

Student self-assessment: a meta-review of five decades of research

Juuso Henrik Nieminen & David Boud

To cite this article: Juuso Henrik Nieminen & David Boud (2025) Student self-assessment: a meta-review of five decades of research, *Assessment in Education: Principles, Policy & Practice*, 32:2, 127-151, DOI: [10.1080/0969594X.2025.2510211](https://doi.org/10.1080/0969594X.2025.2510211)

To link to this article: <https://doi.org/10.1080/0969594X.2025.2510211>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 27 May 2025.



Submit your article to this journal [↗](#)



Article views: 1472



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)



Student self-assessment: a meta-review of five decades of research

Juuso Henrik Nieminen ^a and David Boud ^{a,b,c}

^aCentre for Research in Assessment and Digital Learning, Deakin University, Melbourne, Australia; ^bFaculty of Arts and Social Sciences, University of Technology Sydney, Sydney, Australia; ^cWork and Learning Research Centre, Middlesex University, London, UK

ABSTRACT

Student self-assessment has received considerable research interest. While several review articles have mapped self-assessment studies, there remains a need to synthesise the insights from field and the differing views and approaches within. In this critical meta-review, we conduct a discourse analysis of 28 earlier review studies concerning self-assessment, spanning five decades of research. We ask: *What is known about student self-assessment? How have review studies produced knowledge about self-assessment?* Analysing our data corpus, we deconstruct four discourses: psychological discourse (self-assessment as a clinical intervention); educational discourse (self-assessment as a contextual, relational practice); performance discourse (self-assessment as a performative practice); and societal discourse (self-assessment as a societal practice). These four discourses represent different forms of knowledge production. Through these discourses, self-assessment research has continued the history of promoting a reflective, rational and autonomous human subject in educational settings, but ironically, often through the occlusion of the self.

ARTICLE HISTORY

Received 28 November 2023
Accepted 17 May 2025

KEYWORDS

Self-assessment; meta-review; critical review; discourse analysis; student positioning

Introduction

What is meant by student self-assessment, why should students take part in it, and how could it best be designed and implemented? A glance at contemporary literature reveals that there is a vastly held consensus that self-assessment involves ‘a wide variety of mechanisms and techniques through which students describe . . . and possibly assign merit or worth to . . . the qualities of their own learning processes and products’ (Panadero et al., 2016, p. 804; see also Boud, 2013; Brown & Harris, 2013). The ‘selves’ of self-assessment, students, are seen as central to this practice as in self-assessment, students are asked to produce judgements about their own work. Moreover, self-assessment is largely portrayed as something desirable due to its potential benefits for student learning and metacognition. It is promoted worldwide through national and international research agendas, assessment policies and curriculum statements. Self-assessment

CONTACT Juuso Henrik Nieminen  juuso.nieminen@deakin.edu.au  Centre for Research in Assessment and Digital Learning, Deakin University, 727 Collins Street, Melbourne, Victoria 3008, Australia

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

is commonly associated with formative assessment and Assessment for Learning initiatives that have moved the focal point of assessment from measuring learning outcomes to supporting student learning (Andrade, 2019; Falchikov & Boud, 1989; Harris & Brown, 2018).

Yet there are also fundamental tensions in how research has conceptualised self-assessment. For example, conceptualising self-assessment as a design practice differs significantly from seeing it as a competence (Eva & Regehr, 2008). Research has simultaneously portrayed self-assessment as a clinical intervention, an ethical practice and a critical praxis, only to name a few examples. Overall, the role that self-assessment plays within the ecosystems of assessment is somewhat contested in assessment research. Does self-assessment produce legitimate knowledge about student learning – can it validly be considered ‘assessment’? Whereas many scholars and research approaches recognise the unique information that self-assessment reveals about those aspects of learning that cannot be rendered visible by other means, there are opposing views. For example, Brown and Harris (2014) stated that ‘self-assessment is not a robust assessment practice’ and that ‘its real place in schooling is as a teachable and learnable component of self-regulated learning’ (p. 23).

These are some key examples of the natural ruptures and tractions in self-assessment research: Given the interdisciplinary nature of this research area, it is hardly surprising that there are differing views. However, making the onto-epistemological alignments, ruptures and overlaps visible in self-assessment research would help the research community in making sense of such differing views. A ‘roadmap’ of different logics of thinking and talking about self-assessment could increase synergy and dialogue between various research agendas. Such logics can be characterised as discourses that, according to poststructural thinking and the work of Michel Foucault in particular, refer to practices that systematically form the objects of which they speak (Foucault, 1972, p. 54). Whereas several literature reviews have successfully synthesised empirical evidence and conceptualisations of self-assessment (e.g. Andrade, 2019; Panadero et al., 2016), we argue there is a need to synthesise the different forms of knowledge production that stem from varying underlying assumptions of what counts as legitimate knowledge about self-assessment.

Scholarly debates do not occur in vacuum but are situated in a historical context where self-assessment is seen to have a largely secondary role in global assessment systems: it remains ‘a nice idea and second runner in assessment’ in most educational contexts (Panadero et al., 2016, p. 824). While the benefits of self-assessment are well known, in practice, self-assessment is seen as an illegitimate form of assessment in comparison to more ‘objective’ forms of assessment, such as tests. As Richardson (2022, p. 22) phrases it, ‘the test is still the king’ – ‘an international gold standard’ – in assessment knowledge production in educational policy and practice amid the global testing cultures (Smith, 2016). Of course, similar issues underlie the implementation of other forms of formative assessment practices, too (Birenbaum et al., 2015; Volante et al., 2025; Wylie & Lyon, 2015). However, the largely secondary role of self-assessment in assessment policy and practice might signal to educational stakeholders that students themselves cannot legitimately produce assessment information despite being seemingly centred in ‘self-assessment’.

In this study, we conduct a meta-review of earlier self-assessment reviews to understand the discourses of self-assessment within such a socio-historical context. As the societal discourses around assessment are in flux, how does self-assessment research position itself, its research object, and students amid these shifting discourses? In the words of Panadero et al. (2017), self-assessment research is currently ‘going through an important phase, as shown by recent publications reviewing the accumulated empirical evidence and proposing a new agenda for self-assessment’ (Panadero et al., 2017, p. 75). We argue that the time is ripe for a synthesis of the field as a whole. Such an approach would reach beyond summarising earlier empirical findings and help us understand the paradigms of self-assessment research; not just *what* is known about self-assessment but *how* it is known. To analyse the knowledge production processes of self-assessment research, we rely on discourse analysis that ‘opens the possibility of calling into question the very structuring theses of discourse orienting research practice’ (Pais & Valero, 2012, p. 20).

Research objective

In this study, we conduct a meta-synthesis of earlier self-assessment reviews. Rather than summarising the themes or findings of earlier studies, as many reviews before us have done successfully, we ask: What kinds of discourses has research drawn on while producing knowledge about self-assessment? We focus on earlier review studies given that reviews are considered a powerful form of expert knowledge production: they are highly respected and cited and have considerable power over educational policy and practices. How reviews have conceptualised and studied student self-assessment indirectly influences public discourses about how self-assessment is understood, designed and implemented. This reminds us of the ethics of knowledge production in self-assessment research and underscores the importance of understanding different paradigms and approaches in such knowledge production processes. By conducting a discourse analysis, we not only synthesise evidence or methods of self-assessment research but ‘redefine the coordinates we use to make sense of the problems of the field’ (Pais & Valero, 2012, p. 23).

We do not focus on any particular educational setting but synthesise the knowledge production processes at all levels of education. This way, we can locate where such processes have taken place in the literature, such as whether certain research areas (e.g. higher education) may dominate knowledge production concerning self-assessment.

Our research questions are:

- Through what kinds of discourses has self-assessment been conceptualised in earlier reviews? (*How* is self-assessment known?)
- What are the empirical findings as framed through these discourses? (*What* is known about self-assessment?)
- How are students positioned within these discourses?

A brief history of self-assessment research

Before embarking on our research endeavour, we provide a brief history of self-assessment research. This section provides context for our study by uncovering how the discourses of self-assessment have taken shape in scholarly work. Overall, self-assessment, ‘knowing thyself’ in a reflective manner, is a highly valued virtue in both Western and Eastern philosophical traditions (Yan, 2022). Self-assessment research carries on these traditions and has bent them in the context of education.

The first reported study on whether students could estimate the marks by teachers was over 90 years ago (Sumner, 1932). A focus on comparing student and teacher marks has continued until today. In the 1970s, interest focused on whether student marks could replace teacher marks. This was contingent on students being able to make accurate estimates of the appropriate marks for their work. This move foundered for two reasons. First, though student marks correlated modestly with teacher marks, they were not considered sufficiently accurate. Second, but more importantly, using marks for summative purposes incentivised students to not accurately report their appraisals even if they could make good judgements. Later, Boud (1989) argued that there were conditions under which student marks could be used. Little work that explicitly focused on summative self-assessment was subsequently published. However, the focus on the equivalence of marking remained strong, and the summative dimension of self-assessment appeared in discussions occasionally.

The next phase can be characterised as seeing self-assessment as necessary for students to develop. If students were to be active agents in their own learning and if graduates were to be equipped to operate effectively in society, they must learn to develop the capability of assessing their work (Heron, 1981). Studies from this perspective often focused on comparing student and teacher marks, but student marks were no longer part of a formal grading regime. It was still important that students should be accurate self-assessors; otherwise, they would make poor decisions about their subsequent learning and practice.

Accuracy has remained a key topic in self-assessment research ever since (Andrade, 2019). Reviews in the 1980s identified medium to good correlations that students improved their performance in self-assessing as they progressed and that higher-performing students were better self-assessors than poorer-performing students (Boud & Falchikov, 1989; Boud, 1989). Many studies have replicated this finding since, not least in the much-cited paper identifying the Dunning-Kruger effect (Kruger & Dunning, 1999). However, scholars in some fields, particularly medical education, did not regard correlations between student and teacher marks as sufficiently high and questioned whether individuals are inherently incapable of judging the strength of their own abilities, a direction that leads to self-assessment as a practice to be dropped (Eva et al., 2012). This was not a view taken up much beyond medicine.

Currently, self-assessment research has developed in two overall directions. The first accepts that fostering students’ self-assessment capabilities is necessary but argues that labelling this practice as self-assessment locates it in the field of measurement, not pedagogy. It has sought to shift the discourse and proposed the notion of evaluative judgement to frame a pedagogical purpose (Tai et al., 2018). This move seeks to position self-assessment as a normal part of course design with interventions designed to enhance evaluative judgement fully integrated into other teaching and

learning practices. The second direction takes self-assessment into the realm of educational psychology. It sees self-assessment as an element of models to develop students' metacognitive skills (Brown & Harris, 2014). It is less pragmatic in focus and typically uses less naturalistic research designs. Yet, there seems to be a consensus between these two directions that students should have an active role in self-assessment. Ideas such as 'lifelong learning' and 'self-regulation' are typical in both contemporary research traditions, following a wider turn in international assessment research to promote these ideals (Van der Kleij et al., 2019), albeit in different ways as our review will demonstrate.

Methods

Researching research through a meta-review

A meta-review is a method for reviewing earlier literature reviews (Gough et al., 2012). Review studies are influential in informing educational practices and policies (Davies, 2000). Review studies are crucial for the reflexivity of research fields as they summarise, synthesise and criticise research areas. In doing so, they have particular power in shaping scholarly discourses. Reviews may strengthen certain voices, methods and approaches and silence others; they may shape the idea of 'normal' and 'usual' research approaches and influence the overall paradigms surrounding individual research outputs and projects.

We draw on systematic techniques in our literature search. This ensures that our analysis covers the most relevant references and remains truthful to its cause. However, we emphasise that our study does not rely on typical systematic review methods when synthesising our data material. Rather than looking for themes or patterns, we take the approach of 'researching research' (Pais & Valero, 2012). This means that we analyse the knowledge production mechanisms in scholarly research outputs through the discourses they take part in. This way, we can discuss the functionalities and potential consequences of such discourses. As Pais and Valero (2012) phrased it, '[r]esearch produces languages and tools, which shape what we see and say about the very same world of (...) education' (p. 11). While many of the reviews analysed have done an excellent job of 'providing an overview of what the field knows about its research objects', our meta-review focuses on 'the effects of research in generating particular discourses, with its encompassing ideologies' (Pais & Valero, 2012, p. 10). As such, our study is not 'yet another review of self-assessment' but an attempt to understand self-assessment research and its effects in educational contexts through the expert discourses they produce.

Literature search

For our discourse analysis, we sought to identify a 'corpus of statements' (Arribas-Ayllon & Walkerdine, 2008, p. 115) that allows us to make substantial truth claims about self-assessment research. This is why we chose to undertake traditional literature search methods.

The review process was conducted in Covidence systematic review software (2023), a web-based software for managing reviews. The literature search consisted of three

phases: two systematic search processes and a further snowballing process of reference lists. The first phase started with a systematic search in three core education databases: Scopus, Education Resources Information Center (ERIC) and Web of Science. This search was conducted in June 2022. The search terms concerning self-assessment were constituted by the authors, whereas the search terms concerning review studies drew on Lum et al. (2017). The search string was:

(Self-assess* OR self-evaluat* OR self-grading OR evaluative judg* OR dunning OR kruger) AND (review OR meta-analy* OR 'literature review' OR 'systematic review' OR 'research synthesis' OR 'narrative review' OR 'scoping review' OR 'realist review' OR 'intrinsic review' OR 'critical review')

This search was restricted for title, abstract and keywords in Scopus and Web of Science. The search was restricted to SSCI-indexed journals in Web Of Science. The first author screened the titles and abstracts of 3590 studies against the inclusion and exclusion criteria (Table 1). A total of 119 studies were included for a full-text screen. The first author screened the full articles against the inclusion and exclusion criteria. All unclear cases were discussed and resolved mutually. Ninety-seven studies did not fit our inclusion criteria, mostly because the studies were not conducted in educational settings (e.g. workplace settings; $n = 31$), because they did not focus on self-assessment (e.g. general formative assessment reviews; $n = 27$) and because they were not literature reviews (e.g. conceptual studies; $n = 21$). We included 19 studies.

The second search was conducted in May 2023 since we anticipated we had not identified all relevant sources. The same search protocol was used for potential new sources. We used the same search string in the three databases as in the first phase but updated the review to include studies from 2022 to 2023. This yielded 123 studies, of which 14 were included in full-text review. Finally, two new studies were included in data extraction. Next, we conducted the same search in two new databases (PsycINFO and

Table 1. Inclusion and exclusion criteria.

| | Inclusion criteria | Exclusion criteria |
|----------------|--|--|
| Phenomenon | <ul style="list-style-type: none"> ● Focus on self-assessment in educational contexts. ● Any level of education. ● Focus on assessment related to learning (Assessment of, for and as Learning) | <ul style="list-style-type: none"> ● Focus on student competencies rather than on assessment practices. ● Self-assessment reviews in contexts beyond education (e.g. workplace). ● Studies that included self-assessment as one reviewed practice but that did not provide new information about self-assessment in particular (excluding, e.g. reviews on formative assessment). |
| Type of record | <ul style="list-style-type: none"> ● Peer-reviewed studies. ● Journal articles. ● Literature reviews and syntheses that reported the search process and the analytical process. ● Any year. ● Studies published in English. | <ul style="list-style-type: none"> ● Non-peer-reviewed studies. ● Book chapters, conference proceedings, reports, dissertations. ● Studies that did not focus on a literature synthesis. ● Conceptual overview studies with no review methods. ● Studies in languages other than English. |
| Population | <ul style="list-style-type: none"> ● Student self-assessment at any educational level. | <ul style="list-style-type: none"> ● Self-assessment of other stakeholders (e.g. teachers). |

PubMed) to identify further relevant sources from psychological and medical fields. This search yielded 309 hits. After an abstract screening, 24 studies were included in a full-text review. We included three new studies in data extraction.

In the third phase, we screened the reference lists of the studies included in the full-text review in the search protocol's previous phases. Studies whose titles seemed relevant were sought, and their abstracts were read against the inclusion and exclusion criteria. The full texts of 30 new studies were read against the inclusion and exclusion criteria. This resulted in three newly identified studies. Finally, we added a newly published study by Yan, Panadero et al. (2023) in data extraction in August 2023.

Our final set of articles included 28 studies (19 from the first phase, five from the second one, and four from the third one).

Data extraction and analysis

The following information was extracted from each study:

- Author and study details such as year of publication, discipline and level of education
- Methodological aspects such as the type of review, literature search techniques and data analysis methods
- Both explicit and implicit conceptualisations of student self-assessment
- Stated research gap and contribution
- Findings of the study
- Implications of the study for research and practice

To analyse the discourses of self-assessment, we deployed a Foucault-inspired discourse analysis (Arribas-Ayllon & Walkerdine, 2008). The analysis follows the analytical guidelines for combining review techniques with poststructural discourse analysis, provided by Nieminen et al. (2023). A poststructural analysis treats 'discourse' rather differently to linguistic analyses of discourse in its focus on institutionalised patterns of knowledge: it seeks to analyse how 'truths' are made rather than aiming to uncover such 'truths' themselves. An analysis of discourse within scholarly research shows how certain objects and practices can be *known* in a certain social, cultural and historical contexts. Discourse analysis can be used to 'understan[d]' the contemporary practices through which individuals constitute themselves as subjects of knowledge' (Arribas-Ayllon & Walkerdine, 2008, p. 111).

A discursive approach to self-assessment research asks: through what kinds of discourses is self-assessment known, and, how has it become possible to know self-assessment in this way, at this specific time? (Nieminen, 2021; Pitkänen, 2023). In practice, we focused on the enunciations of *how* self-assessment is made known (RQ1), and *what* is known about self-assessment (RQ2). We sought for statements of truth in the review studies: what they presented as legitimate truths about self-assessment, and how these truths were legitimised. For example, many review studies tested the accuracy of self-assessment and produced knowledge claims about whether or not self-assessment has been shown to be accurate, and to what extent. We then focused on enunciations about how such truths were

made legitimate. For example, the accuracy of self-assessment was made tangible by testing statistical correlations between self-assessment and other forms of assessment. In such situations, we were able to analyse how statistics was used to legitimate knowledge claims, and how comparison operated as a legitimising tool (namely, the legitimacy of self-assessment derived from its correlations with *more legitimate* assessment practices). We also paid attention to how research itself was positioned within discourse. As an apt example, we sought enunciations of how researchers positioned themselves in relation to teachers by asking the documents: who is positioned as a legitimate knowledge producer about self-assessment? Finally, we analysed how these discourses positioned students (RQ3). Following Nieminen et al. (2023), we focused on the roles and responsibilities assigned to students, both implicitly and explicitly. Namely, through discourse, subjects can occupy certain knower positions: discourses ‘offer positions from which a person may speak the truth’ (Arribas-Ayllon & Walkerdine, 2008, p. 117). In practice, we traced how the review studies positioned students concerning self-assessment, knowledge and truth. In other words, we analysed whether and how students were seen as legitimate knowledge producers in self-assessment.

We continuously developed and tested this ‘system’ of discourses so that each provided a unique, albeit not separate nor categorical, process of knowledge production regarding self-assessment. Namely, we revised our analysis until each discourse could be distinguished by its unique ways of producing legitimate knowledge about self-assessment.

An important caveat for our approach is that for a Foucault-inspired analysis, the authors nor the texts are seen as the origin of discourse. Instead, scholarly research is located within broader societal discourses that govern how self-assessment can be made known (Pais & Valero, 2012). Our analysis does not evaluate or judge the studies nor the people behind them. As Valero and Knijnik (2015) note, discourse analysis focuses on ‘the statements in their exteriority, not in what might exist behind what is said’ (p. 34).

What about the quality of our analysis? This is an essential question, as by writing this review, we are making truth statements and taking part in particular discourses ourselves. In this study, we have brought together literature review techniques and discourse analysis. We must first convince our readers that the ‘corpus of statements’ we have generated is meaningful. While we have relied on traditional review methods and reporting mechanisms, there are also limitations. We have excluded texts beyond journal articles (e.g. Brown & Harris, 2013) and only focused on studies published in English. Our study thus takes part in exclusionary scholarly discourses that have systematically omitted ‘other’ avenues for knowledge production, such as national and regional journals in languages beyond English that would have widened the scholarly voices in our corpus.

When it comes to justifying the ‘quality’ of our analysis, we must first emphasise that for Foucault-inspired discourse analysis, what counts as ‘real’ is a historical question about the status of truth. In the words of Arribas-Ayllon and Walkerdine (2008), discourses make objects thinkable. Discourses ‘do not “determine” things but intervene in the relations of what can be known, said, or practiced’ (Arribas-Ayllon & Walkerdine, 2008, p. 120). Thus, the ‘quality’ of our review is measured by whether it convincingly

shows how self-assessment has been made thinkable by review studies. In doing so, we intervene in the relations of truth. We have done our best to provide excerpts from our data corpus and to justify our truth statements to increase the reflexivity of our work.

Findings

Overview of the data corpus

The studies were published between 1982 and 2023 (Figure 1). The studies reviewed spanned from 1961 to 2022. The number of reviewed studies spanned between 11 and 207, with an average of 55 (four studies did not disclose the exact number).

Of the 28 studies, 15 were conducted solely in the higher education context and 10 were general reviews concerning multiple levels of education. Fifteen were disciplinary, and of these studies, 11 focused solely on medical education.

Next, we present our analysis of the discourses of self-assessment (Table 2). We did not classify or categorise the reviews with respect to these four discourses. Most reviews took part in various discourses simultaneously, whereas some of them produced truth statements on a more uniform basis.

Self-assessment as a psychological practice

Self-assessment was understood as a psychological (and psychometric) practice through a psychological discourse. This means that in our data corpus, self-assessment was mainly known through truth statements that gained legitimacy from objective, *scientific* research methodologies. Self-assessment was then seen as a clinical practice; as an *intervention* through which many desirable psychological outcomes could be nurtured. The focus here was on ‘the self’ as a psychological and cognitive

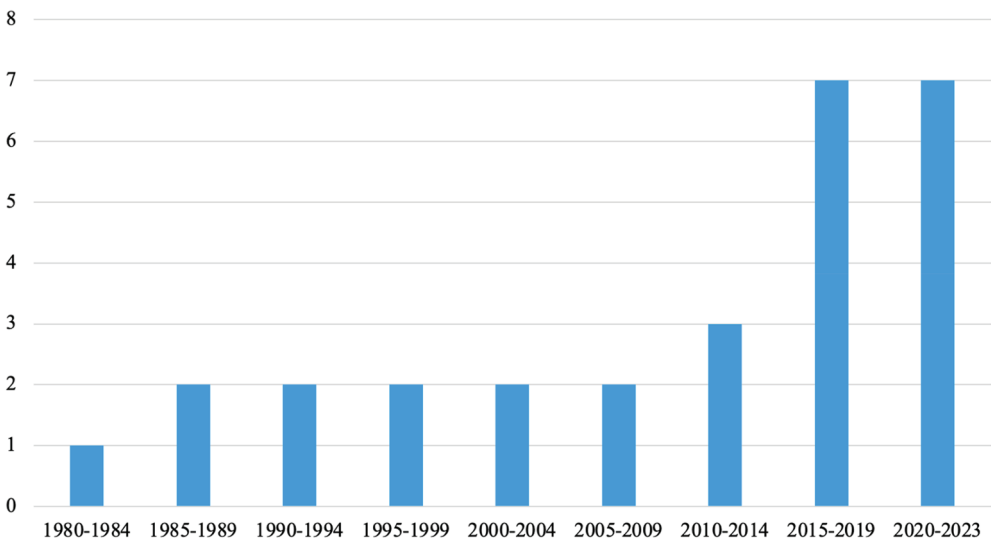


Figure 1. The number of studies published in 5-year intervals. The final year span only considers the four recent years.

Table 2. Overview of the discourses.

| Discourse | Definition | Application | Key findings | Role of research in the ecologies of assessment | Student positioning |
|---------------|--|--|---|--|--|
| Psychological | Self-assessment is an acontextual intervention that aims to nurture certain psychological capabilities and attributes. | Psychological and psychometric tools are used, and statistics in particular. Intervention designs are the gold standard in research. | Self-assessment is an effective intervention to promote students' psychological attributes such as self-regulation. | Research is objective and neutral. Researchers are seen as clinical experts. | Students as partakers of self-assessment interventions (cf. patients). Students as non-experts. Students as individuals. |
| Educational | Self-assessment is a relational practice that ideally builds on the relations between various educational stakeholders. | Research produces pragmatic, naturalistic and contextual knowledge about self-assessment. | Self-assessment is a useful formative process for enhancing the quality of students' work. | Research informs policy and strategies to improve educational practice. Researchers are not seen as knowledgeable agents alongside teachers. | Students as educational agents, connected to the world around them. |
| Performance | Self-assessment as a public performance. Self-assessment is deemed legitimate if it promotes student performance in external measures. | Research focuses on the interrelation between academic performance and self-assessment. | Self-assessment has been largely shown to boost student performance. | Research partakes in performativity cultures itself. | Students as performers. |
| Societal | Self-assessment is a non-neutral practice with societal purposes. | Research is conducted through normative argumentation: how education <i>should</i> be. | Successful self-assessment practices help protect the public from potentially incapable professionals. | Research not only provides implications for policy and practice but takes part in socio-political conversations. | Students as responsible citizens and lifelong learners. |

subject; indeed, research was seen as a way to explore such subjects' inner psychological processes. While such knowledge production is typical to the specific disciplines of educational psychology and cognitive sciences, we emphasise that this discourse was not limited to any particular discipline, but instead, the psychological discourse was used to study self-assessment in disciplines and fields such as medical education and language education. Our review showcases how psychological forms of knowledge production have found their way to various disciplines and fields in self-assessment research, at least when it comes to review studies.

The role of research, as seen through the psychological discourse, was to produce expert knowledge about the psychology of self-assessment. By using scientific methods to ensure the knowledge produced about self-assessment was objective – and in most cases quantitative – the research community could inform assessment policies and practices

and thus play an expert role in the ecosystems of assessment. The psychological discourse is not new, of course, as the roots of educational assessment are in measurement. While it is out of scope for us to introduce such histories, we guide readers to, for example, Broadfoot's (1998a) work that has called this discourse a 'modernist discourse' which is 'informed by notions of science and rationality' (Broadfoot, 1998b, p. 453). Our findings emphasise that self-assessment is tied to this historical continuum of knowledge production of assessment, and measurement in particular.

Scientific and rational methods were seen as essential in ensuring knowledge about self-assessment was rigorous, and that self-assessment itself could be considered *evidence-based*. It was seen as of utmost importance to use 'a robust methodology' to provide 'a solid evidence base for effective self-assessment' (Colthart et al., 2008, p. 14). Amongst rigorous review methodologies (e.g. systematic reviews), the reviews often called for methods borrowed from educational psychology, such as statistics, (quasi-) experimental studies with control groups, and studies conducted not in classroom settings but in laboratories (see e.g. Gordon, 1991, 1992).

Through such knowledge production processes, self-assessment itself was known as an *intervention* which could potentially produce effects for individuals such as enhanced self-regulation (e.g. Davis et al., 2006; Mays & Branch-Mays, 2016; Sanchez et al., 2017; Yan et al., 2022; Yan, Wang, et al., 2023). This view of self-assessment was largely acontextual and universal, meaning that the internal, psychological processes of students were studied in isolation. For example, many studies lamented the 'complex, ill-defined nature of [self-assessment]' and 'universally accepted definition[s]' of self-assessment (Yates et al., 2022, p. 1296). Moreover, interventions should, by definition, lead to positive outcome measures. In many reviews, self-assessment was examined from the viewpoint of *effectiveness*: whether and how well self-assessment was able to promote desirable psychological outcomes (see also the performativity discourse).

The 'black box' metaphor of self-assessment provides an intriguing way to understand the functioning of this discourse. For example, Andrade (2019) called for more research to look into the black box of self-assessment; into the 'cognitive and affective mechanisms of self-assessment' that are 'maddeningly difficult to measure' (p. 10). This metaphor aptly shows how the psychological process of self-assessment was seen as a 'box' with an 'outside' and 'inside'. Research could then aim to isolate the inner processes in order to produce knowledge about 'the box' in neutral ways.

Through the psychological discourse, self-assessment itself garnered its legitimacy by being proved by scholarly research as having psychometric attributes similar to other, more legitimate assessment practices such as tests: '...a key concern is whether the content and construct measured by [self-assessment] reflect the same content and construct measured by external measures' (Li & Zhang, 2021, p. 190). Following this discourse, many studies focused on exploring the psychometric properties of self-assessment practice (e.g. its validity and reliability). A prominent theme in the psychological discourse of self-assessment was accuracy; the 'validity concern' of self-assessment as phrased by Panadero et al. (2016, p. 812). Accuracy is, historically, a dominant theme in self-assessment reviews (Arthur, 1995; Brown et al., 2015; Falchikov & Boud, 1989; Gordon, 1991; Nayar et al., 2020; Wolff et al., 2017). Accuracy of self-assessment was determined 'by comparing a student's self-assessment to the judgements of qualified raters, such as teachers or fellow learners, or to performance on tests' (Brown et al., 2015,

p. 445). These ideas further illustrate the firm grounding of self-assessment in the historical continuum of knowledge production in psychometrics and measurement.

In fact, some review studies noted that we already know enough about accuracy (e.g. Boud & Falchikov, 1989). For example, Andrade (2019) challenged the research community to redirect its gaze on self-assessment as an instrument for learning: 'Future research should focus less on accuracy/consistency/veridicality' (Andrade, 2019, p. 9) Nonetheless, many studies portrayed psychometric properties as the 'sine qua non' of self-assessment:

Because reliability of testing is considered a sine qua non for the validity of assessment interpretations, and research into the human ability to self-evaluate work raises concerns about the quality of students' judgements, it is sensible to investigate the accuracy of students' self-assessments. (Brown et al., 2015, p. 444)

To summarise some key findings produced through the psychological discourse, self-assessment was shown to promote desirable psychological outcomes such as self-regulation and self-efficacy (Panadero et al., 2017) and academic achievement (Yan, Wang, et al., 2023). The reviews about accuracy were rather inconsistent in different educational contexts, yet many noted that student self-assessments are moderately connected with expert ratings. For example, Blanch-Hartigan (2011) reported a weighted mean correlation of $r = .21$ between student and expert ratings in medical education, and Ross (1998) reported an average correlation of $r = 0.63$ in second language education.

Through the psychological discourse, students were positioned as the targets of self-assessment interventions. Students' role was primarily seen as the 'black box' that experts could explore and 'open'. Notably, students were known as individuals; as 'boxes' that were separated from their environments through rigorous and objective research methods. Students were not seen as 'passive' partakers of self-assessment, as self-assessment was largely portrayed to foster student activity and metacognition. Even then, knowledge production about self-assessment gained legitimacy through psychologically sound methods. Students were seen as non-experts in such knowledge production processes, similar to partakers in clinical interventions.

Self-assessment as a relational practice

The psychological discourse was countered by what we call an *educational discourse*. Drawing on Biesta (2009), the word 'education' implies relationality: '...someone educating someone else and the person educating thus having a certain sense of what the purpose of his or her activities is' (p. 39). The educational discourse framed self-assessment as a natural part of student–teacher relations; as a student-centred practice that focuses on 'learner autonomy' and a 'conceptual shift from teacher-centered instruction to learner-centered instruction' (Li & Zhang, 2021, p. 190). Self-assessment was, then, not seen to occur in a social vacuum: '... self-assessment is inherently flawed when used in isolation' (Wolff et al., 2017, p. 116). Teachers' role, then, was to scaffold students into more independent practice (Yan, Panadero, et al., 2023), yet not through interventions but through relational practice.

Review studies taking part in the educational discourse produced practical, *useful* knowledge that could be used in various educational *contexts*. The legitimacy of self-assessment research was determined by how well research would be able to inform educational practice. Many studies had lengthy sections regarding ‘implication for practice’; yet only a small subset of the studies produced primarily educational knowledge (e.g. Andrade, 2019). Some review studies explicitly challenged the cognitivist mainstream literature of self-assessment research. For example, Otero-Saborido et al. (2021) challenged earlier quantitative studies as they did not provide a ‘holistic understanding’ of self-assessment but instead ‘served descriptive purposes that did not necessarily improve students’ teaching – learning processes’. (p. 10) Likewise, Boud and Falchikov (1989) challenged the dominance on measurement studies as ‘educationally misguided’, instead emphasising the ‘practical issue’ of self-assessment:

It is important to acknowledge that although the making of quantitative measurements in order to judge the reliability of student self-marking is a significant practical issue, it is certainly not the most important aspect of self-assessment. Indeed, it can be argued that to focus on this issue is to be educationally misguided. Effort which is directed towards narrowing the gap between student and teacher ratings might more profitably be directed towards developing ways in which systematic formative self-assessment activities can be incorporated into courses to improve student skills in making sensitive and aware judgements on their own work. (Boud & Falchikov, 1989, p. 532)

Self-assessment itself was known through its educational purposes and techniques. For example, Andrade (2019) and Papanthymou and Darra (2019) discussed the summative and formative purposes of self-assessment. Notably, Panadero et al. (2016) synthesised various self-assessment typologies to understand the educational foundations of self-assessment. These typologies focused on, for example, the involvement of students and teachers in the self-assessment process (Tan, 2012), power and transparency (Taras, 2016) and the presence and form of self-assessment criteria (Alonso-Tapia & Panadero, 2010). This is a thoroughly educational approach to self-assessment that acknowledges the relationality of assessment, which, as seen through this discourse, always takes place in a specific social, cultural, political and historical context.

To summarise some key findings, the following design elements were reported to support the self-assessment process:

- Explicit educational purpose for self-assessment (e.g. Boud & Falchikov, 1989)
- Clear criteria for self-assessment (e.g. Panadero et al., 2016)
- Modelling and support (e.g. Yan et al., 2021)
- Instruction and practice (e.g. Sitzmann et al., 2010)
- Feedback during the self-assessment process (e.g. Mays & Branch-Mays, 2016)
- Keeping self-assessment formative (e.g. Andrade, 2019)
- Aligning self-assessment with curricular goals (e.g. Ross, 1998)
- Commitment to the self-assessment process by both students and teachers (e.g. Arthur, 1995)

By taking part in an educational discourse, these reviews explicitly aimed to impact self-assessment practices and policies. The role of self-assessment research communities was,

then, not only to provide external expertise, but to *educate* and take part in relational dialogues. Likewise, through the educational discourse, educational stakeholders such as teachers, parents and students were considered legitimate sources of truth. Consider, for example, how Arthur (1995) explains how self-assessment could train students in providing legitimate knowledge about their own abilities:

Assuming that a goal of independent practitioners is the ability to establish their own criteria, collect their own data, make judgements and decisions, and to take appropriate actions, students may be assisted in moving from dependence (the student stage of their learning) to independence (the professional stage in their learning) by carrying out this process on their own. (Arthur, 1995, p. 275)

The educational discourse framed students as educational agents who operate within their context. This means that students were seen as grounded in their social and cultural contexts alongside their educational system. Students were portrayed as learners who existed within a relationship with teachers, whose role was to scaffold self-assessment practice to gradually increase the independence of students in self-assessment. As Pitkänen (2023) explains, within such discourse, self-assessment takes the role of pastoral care. As students are asked to disclose personal information through self-assessment, they are shaped and shape themselves as particular kinds of educational subjects.

Self-assessment as a (booster of) performance

What we call a *performance discourse* ties student self-assessment with wider cultures of performativity in education. Whereas the psychological discourse produced knowledge about the *internal*, psychological processes of students taking part in self-assessment, the performativity discourse instead focused on the *external* actions, behaviours and outcomes of self-assessment. Through this discourse, research produced knowledge about self-assessment by rendering it into something tangible. Whereas some reviews suggested self-assessment should be *private* (e.g. Brown et al., 2015), most of the reviews focused on analysing whether and how *public* self-assessment practices could foster tangible outcomes. We use the term ‘performance’ to denote that self-assessment was made known as a design practice that was performed by students in visible ways (such as by filling self-assessment forms), often with an initiation by educators (rather than by students themselves) and for the purposes defined by educators (such as for boosting students’ test scores). This mostly extrinsic, visible and public nature of self-assessment has been noted in various earlier definitions of self-assessment (see e.g. Bourke, 2016). It seems to form the backbone of current scholarly understandings of self-assessment.

The performativity discourse legitimised self-assessment as a meaningful practice based on the outcomes it produced. It is possible to connect self-assessment with broader performativity cultures and discourses of education, as discussed by sociologists of education such as Ball (2003). In many contexts, performativity cultures have for decades been ‘reorientating education, educational institutions and students to the competitive needs of the economy’ (Ball, 1998, p. 189) (cf. the societal discourse). In Biesta’s (2009) words, assessment often promotes a culture in which ‘means become ends in themselves’ (p. 35). Here, we do not portray performativity cultures as something necessarily ‘avoidable’. After all, many ‘means’ reached through self-assessment are crucial curricular

goals. Rather than making such normative statements, we will showcase how the reviews in our analysis took part in the performativity discourse by legitimising self-assessment through its performative processes and outcomes, which we argue must be understood while making sense of contemporary knowledge production processes of self-assessment.

In the reviews, the educational benefits of self-assessment were commonly based on the tangible and measurable outcomes of self-assessment. Self-assessment was seen to gain its legitimate, evidence-based status if it was deemed through review methodologies to foster student performance in other, more legitimate measures of achievement, which again often gained legitimacy through measurement and statistics (cf. the psychological discourse). Here, self-assessment could be connected with educational performativity cultures by examining it as a *booster* for students' external performance measures: self-assessment was seen to have 'great potential to improve students' academic performance' (Yan et al., 2022, p. 1). Through the performance discourse, self-assessment became known through its *instrumental value*, namely, whether it could promote students' performance in external assessments. The review studies indeed focused largely on whether and how self-assessment could promote student performance in tests (e.g. Li & Zhang, 2021; Mabe & West, 1982; Papanthymou & Darra, 2019; Sanchez et al., 2017; Sitzmann et al., 2010; Yan, Wang et al., 2023). This effectiveness of self-assessment on student performance in many reviews determined its legitimacy as an assessment practice.

It was found in many reviews that self-assessment boosts student performance in external measures (Andrade, 2019; Colthart et al., 2008; Li & Zhang, 2021; Papanthymou & Darra, 2019). Sanchez et al. (2017) found in their meta-analysis that 12 of their 44 effect sizes for both self- and peer-assessment were negative, yet on average 'students who engaged in self-grading performed better ($g = 0.34$) on subsequent tests than did students who did not' (p. 1049). Yan et al. (2022) meta-analysis concluded that for self-assessment interventions, the performance gain had an effect size of $g = 0.585$ (and $g = 0.448$ when self-assessment was used alongside peer-assessment). Various design elements of self-assessment were reported to promote higher performance, such as student training on self-assessment (e.g. Colthart et al., 2008; Li & Zhang, 2021), feedback processes during the self-assessment process and the explicitness of the self-assessment task (Sanchez et al., 2017; Yan, Wang, et al., 2023).

Taking part in the performativity discourse positions self-assessment research itself as a part of wider performativity cultures (Ball, 1998, 2003). As such, self-assessment becomes a tool to promote students' performance for the sake of the students themselves as well as their institutions (cf. the societal discourse). The legitimacy of self-assessment research was determined by whether it is able to promote student performance or reveal new knowledge about how this could be done.

Through the performativity discourse, students' primary responsibility was to perform self-assessment visibly and benefit from it in tangible, predetermined ways. Such a performative way of knowing students does not mean students were positioned as 'performers', as in doing self-assessment 'for show'. Instead, it means students are positioned as partakers in the tangible, mostly public performances of self-assessment who would benefit from these practices in tangible and often measurable ways. Students

are thus positioned within broader educational cultures of performativity as educational subjects who would ‘identify their own strengths and weaknesses aiming for further improvement’ (Yan et al., 2022, p. 2). Such a positioning emphasises the public (rather than private or undisclosed) performances of self-assessment, and the striving for tangible benefits and improvement (rather than for students’ own, undisclosed aspirations).

Self-assessment as a societal practice

Finally, self-assessment was constructed as a socio-political matter via a *societal discourse*. This means that self-assessment was seen as a *normative* practice that should align with wider societal values, discourses and orders. Thus, how self-assessment practices were designed and implemented were questions about functioning societies. This idea was not the main focus of any of the studies, nor was this discourse seen explicitly in the findings of the studies. Instead, research took part in societal discourse by justifying self-assessment (research) in normative ways. Through this discourse, normative statements and ideologies were seen as legitimate parts of knowledge production processes rather than as something to be avoided, perhaps in the name of ‘objectivity’. Many reviews discussed how self-assessment *should be* conducted. In doing so, research itself was not positioned as ‘apolitical’ but transparently took part in wider societal discourses and power struggles related to assessment.

Many studies investigated how self-assessment has promoted student learning outcomes (cf. performativity discourse). However, the ‘performance’ in itself was not only a matter of individual students but about the performance of schools, institutions and societies. Many studies connected self-assessment with lifelong learning that would benefit students and society at large (e.g. Boud & Falchikov, 1989; Gordon, 1992; Ward et al., 2002; Yates et al., 2022). Yan and colleagues tied self-assessment with the Sustainable Development Goals of the United Nations.

A fundamental goal of higher education is to develop students’ capacity for life-long learning (...), which is also one of the United Nation’s Sustainable Development Goals. Student self-assessment plays a pivotal role in achieving this goal because, from a pedagogical perspective, effective self-regulated or life-long learning is more likely to happen when students have a realistic sense of their own performance so that they can direct their further learning (...). (Yan, Wang et al., 2023, p. 1; references omitted)

Studies conducted in medical education understood self-assessment practice as a way to guard the public against poor practitioners. Here, the societal functioning of self-assessment is seen clearly: flawed self-assessment produced risks not only for individuals but for societies at large (Dunning et al., 2004; Nayar et al., 2020; Wolff et al., 2017). For example, without well-designed self-assessment practices, higher education would not be able to produce capable health practitioners:

Accurate self-assessment of performance allows physicians to understand their own strengths and weaknesses and know which areas to focus on in their lifelong medical education (...). (Blanch-Hartigan, 2011, p. 3)

Some studies explicitly addressed the role of self-assessment within societal assessment systems of selection. In such systems, standardised testing apparatuses were seen to

produce legitimate assessment knowledge for the purpose of societal decision-making, whereas ‘self-assessments should not count towards grades and should be private’ (Brown et al., 2015, p. 452). In this way, self-assessment was seen to fit the overall systems of selection and accountability by providing something crucial yet of lesser priority and legitimacy; something essential but secondary.

The societal discourse portrayed students as part of the societies around them. The role of self-assessment was to train students to function as members of these societies; scholarly research could then help students reach their full potential as citizens.

Discussion

Our meta-review of earlier review studies has identified four discourses through which self-assessment has been made *known*: the discourses of psychology, education, performativity and society. These four discourses pervade five decades of self-assessment reviews, including hundreds of original studies since the early 1960s. We have briefly outlined the main findings from these reviews, as produced through the four discourses, and analysed the positioning of students within these discourses. By focusing on review studies, we have synthesised powerful research artefacts; reviews are considered essential for research fields to grow and develop (Panadero et al., 2017). As such, these studies are influential in defining how self-assessment is conceptualised and discussed. Overall, we hope that our study helps self-assessment research navigate the various research approaches and paradigms by providing a ‘roadmap’ of the knowledge production processes in this field.

Shedding light on the discursive alignments and tensions in self-assessment research

Our analysis has shed light on the discursive alignments and tensions within self-assessment research. As such, it enables ‘meta tools’ to discuss various research approaches. Each of the four discourses builds on different ideas concerning the legitimacy of knowledge in self-assessment and self-assessment research. A greater awareness of these discourses could help the self-assessment research community to seek out intersections, dialogues, and collaborations between research projects. After all, what connects these discourses is *care for self-assessment*:: the will to produce knowledge that in the future might help foster meaningful learning experiences for students.

Different discourses are deployed by researchers for different purposes and become dominant in different contexts. As a given discourse limits what can be seen and acted upon in a field, it is crucial to be aware of the compatibility as well as ruptures between various research approaches and paradigms. Our analysis has shown that self-assessment research may pursue the project of educational psychology to model learning processes that transcend any particular context; to generate practical knowledge to design educational opportunities for students to develop and extend their self-assessment capabilities within given courses; to focus on an agenda of performance which prioritises the achievement of student learning outcomes; or, finally, to explore how self-assessment should be conducted as part of the societal role of assessment. Combinations of these discourses may be deployed together either deliberately, or, often unconsciously, with

one practice of self-assessment used for quite different ends. We need to understand better how they operate separately and together and what effects they create.

We thus call for further reflexivity to self-assessment research on the discourses within this field. We are particularly worried about the siloing of academic conversations, following Yates et al. (2022), in our data corpus. We particularly welcome studies that would explore how these discourses operate together and how they may strengthen and weaken each other. Our ‘roadmap’ can be used to understand which discourses dominate scholarly conversations in certain contexts, and simultaneously, which discourses are left in the margins while producing knowledge about self-assessment. Moreover, we believe these discourses could help future studies in self-assessment research in reporting their findings. Future reporting could rely on these four discourses to more transparently ground the studies within specific paradigms, conversations and problem spaces. For example, scholars could ponder their own researcher positioning, the goal of self-assessment being pursued, the scope of claims being identified, the consciousness of contextualisation, and limitations of findings. Any given review needs to acknowledge the discourse from which each study comes and be cautious about assuming that findings can be simply cumulative. Research needs to disaggregate the different orientations of self-assessment and seek to identify which projects they contribute to. Synthesising knowledge from projects with differing ontological and epistemological traditions remains an important task for future review studies.

Self-assessment research within its broader context

By conducting this meta-review, we have situated these literature reviews in their broader context, namely, the ecosystems of assessment worldwide. As these studies produce powerful knowledge about self-assessment, it is necessary to ask how the four discourses of scholarly knowledge production are situated within broader educational discourses. We see this timely given how strongly research has advocated for self-assessment in assessment policy and practice, and how self-assessment remains secondary in practice.

One of our key contributions has been demonstrating how scholarly discourses concerning self-assessment do not operate in a vacuum but connect with and are made possible by broader societal discourses of assessment. Self-assessment has long historical roots in both ‘Western’ and ‘Eastern’ philosophical traditions. Our meta-review has examined how contemporary self-assessment research has continued the ethico-philosophical ideal of ‘knowing thyself’ as a fundamental virtue in education and societies at large. Future research could further unpack the intertwinements between scholarly and ‘other’ forms of discourses. For example, our study has deconstructed a discourse of self-assessment that focuses on performativity and visibility of self-assessment. Whether or not this is the intention of research communities, such discourses may steer work on self-assessment closer to broader educational cultures of performativity, productivity and accountability (Ball, 1998, 2003). Wider educational cultures related to, for example, competitiveness and testing cultures are likely to marginalise self-assessment as something subordinate to external measures. In doing so, the global performativity and testing cultures are both supported and strengthened by educational research. It is this positioning of self-assessment as secondary and illegitimate practice that, we argue, has at times sidelined the impact of self-assessment as

a practice and distorted research that has been conducted on it, again, perhaps regardless of the intentions of researchers. The performativity discourse thus provides an apt example of how research cannot operate separately from broader societal shifts in policy, practice and discourse.

The Foucauldian understanding of discourse emphasises the constant struggle and flux of discourse. One of the ‘discursive struggles’ in our data corpus regarded the tensions between seeing self-assessment as a local *and* universal practice. Namely, the psychological and performativity discourses legitimised self-assessment through their scientific and objective knowledge production processes, whereas the educational and societal discourses valued normative and contextual forms of knowledge. How these different ‘truths’ operate in tandem represents a timely issue for future research to unpack. Our review has shown how self-assessment research is tied simultaneously to the historical roots of psychology and measurement (e.g. Broadfoot, 1998a) and the more recent paradigm of student autonomy, agency and self-regulation, as demonstrated by the Assessment for Learning movement (e.g. Volante et al., 2025). Self-assessment has historically been ‘known’ outside scholarly discourse in policy, curriculum and practice in ways that might differ from academic discourse. For example, Pitkänen (2022) has analysed the discourses of self-assessment in Finnish curriculum documents since the 1950s, noting that self-assessment has been known differently over the history through the values of critical thinking, citizenship, autonomy, lifelong learning, reflectivity and entrepreneurship. Such normative values, however, have in many contexts developed under the influence of the psychometric movement since the early 1900s with its quest towards ‘rationality’. Thus, tensions between local and global knowledge production have always been at the core of self-assessment work, and we see it crucial for future research to situate itself within these discursive formations in flux.

How could self-assessment research have a more meaningful impact on assessment policy and practice? Here, we draw on Bourke’s (2018) work on how self-assessment could promote students’ ontological awareness. Likewise, self-assessment research could demonstrate ontological awareness by recognising how educators and students are positioned differently through different kinds of discourse. For example, the universal understanding of self-assessment via psychological discourse may hinder successful implementations of self-assessment if such universality does not sit within a grounded, situated classroom context. By developing ontological awareness and humility, self-assessment research could recognise and value voices from various approaches and paradigms. Given how self-assessment remains somewhat secondary in policy and practice, we see it crucial for research communities to explore whether and how students and educators could be portrayed as legitimate producers of knowledge in self-assessment. Otherwise, these stakeholders will arguably not willingly participate in self-assessment – at least not in ways determined by research communities.

An ontological awareness in research would also include an awareness of the power struggles and imbalances between various discourses. While we see it as essential to foster collaborations between discursive divisions, there are also discrepancies and incompatibilities between the four discourses. Who has the power to tell truths about student learning? Historically, it was first the teachers who held such power, whereas later on it has been increasingly taken over by the machineries of assessment (e.g. standardised testing). Focusing on scholarly research only, our study has shown how students and

educators are sometimes portrayed as legitimate sources of knowledge (e.g. the educational discourse), yet not always (e.g. the psychological discourse). While we see plenty of value in the knowledge production processes drawn on psychology and cognitive sciences, we also note the risks of what such discourse might produce, namely, a potential devaluation of the local. Currently, ‘neuroeducational’ approaches have considerable rhetorical power in global debates about assessment policy and practice. In this context, we find it crucial to consider how self-assessment could foster non-scientific communities, particularly students, as legitimate sources of knowledge.

Where is the ‘self’ in self-assessment?

Our analysis has discussed how students are positioned in self-assessment within scholarly discourses. While our study has focused on literature reviews, educational research ultimately concerns real students. The discourses that research takes part in and reproduces may find their way into educational systems, schools and classrooms: they render students ‘knowable’ in particular ways. In this final section, we emphasise the ethics of conducting research about students, calling attention to an aspect that connects different research traditions across multiple discourses, namely, the care for the students.

The reviews in our analysis have produced knowledge about how self-assessment could be used to shape students as educational subjects. Since self-assessment ontologically centres ‘the self’, it is ‘particularly powerful in its making and shaping of pupil subjectivities’ (Pitkänen, 2022, p. 729; see also Bourke, 2018). Our analysis has shown that students were not always portrayed as legitimate ‘knowers’ in self-assessment within scholarly discourse. For example, the psychological discourse positioned students as ‘non-experts’ whose self-assessment was deemed ‘truthful’ if its psychometric properties were considered adequate. On the other hand, through the educational discourse, academic research valued students’ ‘truths’ as legitimate statements about their skills and abilities. Following Pitkänen (2022), this illustrates the potential of self-assessment in simultaneously subjecting students to discipline, surveillance, pastoral care and self-formation.

We anticipate that such discursive tensions about whether and how students are seen as legitimate knowers in self-assessment may impact how students come to value self-assessment as a practice, and themselves as ‘truth-sayers’. All four discourses can produce knowledge that positions students as *objects of the actions of others* rather than as meaningful educational subjects themselves. In such situations, the ‘self’ is occluded in self-assessment. This may occur as self-assessment is used to instil predetermined values such as independence, autonomy and diligence (cf. societal discourse), or if self-assessment is deemed legitimate only when it is conducted exactly in ways determined by educators, for the purposes of educators, and when its success is measured by ways determined by educators (cf. performativity discourse). Our meta-review has provided new conceptual tools for research communities to understand how self-assessment (research) may act as a technology to steer and control students through expert practices for goals determined by not ‘the self’ but by others (see DeLuca et al., 2022; Nieminen, 2021; Tan, 2004; Taras, 2016). Indeed, ‘the self’ has been portrayed as an ideally reflective, rational, individualistic and autonomous agent in much of contemporary work on self-

assessment, envisaging a particular historical understanding of ‘the self’ that is rooted for broader societal ideals about individuality and modernity (see Pitkänen, 2022, 2023).

So, where is the ‘self’ in self-assessment research? Answering this question provides an important starting point for future research. This, we argue, should be a central question for self-assessment research. Without the ‘self’ being in the driving seat of assessment, self-assessment can become a mere technology of control by others to achieve ends other than the student’s. This is the common paradox of all educational work on self-direction or self-determination. When the student becomes the subject of interventions regulated by others and researched for others’ perspectives, the key features of the self can be denied and decision-making by the student removed or negotiated away. This is not uncommon in education, and there is always a balance between doing what teachers think is in the best interest of learners and what learners believe it to be. Yet, this paradox becomes more drastic in work on self-assessment given its supposed centring of ‘the self’. Students may only recognise the central role of self-assessment in their own development if they are provided with opportunities to know themselves as legitimate producers of ontological knowledge about themselves (see Bourke, 2018). Whatever the positioning of students has been in the literature to date, we suggest the need for a strong theme of students having agency is necessary if the term self-assessment is to continue to be accepted. Such a goal requires the research community to reflect upon what role the self-assessment knowledge generated by students themselves plays in the academic knowledge production processes. How are students’ ‘truths’ situated amid the scholarly ‘truths’? One opportunity may arise from seeing students as meaningful contributors in self-assessment research. As Bourke (2016) notes, when self-assessment is incorporated from a learner’s perspective, ‘it can lead the conversations on learning, thus potentially liberating learners from the confines of a pre-determined agenda and feasibly allowing them to achieve a greater balance between understanding their learning needs, and that of the curriculum’. (p. 108)

Final words

In this meta-review, we have synthesised the scholarly knowledge production processes regarding self-assessment research through five decades. We have discussed how self-assessment research has contributed to knowledge about the reflective self in particular ways. The four discourses of psychology, education, performativity and society represent four differing ways of producing legitimate knowledge about self-assessment. Awareness of these discourses may enable the research field to reflect upon its methods, goals and values. Where to go from here? We hope our meta-review will help to orient self-assessment research towards the unknown futures.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Dr. Juuso Henrik Nieminen is a Senior Research Fellow at The Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University, Australia. His research concerns the social, cultural and political dimensions of assessment with a particular focus on how assessment shapes student identities in contemporary societies.

David Boud is Deakin Distinguished Professor and Foundation Director of the Centre for Research in Assessment and Digital Learning at Deakin University, Australia. He is also Emeritus Professor in the Faculty of Arts and Social Sciences at the University of Technology Sydney. He has published extensively on teaching, learning and assessment in higher and professional education. He has been a pioneer in developing learning-centred approaches in assessment and in changing conceptions of feedback.

ORCID

Juuso Henrik Nieminen  <http://orcid.org/0000-0003-3087-8933>

David Boud  <http://orcid.org/0000-0002-6883-2722>

References

Studies in the review are marked with an asterisk (*).

- Alonso-Tapia, J., & Panadero, E. (2010). Effects of self-assessment scripts on self-regulation and learning. *Infancia Y Aprendizaje*, 33(3), 385–397. <https://doi.org/10.1174/021037010792215145>
- *Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4(1), 1–13. <https://doi.org/10.3389/educ.2019.00087>
- Arribas-Ayllon, M., & Walkerdine, V. (2008). Foucauldian discourse analysis. In C. Willig & W. Stainton-Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 91–108). Sage.
- *Arthur, H. (1995). Student self-evaluations: How useful? How valid? *International Journal of Nursing Studies*, 32(3), 271–276. [https://doi.org/10.1016/0020-7489\(94\)00043-J](https://doi.org/10.1016/0020-7489(94)00043-J)
- Ball, S. J. (1998). Performativity and fragmentation in ‘postmodern schooling’. In J. Carter (Ed.), *Postmodernity and the fragmentation of welfare* (pp. 187–203). Routledge.
- Ball, S. J. (2003). The teacher’s soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21(1), 33–46. <https://doi.org/10.1007/s11092-008-9064-9>
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Wyatt-Smith, C., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140. <https://doi.org/10.1177/1478210314566733>
- *Blanch-Hartigan, D. (2011). Medical students’ self-assessment of performance: Results from three meta-analyses. *Patient Education & Counseling*, 84(1), 3–9. <https://doi.org/10.1016/j.pec.2010.06.037>
- Boud, D. (1989). The role of self assessment in student grading. *Assessment & Evaluation in Higher Education*, 14(1), 20–30. <https://doi.org/10.1080/0260293890140103>
- Boud, D. (2013). *Enhancing learning through self-assessment*. Routledge.
- *Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), 529–549. <https://doi.org/10.1007/BF00138746>
- Bourke, R. (2016). Liberating the learner through self-assessment. *Cambridge Journal of Education*, 46(1), 97–111. <https://doi.org/10.1080/0305764X.2015.1015963>

- Bourke, R. (2018). Self-assessment to incite learning in higher education: Developing ontological awareness. *Assessment & Evaluation in Higher Education*, 43(5), 827–839. <https://doi.org/10.1080/02602938.2017.1411881>
- Broadfoot, P. (1998a). Quality standards and control in higher education: What price life-long learning? *International Studies in Sociology of Education*, 8(2), 155–180. <https://doi.org/10.1080/0962021980020022>
- Broadfoot, P. (1998b). Records of achievement and the learning society: A tale of two discourses. *Assessment in Education Principles, Policy & Practice*, 5(3), 447–477. <https://doi.org/10.1080/0969595980050307>
- *Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education Principles, Policy & Practice*, 22(4), 444–457. <https://doi.org/10.1080/0969594X.2014.996523>
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Sage.
- Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, 3, 22–30. <https://doi.org/10.14786/flr.v2i1.24>
- *Colthart, I., Bagnall, G., Evans, A., Allbutt, H., Haig, A., Illing, J., & McKinstry, B. (2008). The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME guide no. 10. *Medical Teacher*, 30(2), 124–145. <https://doi.org/10.1080/01421590701881699>
- Covidence Systematic Review Software. (2023). *Veritas health innovation*. www.covidence.org
- Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26(3–4), 365–378. <https://doi.org/10.1080/713688543>
- *Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R. T. K. E., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *JAMA*, 296(9), 1094–1102. <https://doi.org/10.1001/jama.296.9.1094>
- DeLuca, C., Willis, J., Dorji, K., & Sherman, A. (2022). Cultivating reflective teachers: Challenging power and promoting pedagogy of self-assessment in Australian, Bhutanese, and Canadian teacher education programs. *Power & Education*, 15(1), 5–22. <https://doi.org/10.1177/17577438221108240>
- *Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Eva, K. W., & Regehr, G. (2008). “I’ll never play professional football” and other fallacies of self-assessment. *The Journal of Continuing Education in the Health Professions*, 28(1), 14–19.
- Eva, K. W., Regehr, G., & Gruppen, L. D. (2012). Blinded by ‘insight’: Self-assessment and its role in performance improvement. In B. D. Hodges & L. Lingard (Eds.), *The question of competence: Reconsidering medical education in the twenty-first century* (pp. 131–154). Cornell University Press.
- *Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395–430. <https://doi.org/10.3102/00346543059004395>
- Foucault, M. (1972). *The archaeology of knowledge*. Pantheon Books.
- *Gordon, M. J. (1991). A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine*, 66(12), 762–769. <https://doi.org/10.1097/00001888-199112000-00012>
- *Gordon, M. J. (1992). Self-assessment programs and their implications for health professions training. *Academic Medicine*, 67(10), 672–679. <https://doi.org/10.1097/00001888-199210000-00012>
- Gough, D., Oliver, S., & Tomas, J. (2012). *An introduction to systematic reviews*. Sage.
- Harris, L. R., & Brown, G. T. (2018). *Using self-assessment to improve student learning*. Routledge.
- Heron, J. (1981). Assessment revisited. In D. Boud (Ed.), *Developing student autonomy in learning* (pp. 55–68). Routledge.

- Kruger, J. M., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality & Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- *Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, 38(2), 189–218. <https://doi.org/10.1177/0265532220932481>
- Lum, A., Wakefield, C. E., Donnan, B., Burns, M. A., Fardell, J. E., & Marshall, G. M. (2017). Understanding the school experiences of children and adolescents with serious chronic illness: A systematic meta-review. *Child: Care, Health and Development*, 43(5), 645–662. <https://doi.org/10.1111/cch.12475>
- *Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3), 280. <https://doi.org/10.1037/0021-9010.67.3.280>
- *Mays, K. A., & Branch-Mays, G. L. (2016). A systematic review of the use of self-assessment in preclinical and clinical dental education. *Journal of Dental Education*, 80(8), 902–913. <https://doi.org/10.1002/j.0022-0337.2016.80.8.tb06170.x>
- *Nayar, S. K., Musto, L., Baruah, G., Fernandes, R., & Bharathan, R. (2020). Self-assessment of surgical skills: A systematic review. *Journal of Surgical Education*, 77(2), 348–361. <https://doi.org/10.1016/j.jsurg.2019.09.016>
- Nieminen, J. H. (2021). Beyond empowerment: Student self-assessment as a form of resistance. *British Journal of Sociology of Education*, 42(8), 1246–1264. <https://doi.org/10.1080/01425692.2021.1993787>
- Nieminen, J. H., Bagger, A., Padilla, A., & Tan, P. (2023). Student positioning in mathematics assessment research: A critical review. *Journal for Research in Mathematics Education*, 54(5), 317–341. <https://doi.org/10.5951/jresemetheduc-2022-0030>
- *Otero-Saborido, F. M., Torreblanca-Martínez, V., & Gonzalez-Jurado, J. A. (2021). Systematic review of self-assessment in physical education. *International Journal of Environmental Research and Public Health*, 18(2), 766. <https://doi.org/10.3390/ijerph18020766>
- Pais, A., & Valero, P. (2012). Researching research: Mathematics education in the political. *Educational Studies in Mathematics*, 80(1–2), 9–24. <https://doi.org/10.1007/s10649-012-9399-5>
- *Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803–830. <https://doi.org/10.1007/s10648-015-9350-2>
- *Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>
- *Papanthymou, A., & Darra, M. (2019). The contribution of learner self-assessment for improvement of learning and teaching process: A review. *Journal of Education & Learning*, 8(1), 48–64. <https://doi.org/10.5539/jel.v8n1p48>
- Pitkänen, H. (2022). The politics of pupil self-evaluation: A case of Finnish assessment policy discourse. *Journal of Curriculum Studies*, 54(5), 712–732. <https://doi.org/10.1080/00220272.2022.2040596>
- Pitkänen, H. (2023). Reflektioiva oppilas suomalaisessa perusopetuksen opetussuunnitelmallisessa arviointidiskurssissa. *Kasvatus & Aika*, 17(1). <https://doi.org/10.33350/ka.119400>
- Richardson, M. (2022). *Rebuilding public confidence in educational assessment*. UCL Press.
- *Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1–20. <https://doi.org/10.1177/026553229801500101>
- *Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, 109(8), 1049. <https://doi.org/10.1037/edu0000190>

- *Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning and Education*, 9(2), 169–191. <https://doi.org/10.5465/amle.9.2.zqr169>
- Smith, W. C. (Eds.). (2016). *The global testing culture: Shaping education policy, perceptions, and practice*. Symposium Books Ltd.
- Sumner, F. C. (1932). Marks as estimated by students. *Education*, 32, 429.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>
- Tan, K. H. (2004). Does student self-assessment empower or discipline students? *Assessment & Evaluation in Higher Education*, 29(6), 651–662. <https://doi.org/10.1080/0260293042000227209>
- Tan, K. H. K. (2012). *Student self-assessment. Assessment, learning and empowerment*. Research Publishing.
- Taras, M. (2016). Situating power potentials and dynamics of learners and tutors within self-assessment models. *Journal of Further and Higher Education*, 40(6), 846–863. <https://doi.org/10.1080/0309877X.2014.1000283>
- Valero, P., & Knijnik, G. (2015). Governing the modern, neoliberal child through ICT research in mathematics education. *For the Learning of Mathematics*, 35(2), 34–39. <https://www.jstor.org/stable/44382757>
- Van der Kleij, F. M., Adie, L. E., & Cumming, J. J. (2019). A meta-review of the student role in feedback. *International Journal of Educational Research*, 98, 303–323. <https://doi.org/10.1016/j.ijer.2019.09.005>
- Volante, L., DeLuca, C., Barnes, N., Birenbaum, M., Kimber, M., Koch, M., Wyatt-Smith, C., Poskitt, J., Smith, K., & Wyatt-Smith, C. (2025). International trends in the implementation of assessment for learning revisited: Implications for policy and practice in a post-COVID world. *Policy Futures in Education*, 23(1), 224–242. <https://doi.org/10.1177/14782103241255855>
- *Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7(1), 63–80. <https://doi.org/10.1023/A:1014585522084>
- *Wolff, M., Santen, S. A., Hopson, L. R., Hemphill, R. R., & Farrell, S. E. (2017). What's the evidence: Self-assessment implications for life-long learning in emergency medicine. *The Journal of Emergency Medicine*, 53(1), 116–120. <https://doi.org/10.1016/j.jemermed.2017.02.004>
- Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education Principles, Policy & Practice*, 22(1), 140–160. <https://doi.org/10.1080/0969594X.2014.990416>
- Yan, Z. (2022). *Student self-assessment as a process for learning*. Taylor & Francis.
- *Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A pairwise and network meta-analysis. *Educational Research Review*, 37, 100484. <https://doi.org/10.1016/j.edurev.2022.100484>
- *Yan, Z., Panadero, E., Wang, X., & Zhan, Y. (2023). A systematic review on students' perceptions of self-assessment: Usefulness and factors influencing implementation. *Educational Psychology Review*, 35(3), 81. <https://doi.org/10.1007/s10648-023-09799-1>
- *Yan, Z., Wang, X., Boud, D., & Lao, H. (2023). The effect of self-assessment on academic performance and the role of explicitness: A meta-analysis. *Assessment & Evaluation in Higher Education*, 48(1), 1–15. <https://doi.org/10.1080/02602938.2021.2012644>
- *Yates, N., Gough, S., & Brazil, V. (2022). Self-assessment: With all its limitations, why are we still measuring and teaching it? Lessons from a scoping review. *Medical Teacher*, 44(11), 1296–1302. <https://doi.org/10.1080/0142159X.2022.2093704>