompted reflection on the quality of the assessment task and their broader work as teachers. However, teachers highlighted that structured analysis of student work takes time, noting the need for school leaders to set aside space for this kind of activity and consider how to strategically use the process with selected student samples.

Analyses conducted in response to our first two research questions highlighted the current variation in content and specificity in the identified improvement points and proposed next-step teaching strategies. Foregrounded was the importance of both CK and PCK in the application of these professional skills. While commentaries that linked specific areas for improvement in student work with specific teaching strategies were produced, some included only limited connection to CK, which made it difficult to precisely identify student weaknesses and next learning steps. Particularly for teachers who are unsure of the development of knowledge and skills within a particular discipline, learning progressions (LP) and corresponding LP-based assessment tools may be helpful for identifying current levels of understanding and guiding teacher decision-making around the content and key skills to focus on in next-step teaching strategies (Harris et al., 2022).

Some cognitive commentaries evidenced a limited depth of PCK informing next-step teaching strategies. While teachers identified a range of strategies, many were generic strategies and not necessarily aligned with specific areas of need. In addition, the suggested next-step teaching strategies focused more frequently on core rather than extended aspects of curriculum, a finding consistent with some other studies (e.g. Cisterna & Gotwals, 2018). This suggests that teachers may benefit from support on how to link improvement points to strategies embedded in the discipline and which develop more complex aspects of curriculum.

Teachers who participated in the study, across all grade levels and subject areas, were able to be more specific about areas for improvement (what) than they were about strategies to address identified weaknesses (how). This finding aligns with research suggesting that teachers are better at noticing a problem in student work than they are of devising a specific strategy for addressing the concern (e.g. Gupta et al., 2018; Heritage et al., 2009; Schneider & Gowan, 2013), making this an important area for future teacher professional development. As it was outside the scope of this study to observe teachers enacting their proposed strategies, it is also important for future work to examine whether and how strategies are implemented in effective ways.

The study's findings and the reviewed literature build the basis for a model of teachers' qualitative, evidence-based judgement and decision-making in complex performance assessments (Figure 7). This model suggests that making evidence-informed decisions involves seeing students in and through their work, recognising when learning has occurred and expressing this in ways that capture features of quality. Teachers can use these data to customise current and future learning and feedback in ways that correspond with a student's learning goals and needs and which support progress towards expected educational standards.

The proposed model expands Kress's (2000, 2009) notion of seeing learning into components of the work sample, the learner and teaching strategies, in which teachers:

1. *see* features of quality in the work, discerning how these relate to standards;
2. *see* the learner, knowing the child and their learning history; and

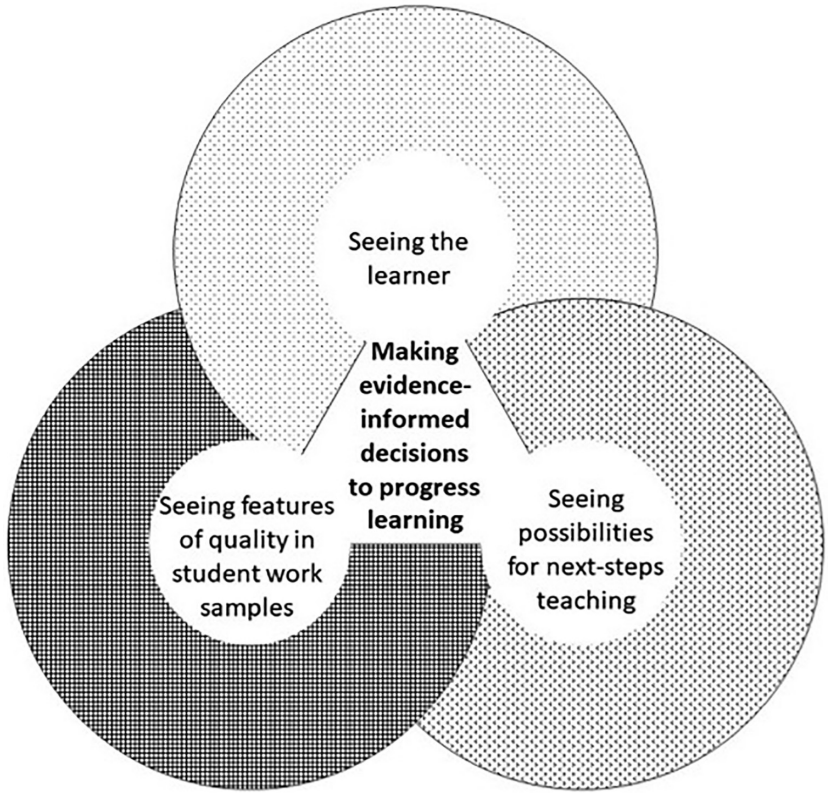

FIGURE 7 Model of teacher professional judgement and decision making.

3. *see* possibilities for next-step teaching, identifying appropriate strategies effective for this learner.

The model suggests that the combination of these three components has the potential to connect assessment to teaching in focussed ways to progress learning.

To be able to see quality in the work, teachers require knowledge of standards and how to use them when appraising student work. Data from this study and others (e.g. Black et al., 2011; Chong, 2021; DeLuca et al., 2017; Smaill, 2020) have highlighted the importance of artefacts (e.g. exemplars and cognitive commentaries) and professional discussion (e.g. focussed learning groups or social moderation) for helping teachers make sense of stated standards and anchor their judgements in the student's work.

Seeing the learner is central in the proposed model. Teachers come to see the learner in and through the student's work. To do this, a strong foundation in CK can support teachers to identify the learner's current level of performance relating to both core and extended aspects of curriculum and how they combine within complex performances. In normal classroom contexts, this CK must then be drawn together with knowledge of the learner (e.g. How does this performance align with previous performances? What teaching and learning strategies have/have not been successful in the past?). This includes teachers being able to consider elements of the performance alongside those from previous assessments, even when the task context has changed (e.g. Can I see growth in use of descriptive language across the student's most recent three tasks?). Teachers must also tap into their own knowledge of the student's motivations and history of engagement to form judgements about if

the work represents the student's learning (i.e. Is this valid evidence of the student's current knowledge and skills?) and consider how proposed strategies may help enhance motivation. Developing a teacher mindset of seeing student learning needs to be made habitual.

Seeing possibilities for next-step teaching requires teachers to draw on PCK particularly around identifying strategies that respond to learning gaps and knowledge of the learner. This study has showed that teachers have a vast repertoire of strategies, however the matching of strategies to learning gaps is more difficult. This study has demonstrated the need for better specification of, and connection between, improvement points and strategies is often couched in general terms. Yet research has called for assessment literate teachers who can use assessment data to tailor their teaching for the student (Black & Wiliam, 2018; Pastore & Andrade, 2019; Xu & Brown, 2016). The school environment and working conditions also need to be conducive for proposed instructional changes to be implemented (Farrell & Marsh, 2016).

This study's findings and the model developed from them have implications for those seeking to support teacher judgement. The model highlights that there is much for a teacher to consider when making dependable human judgements about student learning. Processes of structured analysis of student work, like the cognitive commentary as used in the project, are promising in that they prompt teachers to identify areas for improvement and connect these to next steps for teaching and learning. Teachers in the project suggested that the cognitive commentary could be useful across different pedagogic activities, with the standard as the connecting thread (e.g. during planning to develop shared understanding of the standards, during teaching to illustrate the standards to students, before and during marking as a reminder of the standards, and in moderation meetings as a record of the thinking that informed the judgement).

The variability in specificity and alignment within the cognitive commentaries analysed in the study signals a need for additional professional learning opportunities in this area. We can speculate that gaps in the teacher's disciplinary content knowledge may result in a more general focus within improvement points and next-step teaching strategies. It was evident that teachers who identified generic strategies to progress learning did not connect these directly to the student's work. To make discipline specific responses to data, linking assessment, teaching and learning, teachers require skills in recognising specific areas of strength or weakness, selecting specific strategies, and aligning these with areas of need. Future studies could examine how artefacts like cognitive commentaries could be analysed to identify teacher professional development needs and then tailor learning to support their growth. Skill development would focus on deep understanding of strategies, when they might be applied, and for what purpose (McKnight & Morgan, 2023). This could include the development of student judgement capabilities or evaluative expertise, a need foregrounded in the literature (e.g. Black & Wiliam, 2018; Wyatt-Smith & Adie, 2021).

How to progress student learning from current levels to realistically obtainable learning goals is a crucial professional skill for teachers. While there are many models of assessment literacy highlighting the range of knowledges teachers require for their assessment practice (e.g. Xu & Brown, 2016) and lists of teacher assessment competencies (e.g. Brookhart, 2011), bringing together and defining these knowledges and capabilities into a model of assessment content knowledge may be a useful next step. Assessment content knowledge could focus on the specific knowledges and skills required within assessment contexts, with a need to establish how it may work with CK and PCK to connect teaching, learning and assessment. Professional learning focused on these knowledges may serve as a way for teachers to come to know themselves as assessors and understand their judgements in more informed ways, allowing them to move effectively between teaching the whole class and individual students.

Limitations

This research was conducted during Covid-19 pandemic disruptions, making participant recruitment difficult. While this study drew on a large sample of teachers from diverse school contexts across the state of Queensland, owing to recruitment challenges, samples were unevenly distributed across subject and grade levels. Future studies would benefit from having a stratified, representative sample.

Teachers in this sample had not written cognitive commentaries before; hence, their ability to complete these effectively may increase with practice. Also, teachers were not working with their own students' work, so they could not reflect on the effectiveness of teaching, nor did they have personal understandings of the students' particular learning needs or motivation. While some teachers did note the benefits of blind marking (e.g. avoiding considerations around effort and ability), our model suggests that, within normal classroom situations, teachers draw on knowledge of the learner and learning history along with disciplinary knowledge to make effective pedagogic decisions.

CONCLUSION

There is agreement within the field that data should be used to improve learning, connecting it to teaching and assessment. We know very little about how teachers analyse and use complex performance assessment data and how they tune into their decisions about quality in student work. This study contributed towards filling this gap by analysing teachers' written artefacts, focussing on areas for improvement and their proposed next-step teaching strategies in Queensland, Australia. The process of creating a cognitive commentary engaged teachers in deep analysis of student work and promoted connection to a suite of professional knowledge, including assessment task design, judgement and decision-making around next-step teaching. These data highlighted the variation in, and importance of, teacher CK and PCK in assessment. The study identified a greater teacher focus on (1) core rather than extended thinking, and (2) generic rather than specific teaching strategies. Teacher discussions emphasised the range of considerations when making and acting on judgements of student performance. Teacher reflections highlighted the usefulness of structured analysis to improve their teaching and assessment practices and thereby foster learner growth.

This study also showed the work to be done to better support teachers' use of complex performance data to formulate clear, actionable, next-step teaching strategies. Teachers suggested the portability of the cognitive commentary as a record of judgement decisions to (1) inform collaborative planning meetings, (2) draw on during teaching, (3) refer to when assessing student work, (4) inform feedback and (5) use in moderation meetings. Possible ways forward are to use artefacts such as the cognitive commentary as professional development to mentor teachers new to the field or discipline and promote shared understandings of standards and teaching strategies aligned to improvement areas. They may also be a potentially valuable resource in initial teacher education, allowing preservice teachers to engage in the mechanics of appraising student work to make overall judgements, justify decisions according to stated standards and use this information to plan appropriate next steps for teaching. Through focussed analysis of student work, skills to see quality—recognising learning and expressing this in order to progress learning—may be promoted. Cultivating these judgement skills could help strengthen teacher professionalism and encourage decision making that is embedded in quality features evident in the work alongside knowledge of the learners and their contexts.

ACKNOWLEDGEMENTS

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, Chris De Luca, David Andrich and Terry Gallagher. Open access publishing facilitated by Australian Catholic University, as part of the Wiley - Australian Catholic University agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

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.

CONFLICT OF INTEREST STATEMENT

We have no known conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

Data is available on request from the corresponding author. Materials may be accessed at https://onlinemoderation.com.au/.

ETHICS AND INTEGRITY POLICY STATEMENT

This research has been funded by an Australian Research Council Linkage Project (ARC LP180100046) and conducted under ethics approval by Australian Catholic University (ethics approval number 2019-11H).

ORCID

Claire Wyatt-Smith https://orcid.org/0000-0003-1190-9909
Lenore Adie https://orcid.org/0000-0003-3874-397X
Lois Harris https://orcid.org/0000-0002-7379-3676

REFERENCES

Adie, L., Gallagher, J., Wyatt-Smith, C., Spina, N., & DeLuca, C. (2023). Mediating teachers' assessment work. *Australian Educational Researcher*, 1–21. https://doi.org/10.1007/s13384-023-00675-z

Adie, L., Harris, L., & Wyatt-Smith, C. (2020). Examining research into the use of data walls for teaching and learning: How are they being implemented within data use cycles? *Teaching and Teacher Education*, *89*(103012), 1–16. https://doi.org/10.1016/j.tate.2019.103012

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 20–34. https://doi.org/10.1080/0969594X.2012.736364

Andrade, H. L., & Heritage, M. (2017). *Using formative assessment to enhance learning, achievement, and academic self-regulation*. Routledge. https://doi.org/10.4324/9781315623856

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2014). *What does the ICSEA value mean?* https://docs.acara.edu.au/resources/About_icsea_2014.pdf

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2023). *Australian Curriculum (Version 9)*. https://v9.australiancurriculum.edu.au/

Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Barnes, N., Brighton, C. M., Fives, H., Meyers, C., & Moon, T. R. (2022). Where's the data to support educators' data use for instructional practice? *Theory Into Practice*, *61*(3), 277–287. https://doi.org/10.1080/00405841.2022.2096379

Biesta, G. (2015). What is education for? On good education, teacher judgement, and educational professionalism. *European Journal of Education*, *50*(1), 75–87. https://doi.org/10.1111/ejed.12109
Biggs, J. B., & Collis, K. F. (1982). Evaluating the quality of learning: The SOLO taxonomy. *Academic Press*. https://doi.org/10.1016/C2013-0-10375-3
Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, *18*(4), 451–469. https://doi.org/10.1080/0969594X.2011.557020
Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGraw-Hill Education (UK).
Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102
Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807
Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x
Brown, G. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. Nova Science Publishers.
Castleton, G., Wyatt-Smith, C., Cooksey, R., & Freebody, P. (2003). The nature of teachers' qualitative judgements: A matter of context and salience. Part two: Out-of-context judgements. *The Australian Journal of Language and Literacy*, *26*(2), 33–42.
Choi, K. M., Hwang, J., Jensen, J., & Hong, D. S. (2022). Teachers' use of assessment data for instructional decision making. *International Journal of Mathematical Education in Science and Technology*, *53*(4), 1010–1017. https://doi.org/10.1080/0020739X.2021.1880653
Chong, S. W. (2021). University students' perceptions towards using exemplars dialogically to develop evaluative judgement: The case of a high-stakes language test. *Asian-Pacific Journal of Second and Foreign Language Education*, *6*(1), 1–27. https://doi.org/10.1186/s40862-021-00115-4
Cisterna, D., & Gotwals, A. W. (2018). Enactment of ongoing formative assessment: Challenges and opportunities for professional development and practice. *Journal of Science Teacher Education*, *29*(3), 200–222. https://doi.org/10.1080/1046560X.2018.1432227
Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, *20*(4), 19–27. https://doi.org/10.1111/j.1745-3992.2001.tb00072.x
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104
Collins, D., & Collins, L. (2021). Developing coaches' professional judgement and decision making: Using the 'big 5'. *Journal of Sports Sciences*, *39*(1), 115–119. https://doi.org/10.1080/02640414.2020.1809053
Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, *117*(4), 1–26. https://doi.org/10.1177/016146811511700408
DeLuca, C., Bolden, B., & Chan, J. (2017). Systemic professional learning through collaborative inquiry: Examining teachers' perspectives. *Teaching and Teacher Education*, *67*, 67–78. https://doi.org/10.1016/j.tate.2017.05.014
Farrell, C. C., & Marsh, J. A. (2016). Contributing conditions: A qualitative comparative analysis of teachers' instructional responses to data. *Teaching and Teacher Education*, *60*, 398–412. https://doi.org/10.1016/j.tate.2016.07.010
Fives, H., & Barnes, N. (2020). Navigating the complex cognitive task of classroom assessment. *Teaching and Teacher Education*, *92*, 1–17. https://doi.org/10.1016/j.tate.2020.103063
Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science*, *44*, 267–291. https://doi.org/10.1007/s11251-016-9371-3
Gupta, D., Soto, M., Dick, L., Broderick, S. D., & Appelgate, M. (2018). Noticing and deciding the next steps for teaching: A cross-university study with elementary pre-service teachers. In G. J. Stylianides & K. Hino (Eds.), *Research advances in the mathematical education of pre-service elementary teachers: An international perspective* (pp. 261–275). Springer. https://doi.org/10.1007/978-3-319-68342-3
Harris, L. R., Adie, L., & Wyatt-Smith, C. (2022). Learning progression–based assessments: A systematic review of student and teacher uses. *Review of Educational Research*, *92*(6), 996–1040. https://doi.org/10.3102/00346543221081552
Harris, L. R., & Brown, G. T. (2018). *Using self-assessment to improve student learning*. Routledge.
Haug, B. S., & Ødegaard, M. (2015). Formative assessment and teachers' sensitivity to student responses. *International Journal of Science Education*, *37*(4), 629–654. https://doi.org/10.1080/09500693.2014.1003262

Heldsinger, S. A., & Humphry, S. M. (2013). Using calibrated exemplars in the teacher-assessment of writing: An empirical study. *Educational Research*, *55*(3), 219–235. https://doi.org/10.1080/00131881.2013.825159

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, *28*(3), 24–31. https://doi.org/10.1111/j.1745-3992.2009.00151.x

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Boehmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Boehmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education*, *26*(3), 219–231. https://doi.org/10.1080/08957347.2013.793187

Horn, I. S., Kane, B. D., & Wilson, J. (2015). Making sense of student performance data: Data use logics and mathematics teachers' learning opportunities. *American Educational Research Journal*, *52*(2), 208–242. https://doi.org/10.3102/0002831215573773

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/feduc.2023.1183695

Jansen, T., & Möller, J. (2022). Teacher judgments in school exams: Influences of students' lower-order-thinking skills on the assessment of students' higher-order-thinking skills. *Teaching and Teacher Education*, *111*, 103616. https://doi.org/10.1016/j.tate.2021.103616

Jönsson, A., Balan, A., & Hartell, E. (2021). Analytic or holistic? A study about how to increase the agreement in teachers' grading. *Assessment in Education: Principles, Policy & Practice*, *28*(3), 212–227. https://doi.org/10.1080/0969594X.2021.1884041

Kahneman, D., Sibony, O., & Sunstein, C. R. (2022). *Noise: A flaw in human judgement*. HarperCollins Publishers.

Kress, G. (2000). "You've just got to learn how to see": Curriculum subjects, young people and schooled engagement with the world. *Linguistics and Education*, *11*(4), 401–415. https://doi.org/10.1016/S0898-5898(00)00030-9

Kress, G. (2009). Assessment in the perspective of a social semiotic theory of multimodal teaching and learning. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 19–41). Springer. https://doi.org/10.1007/978-1-4020-9964-9_2

Lai, M. K., & Schildkamp, K. (2016). In-service teacher professional learning: Use of assessment in data-based decision-making. In G. T. L. Brown & L. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 77–94). Routledge. https://doi.org/10.4324/9781315749136

Ludolph, R., & Schulz, P. J. (2018). Debiasing health-related judgments and decision making: A systematic review. *Medical Decision Making*, *38*(1), 3–13. https://doi.org/10.1177/0272989X17716672

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011

Matre, S., & Solheim, R. (2016). Opening dialogic spaces: teachers' metatalk on writing assessment. *International Journal of Educational Research*, *80*, 188–203. https://doi.org/10.1016/j.ijer.2016.07.001

Maxwell, G. S. (2021). *Using data to improve student learning: Theory, research and practice*. Springer. https://doi.org/10.1007/978-3-030-63539-8

Maxwell, G. S., & Cumming, J. J. (2011). Managing without public examinations: Successful and sustained curriculum and assessment reform in Queensland. In L. Yates, C. Collins, & K. O'Connor (Eds.), *Australia's curriculum dilemmas: State cultures and the big issues* (pp. 186–206). Melbourne University Press.

Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 365–380). Springer. https://doi.org/10.1007/978-94-017-9181-6_13

McKnight, L., & Morgan, A. (2023). "Good" choices vs "what really works": A comparison of evidence-based practice in medicine and education. *The Australian Educational Researcher*, *50*(3), 643–659. https://doi.org/10.1007/s13384-022-00509-4

McMillan, J. (2023). Classroom assessment. In R. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International encyclopedia of education* (4th ed., pp. 519–528). Elsevier Science.

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). Qualitative data analysis: A methods. In *Sourcebook*. SAGE.

Nibbelink, C. W., & Brewer, B. B. (2018). Decision-making in nursing practice: An integrative literature review. *Journal of Clinical Nursing*, *27*(5–6), 917–928. https://doi.org/10.1111/jocn.14151

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 1–28. https://doi.org/10.3389/fpsyg.2017.00422
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
Pearsons. (2023). *Automated scoring*. https://www.pearsonassessments.com/large-scale-assessments/k-12-large-scale-assessments/automated-scoring.html
Phung, D. V., & Michell, M. (2022). Inside teacher assessment decision-making: From judgement gestalts to assessment pathways. *Frontiers in Education*, *7*, 1–20. https://doi.org/10.3389/feduc.2022.830311
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
Queensland Government. (2014). *Education (Queensland curriculum and assessment authority) act 2014*. https://www.legislation.qld.gov.au/view/pdf/2017-06-30/act-2014-003
Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, *26*(7), 1372–1380. https://doi.org/10.1016/j.tate.2010.03.008
Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, *13*(2), 191–209. https://doi.org/10.1080/0305498870130207
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714
Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, *34*, 807–826. https://doi.org/10.1080/03075070802706553
Schneider, M. C., & Gowan, P. (2013). Investigating teachers' skills in interpreting evidence of student learning. *Applied Measurement in Education*, *26*(3), 191–204. https://doi.org/10.1080/08957347.2013.793185
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004
Smaill, E. (2020). Using involvement in moderation to strengthen teachers' assessment for learning capability. *Assessment in Education: Principles, Policy & Practice*, *27*(5), 522–543. https://doi.org/10.1080/0969594X.2020.1777087
Smith, C. M. (1989). *A study of standards specifications in English. Master's thesis*. University of Queensland.
Smith, C. M. (1995). *Teachers' reading practices in the secondary school writing classroom: A reappraisal of the nature and function of pre-specified assessment criteria*. Doctoral thesis. University of Queensland.
Sun, J., Przybylski, R., & Johnson, B. J. (2016). A review of research on teachers' use of student data: From the perspective of school leadership. *Educational Assessment, Evaluation and Accountability*, *28*, 5–33. https://doi.org/10.1007/s11092-016-9238-9
Suto, W. I., & Greatorex, J. (2008). What goes through an examiner's mind? Using verbal protocols to gain insights into the GCSE marking process. *British Educational Research Journal*, *34*(2), 213–233. https://doi.org/10.1080/01411920701492050
Taylor, B., & Whittaker, A. (2018). Professional judgement and decision-making in social work. *Journal of Social Work Practice*, *32*(2), 105–109. https://doi.org/10.1080/02650533.2018.1462780
To, J., Panadero, E., & Carless, D. (2022). A systematic review of the educational uses and effects of exemplars. *Assessment and Evaluation in Higher Education*, *47*(8), 1167–1182. https://doi.org/10.1080/02602938.2021.2011134
Valentine, N., Durning, S., Shanahan, E. M., & Schuwirth, L. (2021). Fairness in human judgement in assessment: A hermeneutic literature review and conceptual framework. *Advances in Health Sciences Education*, *26*, 713–738. https://doi.org/10.1007/s10459-020-10002-1
Willis, J., & Adie, L. (2013). Negotiating the meaning of achievement standards in the Australian curriculum. *Curriculum Perspectives*, *33*(1), 52–62.
Wise, S. L., & Smith, L. F. (2016). The validity of assessment when students don't give good effort. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 204–220). Routledge. https://doi.org/10.4324/9781315749136
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., & Bridges, S. (2008). Meeting in the Middle: assessment, pedagogy, learning and educational disadvantage. http://education.qld.gov.au/literacy/docs/deewr-myp-final-report.pdf
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
Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, *17*(1), 59–75. https://doi.org/10.1080/09695940903565610

Wyatt-Smith, C., Lingard, B., & Heck, E. (Eds.). (2021). *Digital disruption in teaching and testing: Assessments, big data, and the transformation of schooling*. Routledge. https://doi.org/10.4324/9781003045793

Wyatt-Smith, C. M., & Gunn, S. (2009). Towards theorising assessment as critical inquiry. In C. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 83–102). Springer. https://doi.org/10.1007/978-1-4020-9964-9

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: 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*, *50*, 1420–1448. https://doi.org/10.1002/berj.3984