How does assessment drive learning? A focus on students' development of evaluative judgement

Juan Fischer, Margaret Bearman, David Boud & Joanna Tai

Published online: 04 May 2023.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

OPEN ACCESS

Check for updates

How does assessment drive learning? A focus on students' development of evaluative judgement

Juan Fischer[a] (ID), Margaret Bearman[a] (ID), David Boud[a,b,c] (ID) and Joanna Tai[a] (ID)

[a]Centre for Research in Assessment and Digital Learning, Deakin University; [b]Faculty of Arts and Social Sciences, University of Technology Sydney, Sydney, Australia; [c]Work and Learning Research Centre, Middlesex University, London, UK

ABSTRACT

Summative assessment is often considered a motivator that drives students' learning. Higher education has a responsibility in promoting lifelong learning and assessment plays an important role in supporting students' capability to make evaluative judgements about their work and that of others. However, as research often focuses on formal pedagogical design, it is unclear what behaviours summative assessment prompts, thus the relationship between summative assessment, learning and evaluative judgement requires further investigation. Drawing on a small-scale ethnography-informed study, this paper adopts a practice theory approach to explore how undergraduate physics students from three year levels make evaluative judgements in the context of summative assessment tasks. The contexts explored through observations and interviews include a graded in-class tutorial, an out-of-class study group for an in-semester assignment, and individual preparation for examinations. The findings suggest that while summative assessment is a crucial aspect of students' learning context, it does not fully shape students' practices. Instead, students engage in incidental conversations about the quality of their work and how to do things in their studies. By focusing on what students *actually do*, this study integrates formal and informal aspects of students' learning, highlighting the tensions between undergraduate practices and intended learning outcomes.

Introduction

Researchers and practitioners in higher education widely accept that summative assessment plays an important role in student learning. It is a longstanding truism that assessment drives learning: if grades prompt students to focus on certain tasks (Rust 2002), summative assessment should align with the intended outcomes of a given curriculum (Biggs 1999). However, while the literature on assessment often frames summative assessment as a motivational device to drive students' behaviours, it is not necessarily clear what these behaviours are or what signals a task sends to the student. From a learning perspective, summative assessment can offer students the opportunity to make judgements about their work, building lifelong learning capabilities and reducing their dependence on teachers or other authority figures (Boud 2000; Tai et al. 2018). In short, more nuanced investigation of the relationship between assessment and learning may be necessary (Joughin 2010).

Learning is not just a matter of what is intended in a pedagogical design, but rather it occurs as a consequence of what students *actually do* (Goodyear, Carvalho, and Yeoman 2021). Learning therefore may bear little resemblance to articulated learning outcomes or even tacit professional skills objectives. A valuable approach to exploring the relationship between summative assessment and learning is to focus on what students do when they engage in different graded tasks, paying close attention to how multiple layers of context frame students' work, avoiding the implicit assumption that the main influence on learning resides in the quality of the task. This kind of inquiry offers a means to apprehend what students de facto learn when they work towards summative tasks, and whether this includes other (un)desirable outcomes besides those expressed in the course guide.

In this paper, we empirically explore the relationship between summative assessment and student learning activities, particularly how students make evaluative judgements in connection with graded tasks. By evaluative judgement we mean 'the capability to make decisions about the quality of work of self and others' (Tai et al. 2018, 471). We propose evaluative judgement as a useful anchor for understanding how students engage with summative assessment because it is not tied to the formal curriculum and represents a key means of learning into the future. We outline literature on assessment, learning and evaluative judgement before providing an in-depth analysis of what a group of undergraduate physics students do in the context of three assessed tasks. These contexts are an in-class tutorial with a graded worksheet, a study group for an in-semester assignment, and individual preparation for final examinations. Our qualitative analysis draws on a quasi-ethnographic study and is informed by practice theories (Nicolini 2012; Schatzki 2012). It pays specific attention to students' decisions about what good work should look like and how to accomplish it. The aim is to conceptualise how assessment influences the development of evaluative judgement and, by focusing on what students do, we reframe summative assessment as an important element of the context for students' learning while highlighting the open-ended nature of student practices and evaluative judgements.

Background: assessment, learning and evaluative judgement

Graded assessment tasks in higher education play a key role in assuring that students have achieved the learning outcomes prescribed by the university, the degree or the unit of study. Summative assessment is commonly described as directing learning (Joughin 2010), mostly in terms of communicating what students need to learn (Boud 2000), mediating their learning approaches (Joughin 2010), and even as a motivational device necessary to prompt students to engage with specific tasks (Rust 2002). Summative assessment may also play a formative role as task design can prompt students to engage in activities that align with the intended outcomes of the course or unit. Notions like constructive alignment emphasise the necessity of aligning different aspects of the curriculum to optimise the chances of students carrying out tasks conducive to the intended learning (Biggs 1999). Under ideal circumstances, it could be argued that 'teaching to the test is exactly what you want because it is teaching the intended curriculum' (Biggs 1998, 108).

However, Joughin (2010) argues that the view of assessment as a motivational tool is based on studies from the 1950s-1970s, and the relationship between assessment and learning needs contemporary insights. Others articulate that the relationship between assessment and learning needs to be understood beyond the immediate aspects of task design. For example, Cilliers et al. (2012) showed how specific task requirements and general curriculum design have effects on student planning and behaviours before engaging in the task itself. Similarly, Harland et al. (2015) showed how, at an institutional level, courses studied in parallel may compete for students' attention, resulting in an unsustainable assessment 'arms race', in which students would not work without grades while feeling generally stressed by constant assessment. Conversely,

Häsä, Rämö, and Nieminen (2021) showed that, in a context where assessment and grades are not high stakes and students receive relevant training and feedback, summative self-assessment can be used for learning and certification purposes, without fear of cheating. Brown (2021) adds a layer of complexity by arguing that all assessment tasks involve learning, since students either need to prepare for a test or learn something from the experience of being tested. Collectively, these studies suggest that summative assessment can have a deeper impact on student learning beyond prompting engagement with specific activities, and that context needs to be understood as more complex than a short series of tasks.

From a conceptual perspective, understanding how assessment influences learning also requires exploring how it may contribute to equipping students with the necessary tools to drive their own lifelong learning. The concept of *sustainable assessment* was introduced in awareness that evaluating students' work needs to not be the monopoly of teachers and institutions, as students need to become the judges of their own work when teachers or other authority figures are not available (Boud 2000).

The notion of sustainable assessment contributed to the subsequent conceptualisation of *evaluative judgement* as defined above. These concepts have helped reframe rubrics and exemplars as artefacts that mediate students' discussion about quality (To, Panadero, and Carless 2022); feedback as a process in which learners use information from diverse sources to improve their work moving away from feedback as the comments teachers make (Henderson et al. 2018); and self and peer assessment as qualitative judgements about how students' work reflects quality standards (Adachi, Tai, and Dawson 2018). However, the relationship between summative assessment, learning and the development of evaluative judgement remains empirically unexplored.

The research mostly focuses on the effectiveness of assessment design in promoting learning, including evaluative judgement. Research on the development of students' evaluative judgement might fruitfully benefit from adopting an alternative view that distinguishes educational design from what students *actually do*, which 'for better or worse… may differ considerably from what their teachers *think* they are doing or what their teachers *intend* them to do' (Goodyear, Carvalho, and Yeoman 2021, 446, emphasis in the original). Some studies have moved in this direction by adopting observational methods to explore how teachers and students use rubrics and exemplars in specific contexts (Hawe et al. 2021), by qualitatively analysing the content of students' judgements in peer-review tasks (Chen et al. 2022) or reflective portfolios (Gladovic, Tai, and Dawson 2022), or by exploring how training doctors develop their evaluative judgement in clinical settings (Bearman et al. 2022). However, they do not consider the effects of summative assessment itself.

Considering the importance of understanding the role that summative assessment plays in learning and the key goal of higher education in supporting lifelong learning, we aim to answer the question 'how do students come to make evaluative judgements in the context of summative assessment?'

Theoretical framework: practice theories

To explore the relationship between evaluative judgement and summative assessment, we identified the general lens of practice theories as a useful theoretical framework for examining students' actions and learning in terms of how these emerge in socially and materially situated contexts. This perspective helps reframe assessment as 'sets of complex relations between people and material things' (Boud et al. 2018, 1115) that need to be understood as they happen in everyday life and not in terms of what good assessment claims to be. In this sense, practice theories allow us to examine what people do *in situ*, without assuming replicability. For this paper, we have adopted two major features of practice theories: 1) the entanglement of

practices and the contexts in which they emerge, and 2) a conception of learning as mastery of socially shared practices.

We talk about a 'general lens' because practice theories are a diverse group of theories that seek to explain how social life is (re)produced through everyday human practices. According to Hager, Lee, and Reich (2012), these theories generally understand practice as a socio-material phenomenon that involves embodied interactions involving humans, objects (books, chairs) and places (classrooms, landscapes). Rather than a directed application of knowledge, ideas and beliefs, practices emerge and evolve in undetermined ways. Practices are historical and contextual instead of being stable or homogeneous routines, and knowledge is viewed as a product of collective practice, not just information that can be transmitted to others. That is, knowledge is considered 'a form of mastery that is expressed in the capacity to carry out a social and material activity' (Nicolini 2012, 5). Thus, knowledge and learning are considered social not just because they often happen in direct interaction with other people but because what is learnt is 'how to go on in practices' that precede the individual and are shared with others (Kemmis et al. 2014).

From this perspective, learning is conceptualised as a process that involves the 'whole person acting in the world' (Lave and Wenger 1991, 49) that results in an *ontological transformation* from participating in social practices (Kemmis 2021). In other words, it involves people learning how to do things and recognise the world around them, learning how to think, talk, relate to one another and, in general, 'come to practise differently' (Kemmis 2021, 282). Kemmis et al. (2014) describe this process as *being stirred into practices*, conveying the happenstance, movement and dynamism of learning as something that occurs by participating in everyday practices, where the learner may be guided by others and themselves, not always deliberately but always mediated by the local context in which practices occur. In this sense, formal teaching and other educational arrangements are reframed as 'the landscape across which students walk' (Goodyear 2015, 34), distinguishing them from the actual learning that results from what students do during and beyond the activities teachers propose. Therefore, practice-based research on learning requires focusing on how students' practices emerge within specific socio-material contexts, including but not limited to assessment design, besides exploring students' accounts of what occurs during their studies.

The value of practice theories has been demonstrated in recent higher education research on issues of teaching and learning. These works have explored issues like how teachers design assessment tasks (Boud et al. 2018), how feedback can be framed as a socio-material practice beyond individual learners (Gravett 2022), and even conceptualisations of evaluative judgement as entangled in disciplinary practices (Bearman 2018). However, these have been largely conceptual or have focused on teacher practices and formal educational arrangements. Hence, we use ethnographic methods to explore student learning and evaluative judgement through direct observation and interaction with students.

Methodology

In our study, we adopted an ethnography-informed approach, as it is widely accepted among practice theorists that practice-based research requires proximal methodologies like ethnography or case studies (Nicolini 2012; Schatzki 2012). This approach focuses on observing how practices unfold in particular times and spaces, how people interact and how participants experience them, resulting in 'thick descriptions' (Hammersley and Atkinson 2007). These descriptions address the complexity of social life through its relationships, processes and histories, and aim for conceptual development rather than generalisations from specific cases. For this paper, we present an in-depth analysis of the practices of a small group of students during and in preparation for graded assessment tasks, to uncover how summative assessment may prompt students' evaluative judgements *in situ*.

Context

This study took place in an undergraduate physics program at a large Australian university, which offers core physics units across three year-levels in physics major degrees, and two year-levels for minor degrees. The units include theoretical and laboratory-based components, with the former offered through lectures and tutorials, which in first year are combined in single sessions, while remaining separate in following years. The summative assessment tasks in the theoretical component are largely a series of written assignments and a final examination covering physics questions and problems, while second and third-year tutorials also include graded worksheets despite having a primarily formative intention. For this paper, we focused on the summative tasks of the theoretical component.

Data collection and participants

The first author collected data during the second semester of 2019 and the first semester of 2020, through participant observations on campus (only during 2019 due to restrictions from Covid-19) and repeated interviews with key student informants. Students were invited to participate early in their semester, resulting in five main participants agreeing to be interviewed and, when possible, observed as they studied. Some of these participants became informants (Glesne and Peshkin 1992), who shared important information about the evolution of the semester and invited other students to participate in observation sessions. Table 1 summarises the students' participation as main or secondary participants by the period of data collection and year level, using pseudonyms to keep their anonymity. All participants received oral and written explanations about the study and signed consent forms.

Observations were captured through fieldnotes, and semi-structured interviews with the main participants were recorded and transcribed for subsequent analysis. Depending on their availability, students were interviewed between one and six times during the semester, with one student participating across two semesters. Interviews covered general aspects of the physics course as well as clarifications of previous observation sessions and descriptions of practices that were not accessible during fieldwork (i.e. off-campus practices or events that occurred without the presence of the researcher). For this paper, we focus on observations of one second-year tutorial session and one first-year study group in preparation for a written assignment, and interviews about how students prepare for examinations throughout the year levels. We chose these activities because they illustrate student practices that emerged during low and high-stakes tasks, while also being restricted to one type of student work (i.e. mathematical problem-solving, as opposed to laboratory work).

Data analysis

Data analysis started as the first author's reflective diary, which was a part of the fieldnotes, and it included discussions with the rest of the team during the longitudinal data collection. Accounts of student practice in response to assessment were constructed, with explicit attention to the socio-material arrangements involved (Kemmis et al. 2014). These were interrogated by

Table 1. List of participants.

Collection period	Year level	Main participants	Secondary participants
2019-2	1st year	Malcolm	Linnea, Carina, James (Study group)
		Tony	—
	2nd year	Steve	Tarja, Ronnie (Tutorial)
		Nita	
2020-1	2nd year	Malcolm (continued)	—
	3rd year	Ian	—

the research team to identify the points where students were making decisions about quality. Recurring contexts and scenarios were noted, and the three most significant for evaluative judgement were selected for this paper.

Findings

In this section, we explore how student practices emerge among socio-material affordances as they relate to prescribed summative activities, with special attention to how they make judgements about their work. These findings derive from observations of collaborative work during a tutorial and towards a written assignment, and students' accounts of their preparation for final examinations.

Working with peers and teachers in a tutorial

Second-year tutorials in the physics course combine formative and summative assessment. Students work collaboratively on problem sets, with the occasional help of tutors, which are submitted for marking at the end of the class. Observing this activity showed students coming together in physical spaces like the classroom with round tables and mobile whiteboards, which allowed them to share and discuss their solutions with the rest of the group, and using material artefacts like textbooks, lecture notes and computers to consult for information. However, as students worked on the questions at hand, their conversations also moved to discussions about how to approach physics problems more generally, and sometimes about how to go about their studies beyond the current task.

> Nita arrived a few minutes late to the tutorial and joined her friends Tarja and Ronnie, who she met in class in the previous semester and with whom she and Steve usually worked. Nita asked them if the problem set related to magnetism, as she was behind in her studies in this area. Later, she noticed that Ronnie tried to solve a problem by starting on one of the multiple-choice options, and she told him off and asked, '*what will you do in the exams when you don't have the answer options?*' As they continued working, Nita mentioned being stressed by other assignments and forgetting to submit an assignment, and then they moved on to talk about the weight of pre-workshop tasks on their final grade. Later, Steve (who joined the group later than Nita) checked the upcoming assignment task on his phone and pointed out that '*the first question is only 3 marks. They basically want you to write anything*', and eventually, the group discussed the grades needed for applying to an honour's degree. At this stage, Steve and Tarja were working on a mobile whiteboard next to their table, and Ronnie and Nita were trying out solutions on their notebooks.

These conversations revolved around what to prioritise in their studies and what constitutes 'good enough' work, decisions that seemed mediated by the marks awarded by each task or question. However, when focusing on the task at hand, their conversation seemed to evolve into how to do things in physics.

> When the tutors approach the table, Steve and Tarja explained their current approach before the teachers offered their views. On one occasion, the tutor offered an alternative approach and Tarja responded, '*I find it not that intuitive. I know where you came from, but I wouldn't have done that*'. This prompted the tutor to explain that it is a complex solution in electromagnetism that someone creatively developed in the past and it is now accepted as an effective approach, while not necessarily an obvious one. This mirrored another conversation heard in a first-year workshop when Tony explained his point of view to a tutor, and the tutor recommended a common form of looking at a problem in physics: '*take a limit in your head: what if this angle gets closer to 0?*' or '*take the approximation when Z is way, way bigger than R, when it tends to infinity*'. Tony described this as the 'physicist mindset', which involves '*how they look at problems and solving problems and how they conceptualise and stuff like that*'.

Overall, tutorials seemed to privilege opportunities for engaging in incidental conversations about quality. But rather than talking about how to do things well in abstract terms, students made evaluative judgements in real time about how to address the problems at hand and how

these approaches may be applied in other situations. Moreover, they also seemed to discuss quality in terms of how to do things well as they navigate their course. Just as tutorials seemed to support students' learning despite resulting in graded work, we show in the next sections that students may engage in similar processes during summative tasks with higher stakes.

Working with peers towards a summative assignment

In a formal sense, assignments focus more on the summative aspects of assessment, as they are submitted individually for marking with less guidance from teachers than tutorial work. However, as students engage in high-stakes assessments, they seem to do more than what the finished assignment shows. During the observed study session, the first-year students shared and discussed solutions, found roadblocks, proposed alternative approaches, double-checked each other's work, retrieved examples from textbooks, lecture slides or other assignments, recommended online resources, and put all of this together trying to work out problems they have not solved before.

> Malcolm and I caught up to talk about the semester and plan future interviews and observations. In this meeting, Malcolm said that his group of friends had agreed to meet up in the student lounge after class to work on the assignment. This was an impromptu study session of a relatively stable group of friends who had met and bonded during previous workshops. At the start of the session, Malcolm explained that they were all going through past assignments' solutions to find similar ways to approach problems. As they shared their individual solutions, Linnea was double-checking Carina's work and yelled, *'this doesn't make sense'.* They discussed the procedure, and Carina replied *'c is just a constant. I can do that with a constant!'* (Constant values can be extracted from integrals, as a lecturer reminded during a previous workshop) Later, Linnea pointed out that if *V* represents a vector, *'you have to say that, because they're gonna read it and ask,* "what is that?!"' At some point, James asked about the solution to a mathematical expression, and Malcolm recommended an online step-by-step calculator, which offered simple calculations but not full solutions to physics problems (for similar purposes, Steve uses a program used in the department for running physics simulations). And like the students in the second-year tutorial, Malcolm pointed out that one of the questions was worth 2 marks but felt much more complex than that.

This episode illustrated that behind the clean assignment that students submit for marking, there is a lot of messy work through which students make different evaluative judgements about their approaches for each problem, how to present their solutions, and what resources are the most effective for their purposes. While assignments represented higher stakes than tutorials, out-of-classroom work seemed more informal, maybe due to the absence of teachers during the activity, allowing students to use a broader range of resources and engage in more heated discussions. However, both activities seemed to offer opportunities for hands-on learning, as Nita (second year) explained:

> It's all good being told these things in lectures, but I think until you look over yourself and do it yourself, you don't really understand what's going on… sometimes just getting another perspective is really helpful or hearing it again in a different way. Sometimes it's by myself, but sometimes, more often than not, it's probably with a group of people.

Thus, when students worked on their assignments, they face new problems that require creative efforts, and in doing so, they need to evaluate their alternatives and how they use different resources to address the task at hand. They may do so individually, but the observation of the study group showed how generative group discussions can be.

Preparing for examinations

End-of-semester examinations are the most typical example of summative assessment as they mark the culmination of a teaching cycle and are usually high-stakes tasks. Like tutorials and in-semester assignments, preparing for examinations can be framed as a particular iteration of

using physics to solve problems. During this preparation, students need to ensure they can solve problems with limited resources and without help from others—even if the revision is done in groups. Students need to anticipate the kind of questions they will face in the examination—'*what is being assessed*' (Tony, first year)—often by consulting previous versions of the same units' examinations and practice questions provided by the lecturers.

The students interviewed explained that they often prepared for examinations individually—although Malcolm talked about a group study session that he could not attend. This preparation involves self-assessing their work by contrasting solutions, preparing notes on the formula sheet they can use during the examination, and eventually making judgements on their practices of revising. Ian (third year) described this approach as follows:

> The only way I've ever studied for an exam is through practice questions… I find that my understanding is improved more when I do practice questions than going over lectures… they can't really throw a curveball at you, they tend to be very similar to the practice questions, so you'll know if you nailed that question or not just because you've done them five times.

Tony described a similar approach, but he added that:

> If I was weak in that area, I would have to supply an 80-20 principle to the situation where you want to study the things that are going to get you the most results for the least amount of time because when your exams come around you have four exams to prepare for in a short period.

Tony's description shows how students need to focus both on learning the content that will be assessed as much as on being able to navigate the examination period and the competing commitments that arise. Sitting and preparing for examinations is something that students need to learn across their degree, as exemplified by Malcolm at the end of his first year: '*In terms of how I prepared this semester, I think I did better. I don't know what I was expecting from first semester, but now I kind of got the idea*'. Nonetheless, when the examination conditions changed in his second year due to the Covid-19 pandemic (i.e. remote, open-book examinations with extra time), Malcolm faced great difficulties: '*it was extremely long and extremely hard. And a lot harder I think than they thought it would be for us*'. In other words, students make evaluative judgements throughout their course about what studying well looks like.

During the examination itself, students cannot talk with other people or consult resources to compare their answers, hence they are less certain about their performance than during their preparation. Student self-assessment during the examination is more complex, as the students interviewed seem to use completion and their general feeling of confidence as indicators of quality of their work, while the final marking seems to be the confirmation they require. For example, after the examination at the end of his first year, Malcolm explained:

> Look, I didn't finish everything. There were some big questions that I wasn't sure where to start, and by the time… the time finished, I hadn't really got an answer for those. But I was a lot better prepared, I could… I've done all the practice exams, so there were some questions that where you just look at them and go 'I've done this three times today' (Malcolm, in first year).

Similarly, Steve after a second-year examination said, '*I think I went very well on the exam. I actually finished all the questions and actually knew what was going for each of the questions*'. Ian (third year) concurred that '*you can always gauge in an exam how well you've done because you recognised the question, and you know how to answer it*'.

These accounts indicate that preparation for examinations involved more than refreshing previously learned content, as the act of revising can be an opportunity for students to gain additional perspectives on the content. Moreover, in terms of evaluative judgement, students seemed to gain expertise in what it is to sit examinations and what preparation it requires, which is an embodied learning of how to work out certain types of problems under the specific conditions of final examinations.

Discussion

We explored the relationship between summative assessment and evaluative judgement by analysing the practices of a group of physics students as they worked towards graded tasks. Our findings showed that summative assessment tasks play a significant role in prompting students on how to engage in specific learning activities (e.g. how much effort to put into a task or question, anticipating relevant questions in preparation for an examination). However, our findings also showed that students do more than what the task at hand prescribes. Whilst these physics students worked in tutorials, towards assignments and in preparation for examinations, they came to make real-time evaluative judgements about the quality of their solutions, their understanding of physics and mathematics, the use of certain tools and resources to enhance their work, and how to go about their studies. Although our findings do not show students' evaluative judgement developing or how they 'came to practice differently', we showed that, across the curriculum, summative assessment tasks require students to make decisions about quality as they navigate their course, even if these decisions are made idiosyncratically. Thus, it is important to consider the unintended role that summative assessments play in evaluative judgement and whether they align to or are in conflict with intended learning outcomes.

During summative activities, social interactions with peers and tutors allowed generative discussions about the correctness of mathematical moves, ways of looking at physics problems, the amount of effort a question requires, and the resources that may help them to address physics and study problems (e.g. step-by-step calculators, books, simulation software). These interactions have a history and the observations we presented involved groups of friends (though Tony did not have a study group and simply joined any table during workshops). Social interactions were enabled by specific material arrangements like the student lounge, round tables and whiteboards, which may be more supportive of collaborative work than lecture theatres. In contrast to these generative discussions, individual preparation for examinations seems to focus on becoming more experienced in dealing with this type of high-stakes task. In this sense, summative assessment seems to be a constitutive element of the context in which students' practices emerge as it invites students to carry out certain tasks, but it does not determine what students do nor does it fully anticipate further affordances and barriers for students' practices. Students facing the same assignment may proceed differently as they encounter different people and artefacts at different times and places.

These findings highlight that students' practices are not abstract procedures, but concrete and purposeful forms of doing and saying under specific circumstances. They illustrate how assessment tasks provide a map for student work, but they do not determine how students will go about those activities (Goodyear 2015; Goodyear, Carvalho, and Yeoman 2021). As Boud et al. (2018) explain, assessment is a 'socially situated phenomenon' (p. 1108) and not a 'unilateral act of teachers or course designers to which students are subjected' (p. 1109). By attending to students' practices in real time, we showed how they idiosyncratically evaluated their study practices as much as their finished products (Fawns and O'Shea 2019), even if the task at hand was not deliberately designed to support their evaluative judgement (Gladovic, Tai, and Dawson 2022). As students engaged in diverse activities, including summative assessment tasks, they encountered unexpected challenges for which they needed to test and discuss different approaches, in a combination of recalling previous knowledge and creative decisions on how to solve a given problem, which in turn may result in the continuous refinement of their evaluative judgement as 'a "feel" for what constitutes quality in the field' (Dall'Alba, 2018, 20)—in this case, the physics curriculum.

Implications

One challenge that this paper raises is that, while educators may seek to support students' evaluative judgement for life post-graduation, students' decisions occur within the context of

the practices of their undergraduate degree. This means that many of their judgements may focus on how things are done in their degree, whether or not the curriculum is designed to be an effective introduction to the actual practices of the discipline. We can argue that students are learning how things are done in their discipline as a secondary effect of *being stirred into* their current study practices. This is not a matter of deep/surface approaches to learning but about students coming to know how to navigate the social and material context of their every-day life, which happens to be a highly structured environment where summative tasks and grades play an important role in teachers' and students' everyday lives (Harland et al. 2015). In other words, for curriculum design, it is important to consider *what* students are learning in pursuit of completing summative assessment tasks, not only *if* they are. For this purpose, we identify two approaches that may contribute to moderating the tensions between learning within the curriculum and disciplinary practices: authentic assessment at a program level and partnering with students for curriculum design.

Authentic assessment is a valuable approach that may help to gradually bring students closer to practices relating to the real and cognitively challenging problems in their field, while making holistic and explicit evaluative judgements about their own work (Villarroel et al. 2018). Careful design of authentic assessment needs to pay attention to the material and social aspects of the learning context as much as the content and requirements of the task (Gulikers, Bastiaens, and Kirschner 2004). For example, assessment could aim to recreate how physicists encounter and formulate problems in their work, rather than relying on self-contained questions typical of traditional assignments. Long-term projects that span one or multiple semesters may also require students to work collaboratively, gather and organise information, and use current specialised technology to formulate complex solutions to ill-defined problems, while documenting the decision-making process and self-evaluating the merits and limitations of the project.

It is worth considering the realism of authentic assessment in terms of its social value, not only in terms of future employment and economics (McArthur 2023). Doing so might promote forms of evaluative judgement that incorporate wider societal issues when making decisions about 'what good work looks like'. For instance, scientific projects could be evaluated not only on their use of disciplinary knowledge and techniques but how well they respond to the needs of a specific community and their potential side-effects.

Recognising and understanding student practices that emerge out of learning and assessment activities could be an onerous undertaking. However, partnering with students to explore ped-agogical approaches could be a valuable strategy as students can provide relevant input for educational designs that better respond to their actual practices, including how social relation-ships are formed, how students use material artefacts and how they make evaluative judgements. Partnerships could be as simple as peer mentoring programs, or as complex as co-producing learning resources or co-designing sequences of tasks considering the material settings that may be more productive for student learning. As Goodyear, Carvalho, and Yeoman (2021) explain, evaluation of educational designs 'sometimes means checking whether an intervention worked' but in a practice-focused perspective, it 'always also involves answering questions about *how* an intervention worked' (p. 460, emphasis in the original), which requires understanding what students *actually did* and what enabled/constrained their actions.

Limitations

One limitation of this paper is that being a small-scale ethnography-informed study, the evidence we presented cannot be generalised as something most students do. The students that partic-ipated in this study cannot be claimed to be representative of the whole student cohort. They cannot encapsulate all students' practises/actions within the course, and further research may offer additional perspectives. However, our in-depth study aimed to explore what some students

did under certain circumstances and provide conceptual insights about the relationship between assessment, learning and evaluative judgement. We recognise that learning what quality looks like and how to navigate the higher education system is a complex longitudinal process and not an automatic effect of simply being there and working on problems. Future research can address this limitation through larger-scale studies or through intentionally recruiting students from specific equity groups.

In this paper, we followed the general view on learning from practice theories; however, conceptualising evaluative judgement may benefit from a focused exploration of the interplay between student practices and the social and material conditions that enable or limit them, for which some practice theories may be better equipped than others. This focus on the situated nature of context can be particularly useful for uncovering how discursive, material, economic and power issues mediate students' evaluative judgement.

Conclusion

We used an ethnographic approach to explore how students come to make evaluative judgements in the context of summative assessment tasks. We showed that summative assessment plays an important role as part of the context in which students' practices emerge, but task design does not fully anticipate what students do, as they tend to engage in incidental conversations about the quality of their work and about how to do things in their studies. This exploration also uncovered how students make decisions about quality in diverse contexts beyond the formative tasks oriented to support their evaluative judgement. The insights from this study contribute to bringing together the formal and informal aspects of students' learning, highlighting the open-ended nature of students' practices and how their decisions about quality often target how to go about undergraduate study practices. This study contributes to locating learning from summative assessment and evaluative judgement in what students *actually do* and highlights the tensions between undergraduate practices and intended learning outcomes.

Disclosure statement

No potential conflict of interest was reported by the authors.

Ethical approval

This research, identified by project number HAE-19-121, received ethics approval from the Human Ethics Advisory Group at Deakin University's Faculty of Arts and Education.

Notes on Contributors

Juan Fischer is a PhD graduate from the Centre for Research Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University. His research interests include evaluative judgement, assessment, curriculum development and practice theories.

Margaret Bearman is a Professor (Research) within the Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University. Her research interests include assessment design, work-based feedback, simulation and digital technologies, sociomateriality, and educational workforce development.

David Boud is Alfred Deakin Professor and Foundation Director of the Centre for Research in Assessment and Digital Learning at Deakin University, Australia. He is also Professor of Work and Learning at Middlesex University, London and Emeritus Professor in the Faculty of Arts and Social Sciences at the University of Technology Sydney. His current work focuses on the areas of assessment for learning in higher education, academic formation and workplace learning.

Joanna Tai is Senior Research Fellow at the Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University. Her interests include student perspectives on learning and assessment from university to the workplace, peer-assisted learning, feedback and assessment literacy, developing capacity for evaluative judgement, and research synthesis.

ORCID

Juan Fischer 🆔 http://orcid.org/0000-0002-8938-3575
Margaret Bearman 🆔 http://orcid.org/0000-0002-6862-9871
David Boud 🆔 http://orcid.org/0000-0002-6883-2722
Joanna Tai 🆔 http://orcid.org/0000-0002-8984-2671

References

Adachi, C., J. H.-M. Tai, and P. Dawson. 2018. "Academics' Perceptions of the Benefits and Challenges of Self and Peer Assessment in Higher Education." *Assessment & Evaluation in Higher Education* 43 (2): 294–306. doi:10.1080/02602938.2017.1339775.

Bearman, M. 2018. "Prefiguration, Identities and Agency: The Disciplinary Nature of Evaluative Judgement." In *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*, edited by D. Boud, R. Ajjawi, P. Dawson, and J. Tai, 147–155. Abingdon, Oxon: Routledge.

Bearman, M., M. Dracup, B. Garth, C. Johnson, and E. Wearne. 2022. "Learning to Recognise What Good Practice Looks like: How General Practice Trainees Develop Evaluative Judgement." *Advances in Health Sciences Education* 27 (1): 215–228. doi:10.1007/s10459-021-10086-3.

Biggs, J. 1998. "Assessment and Classroom Learning: A Role for Summative Assessment?" *Assessment in Education: Principles, Policy & Practice* 5 (1): 103–110. doi:10.1080/0969595980050106.

Biggs, J. 1999. "What the Student Does: Teaching for Enhanced Learning." *Higher Education Research & Development* 18 (1): 57–75. doi:10.1080/0729436990180105.

Boud, D. 2000. "Sustainable Assessment: Rethinking Assessment for the Learning Society." *Studies in Continuing Education* 22 (2): 151–167. doi:10.1080/713695728.

Boud, D., P. Dawson, M. Bearman, S. Bennett, G. Joughin, and E. Molloy. 2018. "Reframing Assessment Research: Through a Practice Perspective." *Studies in Higher Education* 43 (7): 1107–1118. doi:10.1080/03075079.2016.1202913.

Brown, G. T. 2021. "Assessments Cause and Contribute to Learning: If Only we Let Them." In *Assessment as Learning: Maximising Opportunities for Student Learning and Achievement*, edited by Z. Yan and L. Yang, 38–52. London: Routledge.

Chen, L., S. Howitt, D. Higgins, and S. Murray. 2022. "Students' Use of Evaluative Judgement in an Online Peer Learning Community." *Assessment & Evaluation in Higher Education* 47 (4): 493–506. doi:10.1080/02602938.2021.1933378.

Cilliers, F. J., L. W. T. Schuwirth, N. Herman, H. J. Adendorff, and C. P. M. van der Vleuten. 2012. "A Model of the Pre-Assessment Learning Effects of Summative Assessment in Medical Education." *Advances in Health Sciences Education : theory and Practice* 17 (1): 39–53. doi:10.1007/s10459-011-9292-5.

Dall'Alba, G. 2018. "Evaluative Judgement for Learning to Be in a Digital World." In *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*, edited by D. Boud, R. Ajjawi, P. Dawson, and J. Tai, 18–27. Abingdon: Routledge.

Fawns, T., and C. O'Shea. 2019. "Evaluative Judgement of Working Practices: Reconfiguring Assessment to Support Student Adaptability and Agency across Complex Settings." *Italian Journal of Educational Technology*, 27 (1): 5–18. doi:10.17471/2499-4324/1027.

Gladovic, C., J. H.-M. Tai, and P. Dawson. 2022. "Qualitative Approaches to Researching Evaluative Judgement in Pedagogical Activities: A Case Study." *Assessment & Evaluation in Higher Education* 47 (2): 231–244. doi:10.1080/02602938.2021.1901854.

Glesne, C., and A. Peshkin. 1992. "Being There: Developing Understanding through Participant Observation." In *Becoming Qualitative Researchers: An Introduction*, 39–62. White Plains, N.Y.: Longman.

Goodyear, P. 2015. "Teaching as Design." *HERDSA Review of Higher Education*, 2 (2): 27–50.

Goodyear, P., L. Carvalho, and P. Yeoman. 2021. "Activity-Centred Analysis and Design (ACAD): Core Purposes, Distinctive Qualities and Current Developments." *Educational Technology Research and Development : ETR & D* 69 (2): 445–464. doi:10.1007/s11423-020-09926-7.

Gravett, K. 2022. "Feedback Literacies as Sociomaterial Practice." *Critical Studies in Education* 63 (2): 261–274. doi:10.1080/17508487.2020.1747099.

Gulikers, J., T. Bastiaens, and P. Kirschner. 2004. "A Five-Dimensional Framework for Authentic Assessment." *Educational Technology Research and Development* 52 (3): 67–86. doi:10.1007/BF02504676.

Hager, P., A. Lee, and A. Reich. 2012. "Problematising Practice, Reconceptualising Learning and Imagining Change." In P. Hager, A. Lee, & A. Reich (Eds.), *Practice, Learning and Change: Practice-Theory Perspectives on Professional Learning*. Dordrecht, Netherlands: Springer.

Hammersley, M., and P. Atkinson. 2007. *Ethnography: Principles in Practice* (3rd ed.). London: Routledge.

Harland, T., A. McLean, R. Wass, E. Miller, and K. N. Sim. 2015. "An Assessment Arms Race and Its Fallout: Highstakes Grading and the Case for Slow Scholarship." *Assessment & Evaluation in Higher Education* 40 (4): 528–541. doi:10.1080/02602938.2014.931927.

Häsä, J., J. Rämö, and J. H. Nieminen. 2021. "Supporting Quality of Learning by Letting Students Give Their Own Grades: An Innovative Self-Assessment Model in University Mathematics." In *Assessment as Learning: Maximising Opportunities for Student Learning and Achievement*, edited by Z. Yan and L. Yang, 232–243. London: Routledge.

Hawe, E., H. Dixon, J. Murray, and S. Chandler. 2021. "Using Rubrics and Exemplars to Develop Students' Evaluative and Productive Knowledge and Skill." *Journal of Further and Higher Education* 45 (8): 1033–1047. doi:10.1080/0309877X.2020.1851358.

Henderson, M., D. Boud, E. Molloy, P. Dawson, M. Phillips, T. Ryan, and P. Mahoney. 2018. *Feedback for Learning: Closing the Assessment loop - Final Report*. Canberra: Australian Government Department of Education and Training

Joughin, G. 2010. "The Hidden Curriculum Revisited: A Critical Review of Research into the Influence of Summative Assessment on Learning." *Assessment & Evaluation in Higher Education*, 35 (3): 335–345. doi:10.1080/02602930903221493.

Kemmis, S. 2021. "A Practice Theory Perspective on Learning: Beyond a 'Standard' View." *Studies in Continuing Education* 43 (3): 280–295. doi:10.1080/0158037X.2021.1920384.

Kemmis, S., J. Wilkinson, C. Edwards-Groves, I. Hardy, P. Grootenboer, and L. Bristol. 2014. *Changing Practices, Changing Education*. Singapore: Springer.

Lave, J., and E. Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.

McArthur, J. 2023. "Rethinking Authentic Assessment: Work, Well-Being, and Society." *Higher Education* 85 (1): 85–101. doi:10.1007/s10734-022-00822-y.

Nicolini, D. 2012. *Practice Theory, Work & Organization: An Introduction*. Oxford: Oxford University Press.

Rust, C. 2002. "The Impact of Assessment on Student Learning:How Can the Research Literature Practically Help to Inform the Development of Departmental Assessment Strategies and Learner-Centred Assessment Practices?" *Active Learning in Higher Education* 3 (2): 145–158. doi:10.1177/1469787402003002004.

Schatzki, T. 2012. "A Primer on Practices: Theory and Research." In J. Higgs, R. Barnett, S. Billet, M. Hutchings, & F. Trede (Eds.), *Practice-Based Education: Perspectives and Strategies*. Rotterdam, Netherlands: Sense Publishers.

Tai, J., R. Ajjawi, D. Boud, P. Dawson, and E. Panadero. 2018. "Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work." *Higher Education* 76 (3): 467–481. doi:10.1007/s10734-017-0220-3.

To, J., E. Panadero, and D. Carless. 2022. "A Systematic Review of the Educational Uses and Effects of Exemplars." *Assessment & Evaluation in Higher Education* 47 (8): 1167–1182. doi:10.1080/02602938.2021.2011134.

Villarroel, Verónica, Susan Bloxham, Daniela Bruna, Carola Bruna, and Constanza Herrera-Seda. 2018. "Authentic Assessment: Creating a Blueprint for Course Design." *Assessment & Evaluation in Higher Education* 43 (5): 840–854. doi:10.1080/02602938.2017.1412396.