Check for updates

Elements for understanding and fostering self-assessment of learning artifacts in higher education

Christian Köppe[1]*, Roald P. Verhoeff[2] and
Wouter van Joolingen[1]

[1]Freudenthal Institute, Utrecht University, Utrecht, Netherlands, [2]Radboud Teachers Academy,
Radboud University, Nijmegen, Netherlands

Self-assessment skills have long been identified as important graduate attributes. Educational interventions which support students with acquiring these skills are often included in higher education, which is usually the last phase of formal education. However, the literature on self-assessment in higher education still reports mixed results on its effects, particularly in terms of accuracy, but also regarding general academic performance. This indicates that how to foster self-assessment successfully and when it is effective are not yet fully understood. We propose that a better understanding of *why* and *how* self-assessment interventions work can be gained by applying a design-based research perspective. Conjecture mapping is a technique for design-based research which includes features of intervention designs, desired outcomes of the interventions, and mediating processes which are generated by the design features and produce the outcomes. When we look for concrete instances of these elements of self-assessment in the literature, then we find some variety of design features, but only a few desired outcomes related to self-assessment skills (mostly accuracy), and even less information on mediating processes. What is missing is an overview of all these elements. We therefore performed a rapid systematic literature review on self-assessment to identify elements that can help with understanding, and consequently foster an effective self-assessment of learning artifacts in higher education using conjecture mapping as analytical framework. Our review revealed 13 design features and six mediating processes, which can lead to seven desired outcomes specifically focused on self-assessment of learning artifacts. Together they form a model which describes self-assessment and can be used as construct scheme for self-assessment interventions and for research into the how and why self-assessment works.

KEYWORDS

self-assessment, learning artifacts, higher education, conjecture map, assessment design

1 Introduction

Self-assessment skills have long been identified as important attributes of graduates in higher education. They have been shown to positively influence students' learning and academic performance in general (Panadero et al., 2013), self-regulated and lifelong learning (Burgess et al., 1999; Dochy et al., 1999; Nicol and MacFarlane-Dick, 2006), and students' self-efficacy (Sitzmann et al., 2010; Panadero et al., 2023).

Several definitions of self-assessment have been given in the literature over the years. Table 1 provides an overview of some often cited definitions. While all of them are valid, we will use for this study the definition of Panadero et al. (2016) for following reasons:

- it includes different mechanisms and techniques for student judgments,
- it includes assessment *and* evaluation, and
- it includes student products.

Panadero et al. (2016) present and discuss different self-assessment typologies which reflect the distinctions and similarities of various self-assessment mechanisms and techniques such as self-marking, self-rating, self-grading, self-appraisal, or self-estimates. These distinctions are useful for classifications of self-assessment practices, yet for our study we prefer an approach which allows inclusion of all these mechanisms and techniques which is reflected in the self-assessment definition by Panadero et al. (2016).

Many scholars advocate that self-assessment should only be formative [see, e.g., (Brown et al., 2015; Andrade, 2019)], but there is evidence that summative forms such as self-grading or self-evaluation can be beneficial too (Edwards, 2007; Nieminen and Tuohilampi, 2020). Both forms are included in the self-assessment definition of Panadero et al. (2016) and will also be included in our work.

Products, which are created by students and used for evidencing their learning, can be termed *learning artifacts* (Cherner and Kokopeli, 2018). Such artifacts are often the direct results of assignments common in higher education (e.g., essays, design documents, or scientific reports), but sometimes they are especially created to facilitate assessment of learning (e.g., videos of oral presentations or audios of speaking tests).

As these artifacts are usually used by teachers for summative assessments (as evidence for learning), it is helpful for students to know how to assess their themselves and consistent with their teachers. Much self-assessment research focuses on this accuracy of self-assessment in terms of the level of consistency between student self-assessment and an external assessment, usually from a teacher (see, e.g., Brown et al., 2015; Han and Riazi, 2018; González-Betancor et al., 2019; Carroll, 2020).

Although accuracy of self-assessments is an important indicator of their validity and reliability, in our view the main goal of self-assessment is to help students with learning while producing and improving learning artifacts. This is in line with the positive effects on self-regulated learning (Panadero et al., 2017) and lifelong learning (Taranto and Buchanan, 2020) that have been related to self-assessment skills. Higher education, and specifically universities, can play an essential role in contributing to lifelong learning (Atchoarena,

2021). A study at a university is usually the last phase of formal education and therefore the last possible place for addressing the acquisition of self-assessment skills in a formal setting. In our study, we will focus on understanding, and consequently fostering, self-assessment of learning artifacts in higher education.

Elements of interventions intended to support self-assessment can be found in numerous studies. Based on findings from these studies, several authors collected and discussed these elements. Brown et al. (2015) present features that have been shown to improve self-assessment accuracy: clear criteria, models (e.g., comparable work of other students), instruction and practice in self-assessment, feedback on accuracy, rewards, and keeping self-assessments strictly formative. Nielsen (2014) provides a list of strategies for effective implementation of self-assessment methods in writing instruction, e.g., self-assessment training, models, or co-development of criteria. Tai et al. (2018) describe practices common for developing evaluative judgment: self-assessment, peer-feedback/review, feedback (on student's judgments and as dialog), rubrics, and exemplars.

The results of self-assessment studies which used above mentioned intervention elements are providing valuable information, but are in some cases also inconsistent. In a recent meta-analysis, Yan et al. (2022) found that even though the overall effect of SA on academic performance was positive ($g = 0.585$), in 22.79% of the SA interventions negative effects were observed. And also accuracy highly can differ, depending on the creation of optimal conditions while avoiding the many pitfalls (Brown et al., 2015).

Observations on such inconsistencies in the results are not new, and Eva and Regehr (2008) suggested that we should stop addressing questions like "How can we improve self-assessment?," as hundreds of studies have led to the answer "You cannot." In 2019, Andrade confirmed this image in her literature review on student self-assessment (Andrade, 2019), but added:

> "What is not yet clear is why and how self-assessment works. Those of you who like to investigate phenomena that are maddeningly difficult to measure will rejoice to hear that the cognitive and affective mechanisms of self-assessment are the next black box." (Andrade, 2019, p. 10)

In our view, it is valid to continue trying to improve self-assessment. We agree that it is valuable to study the cognitive and affective mechanisms of self-assessment. In addition, we propose that a better understanding of *when*, *why*, and *how* self-assessment interventions work can be gained by a more design-based research approach. This means that empirical educational research should be blended with theory-driven design of learning environments (The

TABLE 1 Definitions of self-assessment.

Source	Definition
Boud and Falchikov (1989, p. 529)	"The involvement of learners in making judgments about their own learning, particularly about their achievements and the outcomes of their learning"
Andrade and Valtcheva (2009, p. 13)	"Self-assessment is a process of formative assessment during which students reflect on the quality of their work, judge the degree to which it reflects explicitly stated goals or criteria, and revise accordingly."
Panadero et al. (2016, p. 804)	Self-assessment "involves a wide variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes or products. This involves retrospective monitoring of previous performance … and reporting, hopefully truthfully, the quality of work completed."

Design-Based Research Collective, 2003). Instead of only looking at the successes or failures of certain intervention elements, we should also focus on the interactions and connections between designed learning environments, processes of enactment, and outcomes of interest (The Design-Based Research Collective, 2003).

Sandoval proposed a technique for design-based research in education called Conjecture Mapping (Sandoval, 2014). This framework intends to specify the "theoretically salient features of a learning environment design and map out how they are predicted to work together to produce desired outcomes" (Sandoval, 2014, p. 19). A conjecture map consists of four core elements:

- some high-level conjecture (s) about how to support a specific kind of learning in a specific context,
- embodiments (or design features) of specific designs in which that conjecture becomes reified,
- mediating processes the embodiments are expected to generate, and
- desired outcomes produced by the mediating processes.

Apart from the high-level conjecture, design conjectures describe how certain design features contribute to the generation of specific mediating processes, while theoretical conjectures describe how mediating processes lead to (or produce) desired outcomes.

If we take for example the provision of a rubric as an often implemented design element in self-assessment interventions, then the pure provision of the rubric does not in itself contribute to a higher accuracy of self-assessment. Students have to learn to distinguish between the different quality levels described in the rubric, to translate the quality descriptions of the rubric to concrete instances of products, and to apply the rubric for self-assessment of their own work. This means that, in order to achieve a higher accuracy, *interaction with the criteria* needs to be generated as a mediating process. A (simplified) conjecture map based on this example is shown in Figure 1. Generating this mediating process likely requires more design features, as otherwise the pure provision of a rubric would be sufficient. On the other hand, the absence of this mediating process is one of the possible explanations for the low accuracy of self-assessments in some studies where a rubric has just been provided to the students without further intervention elements.

Only few studies provide detailed information about the process of self-assessment in educational practice (Andrade, 2019). When we look for elements of self-assessment in the literature, such as the ones in the previous example, then we find some variety of design features, but only a few desired outcomes related to self-assessment

skills (the majority looks at accuracy), and even less information on mediating processes. What we miss is an overview of all these elements. Getting this overview could lead to a model of self-assessment, which can be used as a construct scheme for self-assessment interventions and would support a design-based research approach.

As a basis for such research, this study intends to provide such overview by identifying the most important elements of self-assessment as described in the literature. We address the following research question:

> What are important elements for understanding and fostering effective self-assessment of learning artifacts in higher education?

We will use the elements of conjecture maps as analytical framework, and focus only on elements directly related to–or part of––the self-assessment intervention: which embodiments were implemented, were there descriptions of mediating processes as generated by the embodiments, and which outcomes were produced or desired? When looking at the outcomes, we will only look for outcomes related to self-assessment itself, and not outcomes such as academic performance or the quality of the products.
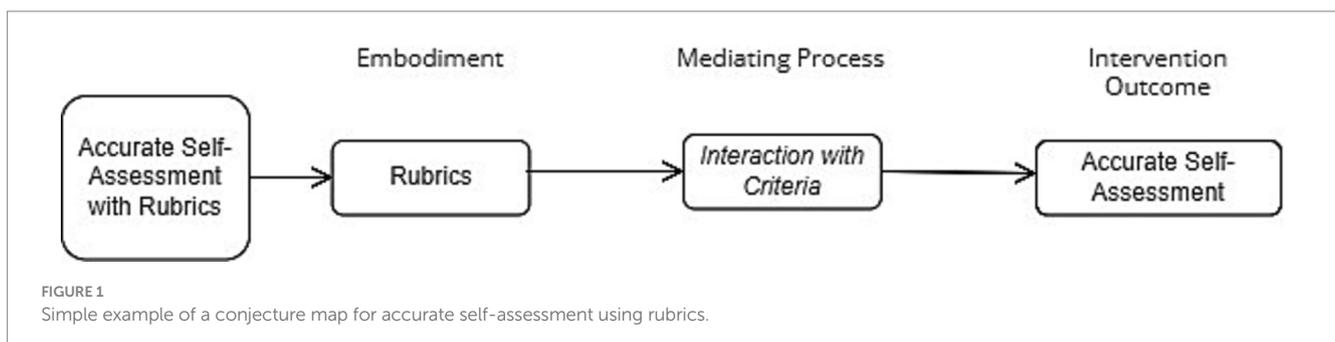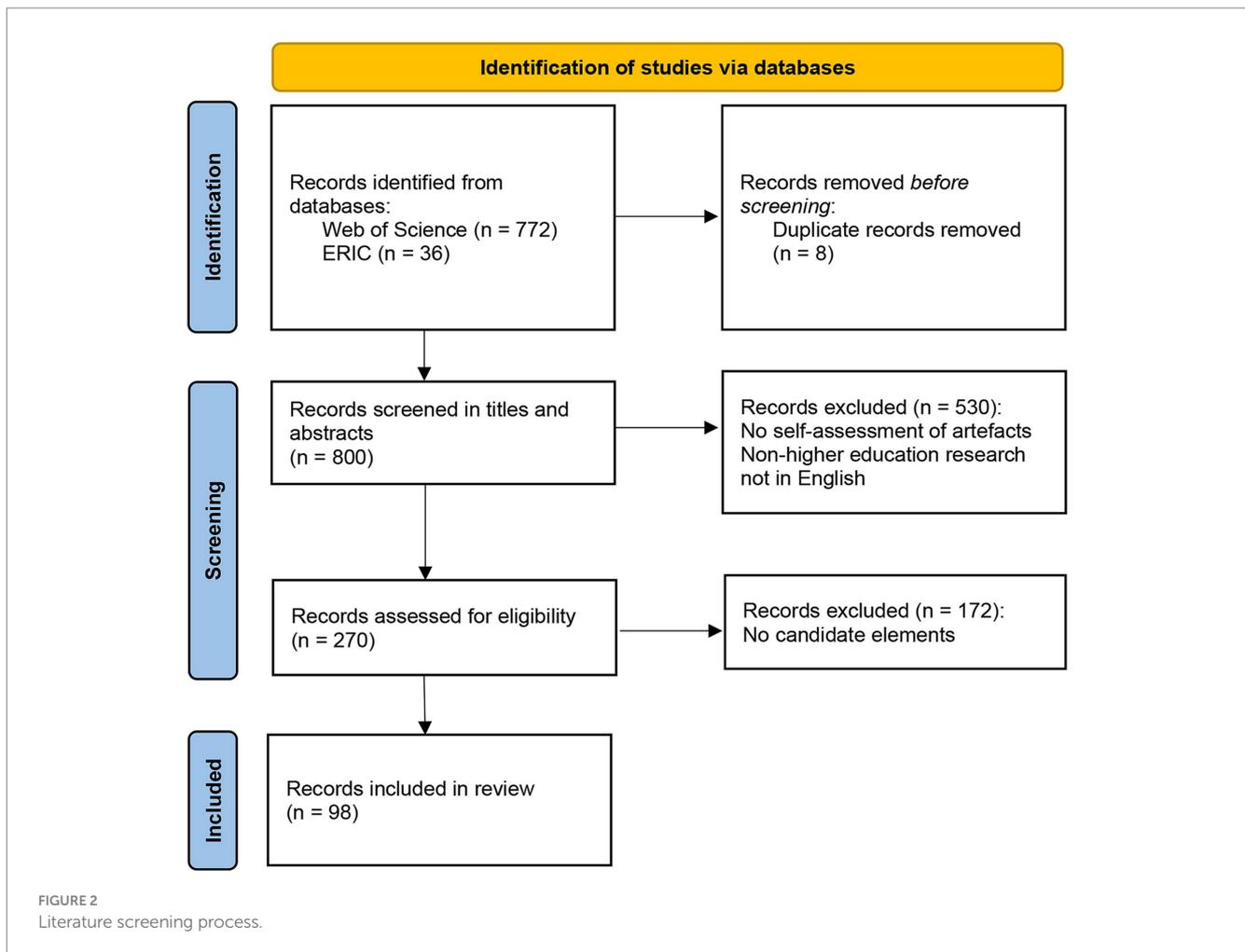
2 Methods

We performed a Rapid Systematic Review (Grant and Booth, 2009) as an alternative to a full systematic review, because we deemed conceptual saturation to be more important than completeness. In our study, the literature search, screening process, analysis, and synthesis were performed by the first author. All results, particularly unclear cases, were discussed and negotiated with the co-authors during all stages of the process.

We included empirical literature presenting qualitative or quantitative accounts of self-assessment of learning artifacts in higher education, as well as theoretical papers in our review, as the aim was to provide an overview of all potentially relevant elements and not a meta-analysis. The literature screening flowchart is shown in Figure 2.

2.1 Search strategy

For our literature search on self-assessment, we used two scientific databases: ERIC and Web of Science. The ERIC database is focused on research in education and is expected to contain publications from educational science addressing self-assessment. However, many

FIGURE 1
Simple example of a conjecture map for accurate self-assessment using rubrics.

FIGURE 2
Literature screening process.

publications are also released in field-specific journals, e.g., self-assessment in language learning or mathematics. To cover such publications, we used as second database the Web of Science Core Collection, as this database indexes a wide range of scholarly journals, books, and proceedings. We included literature from a period of 15 years (publication date between 2006-01-01 and 2021-12-31). As search terms, we used "self-assessment" and "self-grading", as both are used in the literature, in combination with "higher education": "((ALL = (self-assessment)) OR (ALL = (self-grading))) AND ALL = ("higher education")." We did not include search terms related to self-assessment of artifacts/products in this first search in order to also select publications which do not use these terms, but mention only the concrete products, e.g., written essays, assignment solutions etc.

The Web of Science search resulted in 772 publications, and the ERIC educational research database search resulted in 36 publications. After removing the eight duplications, the full initial set contained 800 publications.

2.2 Inclusion criteria

The titles and abstracts of all publications in the full initial set were read to ascertain that they addressed self-assessments of learning

artifacts in higher education. Publications with a context other than higher education (e.g., at workplaces) and publications addressing self-assessment of knowledge or skills that did not require the production of artifacts were excluded. Only publications in English were included in this study. The resulting dataset comprised 270 papers. Furthermore, only papers were included that contained at least one candidate for an embodiment, mediating process, or outcome. This resulted in a set of 92 papers that were used for our data analysis.

2.3 Data analysis

Our data analysis followed the thematical analysis approach (Braun and Clarke, 2006). From the included papers, we collected the publication source details and an extract containing the potential embodiments, processes, and outcomes. Codes were assigned to indicate the type of embodiment, process, or outcome addressed in each paper. In the initial analysis, the potential mediating processes and outcomes were collected in one column because there is often a close connection between them (e.g., understanding quality can be both a process *and* an outcome). The mediating processes and outcomes were distinguished during the qualitative synthesis phase. This initial analysis revealed 451 embodiments and 192 combined mediating processes and outcomes.

The next phase of the thematical analysis (searching for themes) involved inductive analysis of the initial codes to identify candidate themes for embodiments and combined mediating processes and outcomes. The first iteration of the analysis yielded 28 themes for embodiments and 19 themes for combined processes and outcomes.

Candidate themes for both embodiments and processes/outcomes were reviewed and refined in the next phase. This involved a more fine-grained classification of themes according to the types of embodiments, mediating processes, and outcomes, as described by Sandoval (2014). The results of the thematical analysis are presented in section 3.

2.4 Characteristics of included studies

A complete overview of all 92 publications included for analysis is provided in Supplementary Table S1 in Supplementary material. Covered disciplines/fields of study (where mentioned) include educational science and teacher education ($N=21$), English/Language Learning ($N=12$), Biology/Life Sciences ($N=7$), Mathematics ($N=5$), Business ($N=4$), Social Sciences ($N=3$), Accounting ($N=3$), Chemistry ($N=2$), and Engineering ($N=3$). Covered only once are Design, Computer Science, Criminal Justice, Liberal Arts, Health, History, Information Literacy, and Physics. Some studies cover multiple disciplines and some (mostly reviews) do not mention the covered field.

Seventy six studies are empirical and 16 studies are reviews, model buildings or other theoretical work. Of the 76 empirical studies, 31 are quantitative, 16 are qualitative, and 29 mixed qualitative-quantitative. The sample sizes in the studies vary: 2 studies with very small samples ($N<10$), 14 with small samples ($N<30$), 29 with medium samples ($N<100$), and 31 studies with large samples ($N>=100$). Most empirical studies also mentioned the country of data collection with a total of 24 different countries. Countries include Australia ($N=14$), United States ($N=11$), Spain ($N=9$), Great Britain ($N=9$), China ($N=5$), Hong Kong ($N=3$), New Zealand ($N=3$), South Africa ($N=3$) and 1 or 2 times each Taiwan, Finland, Chile, Columbia, Germany, Ireland, Iran, Israel, Lithuania, Mexico, Malaysia, Singapore, Serbia, Thailand, Turkey, and Ukraine. Regarding continent, 24 studies are from Europe, 18 from Asia (incl. Turkey), 17 from Australasia/Oceania, 12 from North-America, 3 from Africa and 2 from South-America.

Sixty-six studies have students as only data source, 2 studies use data from teachers only and 8 studies use data from students and teachers/staff. There are 40+ different types of products/learning artifacts used for self-assessment. The most frequently used ones are essays, (recordings of) oral presentations, scientific reports, and course assignments. The study year (where specified) also varies greatly: 17 studies look at 1st years only, 4 studies look at 2nd years, 14 studies look only at 3rd and/or 4th years, 8 studies look at multiple years and the rest is more unspecific (bachelor students, undergraduate, adult learners etc.). Forty-eight studies used only one source for data collection: 24 used process/performance data (the products, self-assessments and teacher assessments), 13 used surveys, 4 used interviews, 2 focus groups and 1 observation. All other studies triangulated with multiple sources (various combinations, double counts possible): 23 studies combined surveys with other sources, 12 studies combined interviews with other sources, 29 combined process/performance data with other sources, and nine studies used three or more sources for data collection.

3 Results

Our review revealed a total of 26 elements deemed relevant for fostering and understanding self-assessment of learning artifacts. These elements comprise 13 embodiments and six mediating processes, which can lead to seven desired outcomes specifically focused on self-assessment of learning artifacts. Together they form a model which describes self-assessment and can be used as construct scheme for self-assessment interventions and for research into the how and why self-assessment works. In this section, we will first present embodiment, mediating process, and outcome individually, sorted by their number of occurrences. Then the complete model will be presented.

3.1 Embodiments

Embodiments are the concrete elements of educational designs. They are classified and aggregated according to four types, following Sandoval (2014): *Tools and Materials* (software programs, instruments, manipulable materials, media, and other resources), *Task Structures* (the structure of the tasks learners are expected to do—their goals, criteria, standards, and so on), *Participant Structures* (how students and teachers are expected to participate in tasks, the roles and responsibilities participants take on), and *Discursive Practices* (practices of communication and discussion or simply ways of talking). A categorized summary of all embodiments is provided in Table 2.

3.1.1 Tools and materials

Analytic *rubrics* were the most frequently mentioned tool in the literature on self-assessment of learning artifacts (63 publications). They are used to define pre-set assessment criteria for (elements of) learning products, including descriptions of several quality levels per criterion. Analytic rubrics are useful for feedback (Sadler, 2009) and, consequently, for formative self-assessment. Rubrics must be complete, clear, and transparent to guide students' self-assessments (Fastré et al., 2012; Tai et al., 2018).

While rubrics define assessment criteria and several levels of quality, *exemplars* (31 publications) illustrate how these dimensions of quality manifest in concrete work products (Sadler, 1989; Smyth and Carless, 2021). Exemplars can be either authentic student samples, representing a certain quality level, or teacher-constructed examples that make specific features visible to students (Smyth and Carless, 2021). Several publications recommend using a range of exemplars of different qualities (e.g., Jones et al., 2017; Knight et al., 2019). Additionally, assessments of exemplars or comments on their quality can be provided to students as models for their self-assessments (Andrade and Valtcheva, 2009; Fastré et al., 2012).

Only 11 of the reviewed publications provided a description of concrete *self-assessment instruments*, the actual tools to be used while performing self-assessment. The instruments mentioned mostly include generic ones such as assessment sheets/forms (Taras, 2015; Hung, 2019).

TABLE 2 Overview of embodiments found in the reviewed literature, incl. number of publications and short description.

Embodiment	#	Description
Tools and materials		
Rubrics	63	A scoring guide for evaluating the quality of students' artifacts containing evaluative criteria, quality definitions, and a scoring strategy.
Exemplars	31	Samples of work which enable students to understand assessment expectations and illustrate dimensions of quality.
Self-assessment instruments	11	Tools and materials used for actual performance of self-assessment, such as scoring sheets, online forms, or computer-assisted instruments.
Task structures		
Assessment training	44	Explicit training of how to perform an assessment of relevant work using defined criteria. Often includes explanation of how assessment procedures work.
Required self-assessment	22	Students have to perform at least one self-assessment, ideally of work in progress.
Shared grading	21	Shared responsibility of grading - at least partially - with students. This makes self-assessments more relevant for students.
Revision/improvement possibility	18	Provide the possibility of students using self-assessment results for improving or revising their work.
Incentive for accurate self-assessment	15	Make the quality of self-assessments matter for the students by providing incentives that stimulate the accuracy of their self-assessments.
Assessment result justification	13	Students have to provide an explicit justification of their assessment instead of having them only rate the quality without further explanation.
Participant structures		
Teacher feedback on self-assessment results	28	Students get timely feedback on their self-assessments too, not only on their work.
Peer assessments	13	Students assess each other's work provide/receive feedback.
Discursive practices		
Co-creation of criteria	23	Students are involved in the creation or adaptation of assessment criteria (rubrics).
Assessment result discussion	8	Results of self or peer assessments are discussed with peers and/or teachers.

Few publications explicitly mention the technical instruments or digital systems used for self-assessment. Besides standard learning management systems such as Blackboard (e.g., Diefes-Dux, 2019), some papers also report on newly developed systems for the purpose of supporting self-assessment. Li and Zhang (2021) developed and used a computer-assisted adaptive instrument and Agost et al. (2021) introduced computer-based adaptable resources. Both include elements such as rubric annotations, complementary resources, and selectable levels of provided details. Lawson et al. (2012) described the review system designed to facilitate self-assessment.

Despite these reports, our review revealed that relatively little attention has been given to the technical support of self-assessment and how certain instruments can facilitate self-assessment. Further research in this area is warranted.

3.1.2 Task structures

Assessment training was explicitly mentioned as a way to increase the quality of students' self-assessments (44 publications). Some papers only mention that students receive training, without further specification of the kind of training, while others provide a more detailed description of training elements, for example, analysis and discussion of criteria (Lavrysh, 2016; Hung, 2019), discussing assessment results with student peers (Chen, 2008), using multiple exemplars of varying quality to help students develop an appropriate sense of what makes good quality before introducing the rubrics (Smyth and Carless, 2021), or discussing with students how to interpret the quality levels (Tai et al., 2018). Some reviewed studies

explicitly included an explanation of how the assessment process works (self-assessment and final assessment).

Twenty-two publications included at least one *required self-assessment of work-in-progress*. Initial self-assessments are often inaccurate, as they may be biased by (unintentional) self-deception and prior academic success (Buckelew et al., 2013). Required self-assessments of work-in-progress help to discover and address these issues. Additionally, students may perceive self-assessment as beneficial to them when they experience how it helps them close the gap between their current and desired performance.

Thirteen publications included students having to provide an explicit *assessment result justification* that obligates them to substantiate the quality level of (components of) an artifact and how it meets the assessment criteria. This stimulates higher-level cognitive skills (evaluation and analysis) and helps develop an understanding of what quality means (Tai et al., 2018). Some papers include assessment justification also for peer assessments or as part of assessment training. Explicit justification by students also provides information for teachers about the quality of their self-assessment and can serve as input for feedback on how students' self-assessment might be improved (Tai et al., 2018).

Fifteen publications described an *incentive for accurate self-assessment* to stimulate serious and/or accurate self-assessment. In most cases, an accurate self-assessment is counted as a (small) part of the grade or resulted in students receiving extra credits (Cabedo and Maset-Llaudes, 2020). Contrary to rewarding accuracy, some publications mention punishments for inaccurate self-assessments,

such as lower grades or loss of credits (Knight et al., 2019; Seifert and Feliks, 2019). Sometimes, completing a self-assessment itself was rewarded, independent of its accuracy (Davey, 2015; Wanner and Palmer, 2018).

Twenty-one publications reported on *shared responsibility of grading* at least partially between teachers and students to make self-assessment more valuable to students (Bourke, 2018) and to involve them actively in the assessment and grading process. This also benefits the teacher-student relationship (Edwards, 2007). Students reported putting more effort into self-assessment when graded (Jackson and Murff, 2011). Known problems of self-grading, such as grade inflation or social response bias (e.g., Brown et al., 2015), can be prevented by grade negotiation with the assessor (McDonnell and Curtis, 2014; Seifert and Feliks, 2019) and combining self-grading with assessment training or requiring justification of the grade by students (Evans, 2013; Bourke, 2018).

The relevance of self-assessments increases if there is a *revision/ improvement possibility* (18 publications) to close the gap between the actual level of performance and the desired quality standard (Andrade and Valtcheva, 2009). Students can use their self-assessments of draft versions for revisions/improvements of the work before their final submission (Taras, 2015; Wanner and Palmer, 2018). Nielsen (2014) emphasized that it is important to provide sufficient time for revision.

3.1.3 Participant structures

The importance of *teacher feedback on self-assessment results* has been emphasized in 28 publications (e.g., Andrade and Valtcheva, 2009). It should focus on both the quality and accuracy of students' judgments and not only on the quality of students' work (Sitzmann et al., 2010; Tai et al., 2018) and should be given on time. If students provide justifications of their self-assessment, feedback can be tailored to their needs. Ideally, teachers can see whether students really understood the quality criteria and whether they were able to translate and use them for concrete evaluations. Reasons for over-assessment and under-assessment (such as unintentional self-deception or other biases) could be identified and addressed in the feedback (Fastré et al., 2012).

Peer assessment was used in 13 publications to improve self-assessment skills (Bozzkurt, 2020). Assessing the work of peers requires students to apply quality criteria to some work similar to theirs and to justify their assessment. If students have to perform assessments on the works of several peers, they are exposed to different implementations and levels of quality. This helps them gain a better understanding of the quality criteria and how they manifest in concrete artifacts. Getting their own work assessed by peers enables students to calibrate their own assessment with that of others and potentially develop more insights into multiple interpretations of quality criteria and their manifestation in artifacts.

3.1.4 Discursive practices

Involving students in *co-creation of rubrics/criteria* leads to a shared understanding of these criteria and increased ownership (23 publications, e.g., Birjandi and Hadidi Tamjid, 2012; Nielsen, 2014). During the process of creation, students are involved in the social construction and articulation of standards, which helps them with evaluative judgment in new fields and contexts (Tai et al., 2018). Students can also analyze and discuss exemplars to identify quality dimensions and criteria (Smyth and Carless, 2021). Boud and Falchikov (1989) stated that besides rating one's own work, the identification of criteria or standards applied to one's work is another key element of self-assessment.

An *assessment result discussion* with peers and/or teachers increases the understanding of how to accurately self-assess (eight publications, e.g., Lavrysh, 2016). Such discussions can be held after self-assessment, peer assessment, and teacher assessment, and can serve both as feedback and feedforward (Mannion, 2021). These are valuable collective calibration practices (Brown et al., 2015).

3.2 Mediating processes

Identifying the mediating processes helps to understand how and when embodiments do (not) contribute to achieving the desired outcomes. It also helps to clarify why some studies relate positive outcomes to certain embodiments (e.g., rubrics and assessment training), while other studies do not. According to Sandoval (2014), mediating processes manifest as either observable interactions or artifacts that function as proxies for learning processes, indicating the extent of learner engagement in relevant activities. Our review shows that mediating processes are often not explicitly described, although they are likely to have been generated by the embodiments. The mediating processes found in the literature are summarized in Table 3 and described in more detail below.

TABLE 3 Overview of mediating processes found in the reviewed literature, incl. number of publications and short description.

Mediating process	#	Description
Interactions		
Interaction with criteria	3	Students actively engage with criteria and standards.
Assessment of other work	12	Students assess exemplars or the work of others, ideally of varying quality.
Regular self-assessments	14	Students perform regular self-assessments of their work in progress for generating self-feedback and use this for improvement.
Assessment dialog	11	Elements of assessments are discussed between student and teacher, but also among students.
Self-reflection	6	Self-reflection on self-assessments to create awareness and improvement of performance.
Artifacts		
Visible progress (through repeated SA)	2	Students produce artifacts of increasing quality as results of repeated self-assessments.
Regular self-assessments	14	Students document the results of their regular self-assessments.

3.2.1 Interactions

Students should perform *interaction with criteria* in order to understand what they mean and how to apply them to concrete products (3 publications, e.g., Boud et al., 2013; Bird and Yucel, 2015). This requires cognitive engagement, for example, by discussing the criteria (Cowan, 2010) or deconstructing a rubric (Jones et al., 2017).

The practice of *assessment of other work* is often described as essential for improving self-assessment skills (12 publications, e.g., Chen, 2008). Examining and interacting with work similar to what is expected from students serves as an instantiation and representation of good (or varying) quality (Andrade and Du, 2007; Bird and Yucel, 2015). Such similar work can be the products of peers or exemplars provided by the teacher. As the students are not personally attached to the artifacts to be assessed, the risk of biased assessment is much lower.

Students who perform *regular self-assessments* generate formative self-feedback on work-in-progress (drafts) and use it to inform revisions and improvements (14 publications, e.g., Andrade and Valtcheva, 2009). These self-assessments can be obligated (through embodiments) but might also be performed by students' choice as part of monitoring their own performance (e.g., Chen, 2008; Cowan, 2010).

Assessment dialogs about different elements of assessments are key to supporting students' self-assessments (11 publications, e.g., Nielsen, 2014). Communication between teachers and students (and between students) about self-assessment should be reciprocal and focus on learning how to self-assess. Allowing a dialog on assessment tools (Cockett and Jackson, 2018) or assessment results (Lavrysh, 2016) can be considered a democratic strategy, which is often valued by students (McDonnell and Curtis, 2014) and enhances their receptiveness to feedback (Henderson et al., 2019).

Self-reflection on self-assessments (six publications, e.g., Brown et al., 2015; To and Panadero, 2019) means that students become aware of not only the quality of their work products, but also the quality of their self-assessment performance. Therefore, such self-reflection is an important process for identifying strategies for closing the gap between the actual performance of self-assessments and the desired performance (correct and complete self-assessments). These strategies can be applied to improve self-assessment skills.

3.2.2 Artifacts

The different drafts of the learning artifacts present *visible progress through repeated self-assessments* (two publications). This means that self-assessments that indicate deficiencies in these artifacts can lead to traceable improvements. The learning process generated by self-assessments manifests in the progress of learning artifacts (Hung, 2019; Xiang et al., 2021).

The results of the *regular self-assessments* (14 publications) may also be documented as explicit artifacts by the students, which indicate the learner's engagement with the quality of their work-in-progress.

Our review revealed seven mediating processes (counting regular self-assessments twice, both as interaction and artifact) that may be generated by the embodiments described in the previous section. However, most publications do not explicitly pay attention to mediating processes, and further research is needed to validate these processes and identify other potential processes that contribute to achieving the outcomes, as described in the next section.

3.3 Outcomes

The mediating processes, generated by embodiments, contribute to the production of desired outcomes. Although most of the literature on self-assessment focused primarily on high accuracy of self-assessments as outcome, in our review we also found other relevant outcomes.

Some outcomes are explicitly stated as such (e.g., judgment calibration or focus on quality improvement), whereas other expected outcomes remained implicit. For example, an item in a questionnaire such as "Would you like to see self-assessment applied in other courses too" indicates the outcome that students perceive self-assessments as valuable to them. All desired outcomes identified in the literature review, both explicit and implicit, are summarized in Table 4.

The outcome directly related to accuracy of self-assessments is *judgment calibration* (19 publications, e.g., Bozzkurt, 2020), meaning that student assessments are mostly consistent with external assessments, such as that of the teacher (Table 4).

Achieving a *low impact of self-representation* is one of the implicit outcomes (seven publications). Self-assessments are often influenced by self-representation and personality and not by real levels of achievement (Jansen et al., 1998). Unintentional self-deception, often present in low-achieving students, can lead to over-assessment (Agost et al., 2021). Other student characteristics, such as gender and cultural features, have also been reported to influence self-representations that impact self-assessments (González-Betancor et al., 2019; Carroll, 2020).

Gaining a *quality understanding* of relevant standards seems to be a key outcome of self-assessments, as it is the outcome mentioned the most (27 publications). The levels range from a more generic understanding of quality (Kearney, 2013; Carless and

TABLE 4 Overview of outcomes found in the reviewed literature, incl. number of publications and short description.

Outcome	#	Description
Judgment calibration	19	Student assessments are (mostly) in line with external assessments.
Low impact of self-representation	7	Students' self-representation does not have much impact on their self-assessments.
Quality understanding	27	Students understand relevant quality aspects of expected standards, defined as assessment criteria.
Self-assessment appreciation	10	Students understand the value of self-assessment and appreciate it.
Students' assessment ownership	16	Students understand the assessment process and are motivated and self-directed in assessing their work.
Quality improvement focus	22	Students focus on improving the quality of learning.
Self-efficacy growth	5	Students are more confident that they can deliver high-quality work.

Chan, 2017; Scott, 2017) to distinguishing between low- and high-quality work (Lavrysh, 2016; Seifert and Feliks, 2019) or recognition of average but sufficient pieces of work (Taras, 2015). Some publications have focused on an expected standard for concrete artifacts that students should understand (McDonnell and Curtis, 2014; Bird and Yucel, 2015; Adachi et al., 2018). Sadler (1989) emphasized that students should understand that not all quality aspects can always be unambiguously formulated, that criteria may require interpretations, or that there is appraisal of work as a whole and not only of the parts.

The sustainable effect of self-assessment depends on how valuable and beneficial students see its application for their learning progress and the quality improvement of their learning artifacts. *Self-assessment appreciation* is an implicit outcome mentioned in ten publications, helpful for effective self-assessment to occur and leading to students' application of it in following assignments, even though it is not required (Andrade and Valtcheva, 2009). The focus in the literature varies slightly and ranges from understanding *why* self-assessment is beneficial (Yan and Brown, 2017; Adachi et al., 2018) to understanding *how* self-assessment is beneficial (McDonnell and Curtis, 2014; Lavrysh, 2016) to *becoming convinced* of the benefits of self-assessment by experiencing it themselves (Panadero et al., 2016; Deeley and Bovill, 2017).

Students engage more in self-assessment when they assume *assessment ownership* (16 publications, e.g., Adachi et al., 2018). Commitment to, and understanding of, the assessment system helps to develop students' competence in performing accurate and realistic self-assessments (Nielsen, 2014; González-Betancor et al., 2019) and producing better work (McDonnell and Curtis, 2014). A sense of ownership developed by self-assessment activities, such as the co-creation of criteria, motivates students (Nielsen, 2014) and cultivates responsibility and autonomy (Cassidy, 2007).

Having a *focus on quality improvement* helps students identify appropriate actions to close the gap between the actual level of performance and the desired standard (22 publications, e.g., Sadler, 1989; Andrade and Du, 2007). Self-assessments are used to improve work and correct mistakes to get closer to the desired standard (Bourke, 2018).

Five publications reported *self-efficacy growth* regarding judging quality as assessors as a result of self-assessment practices (e.g., Xiang et al., 2021). Some studies have explicitly aimed to build students' confidence through self-assessment activities (e.g., Scott, 2017).

3.4 Model of self-assessment

The elements presented in the previous sections together form a model of self-assessment of learning artifacts in higher education based on the reviewed literature. Figure 3 gives an overview of the model. The elements in the model are structured according to the conjecture mapping approach into embodiments, mediating processes, and (desired) outcomes. The 13 embodiments identified in various self-assessment interventions can be applied in different configurations, depending on the focus of the self-assessment intervention (e.g., formative or summative) or other design decisions (e.g., providing *Assessment Training* versus "training-on-the-job" with multiple *Required Self-Assessments*). Application of the embodiments

can contribute to the generation of the six mediating processes contained in the model. These mediating processes are deemed essential for the production of the desired outcomes and help to understand how and why the overall intervention design works. One mediating process, Regular Self-Assessments, can be categorized as both an interaction and an artifact, it is therefore included twice in the model. The seven desired outcomes described in the model provide an overview of what desirable results are of self-assessment interventions. These outcomes are specifically related to self-assessment itself and not to other (domain-specific) learning outcomes, which may be formulated within the context of the educational intervention.

4 Discussion

What are important elements for understanding and fostering self-assessment (of learning artifacts) in higher education? To answer this question, we performed a rapid systematic review (Grant and Booth, 2009) and used the Conjecture Mapping approach (Sandoval, 2014) as the analytical framework. Using this approach, we created a model linking 13 embodiments, six mediating processes, and seven outcomes, forming an integrated framework for understanding self-assessment, which can also serve as a basis for the design of learning environments that support self-assessment.

In this model, the 13 identified embodiments comprise a variety of design features that were all applied to support the self-assessment of learning artifacts, albeit in various configurations and with sometimes mixed results. All embodiments were applied in successful designs, indicating that they can potentially contribute to effective self-assessment. However, as with any educational intervention, the success of their application depends on context and details of implementation. The main contribution of our model is assisting researchers and practitioners in understanding of and designing for self-assessment by linking the levels of mediating processes and outcomes to the embodiments found.

According to Andrade (2019), research on these processes is essential for understanding *when*, *how* and *why* certain embodiment configurations lead to certain outcomes. However, our review showed only occasional mentions of these mediating processes and hardly any direct relationship with embodiments. Understanding how self-assessment can be successfully implemented requires identifying potential mediating processes and explaining how certain outcomes are related to specific (combinations of) embodiments.

The outcomes described in this paper cover various aspects of what should be achieved in order to make self-assessment effective. In addition to the outcomes explicitly mentioned in the literature on self-assessment, such as accuracy of self-assessments, we identified more implicit outcomes such as self-assessment appreciation or quality understanding. These outcomes are often student-centered and can be regarded as important when it comes to engaging students not only in comparing their work with the standard that is required, but also in appropriate *action* to improve their work toward that standard (Sadler, 1989) also when self-assessment is no longer required by the teacher. A more specific finding of our review is that relatively little research has been conducted on the impact of technological support on the effectiveness of self-assessments (the technical self-assessment
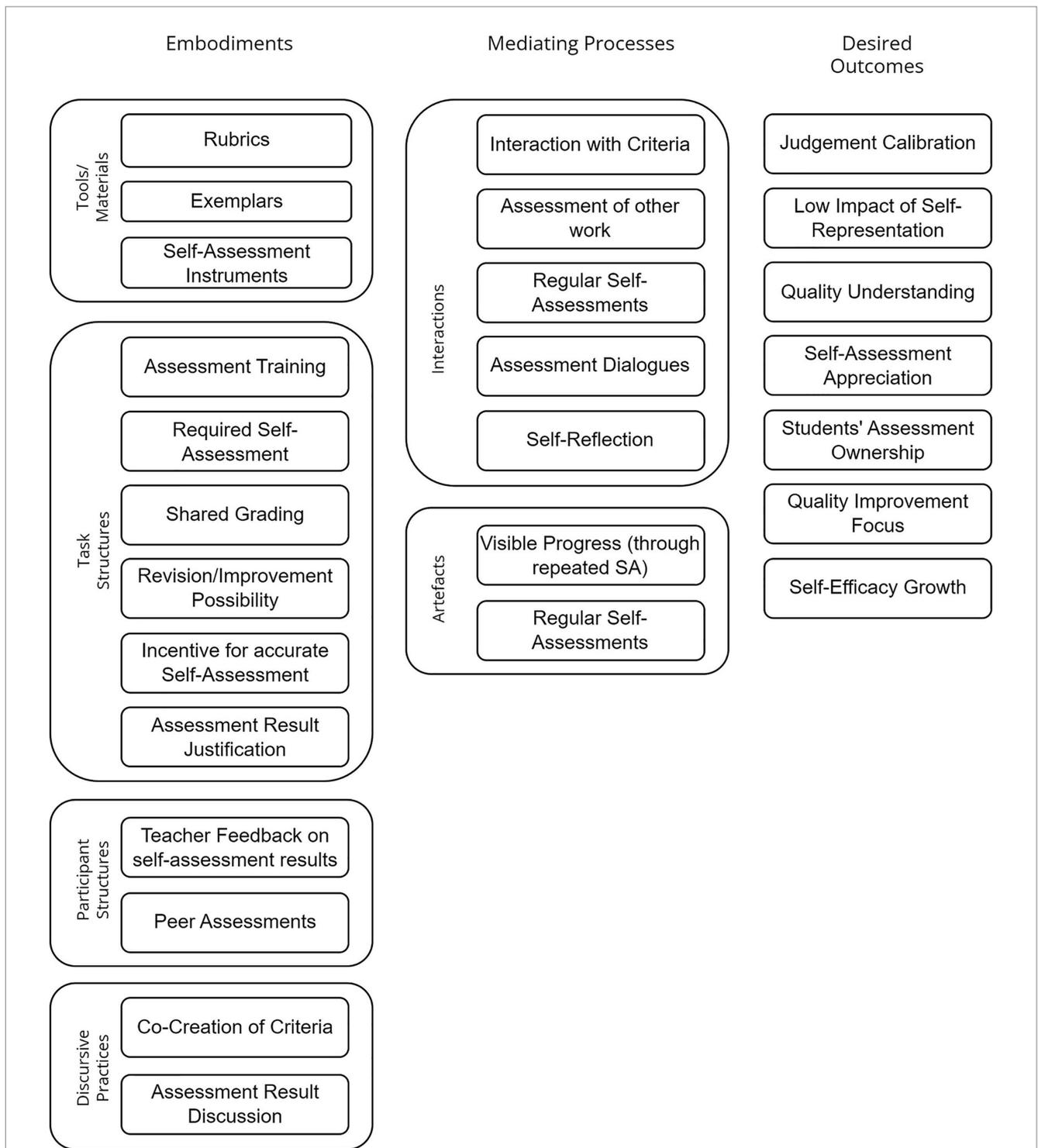
FIGURE 3
Model of self-assessment based on results of our study.

instruments). In addition, the mediating processes generated in educational designs with students' self-assessment of their work seem underexplored in the literature.

The framework is not prescriptive in the sense that it can be used to determine what kinds of self-assessment embodiments should be present in order to achieve specific outcomes. Instead, it provides an interpretative structure for understanding existing practices and

for creating designs that explicitly link design features to expected outcomes, mediated by learning processes. The embodiments, processes and outcomes resulting from our review become building blocks for designers to make and underpin their design choices, in line with Sandoval et al.'s original ideas behind conjecture mapping.

A very relevant question in this field is what improving SA means, and the presented framework can help with an answer to this question.

We propose the effectiveness of SA as level of achievement of one or more of the seven outcomes. For improving SA we argue that outcomes deemed relevant in a specific context should be improved. As argued above, the framework can be used as construct scheme for this. Following the constructive alignment approach (Biggs and Tang, 2011), these should be aligned with teaching and learning activities which can be designed using the elements of the presented model.

4.1 Limitations

The research method we applied was a Rapid Systematic Review (Grant and Booth, 2009) combined with a Thematical Analysis (Braun and Clarke, 2006). We used Conjecture Mapping as framework for the thematical analysis (Sandoval, 2014) aiming for the conceptual saturation of the identified elements. A consequence is that we might have missed publications with additional elements. This means that we cannot consider the collections of elements, processes and outcomes to be final or complete. For the model as a whole this has no consequences, as it will be robust to the addition of more elements, but such additions may be expected as a consequence of future studies.

Studies included in our review were from a variety of fields and countries and covered SA of different types of artifacts. Even though we did not evaluate per element in which field/country/artifact type it was applied/successful (as this was not our goal), we assume that our model principally is applicable for all types of artifacts in various fields. Evaluation of this assumption is subject to future work.

4.2 Future research

This model can be applied to at least three future research directions. All three can contribute to answering questions regarding *how, when,* and *why* specific embodiments contribute to the achievement of desired outcomes through certain mediating processes and consequently promote an understanding of successful self-assessment processes.

4.2.1 Research direction 1: new case studies

For educational design research, new case studies can use this framework for the design of self-assessment interventions and research on the effectiveness of these interventions. Figure 4 shows an example of a conjecture map based on our results. The four embodiments can be implemented and studied to determine to what extent they generate the three mediating processes and produce the intended outcomes.

4.2.2 Research direction 2: mapping onto existing research

Mapping elements of the framework as potential conjectures onto existing research can help to gain more insight into the complex phenomenon of self-assessment. For example, consider the following publication: *Student self-assessment: Results from a research study in a level IV elective course in an accredited bachelor of chemical engineering* (Davey, 2015).

The following embodiments can be discerned in the educational design of the study: assessment rubrics (provided in advance of course starting), required self-assessment (one after submission of assignment), shared grading (self-assessment counts for 10 % of the final grade), example (as idealized solution, provided after assignment submission), and a marking sheet. The course with these embodiments was reported to have the following self-assessment related outcomes: Accuracy was not high (self-assessments were 16% higher than tutor assessments), only 50% of participants valued self-assessment, and only 50% were confident that their sell-assessment was correct. From our conjecture mapping perspective, we note that this study lacks any mention of mediating processes. This information has been added to the conjecture map results in Figure 5.

It is likely that the mediating processes required to produce the outcomes were insufficiently generated. The following suggestions are formulated based on the proposed framework:

The necessary mediating processes that contribute to the production of the three outcomes are *Regular Self-Assessments*, *Assessment Dialogs*, *Self-Reflection*, and *Visible Progress (through repeated SA)*. To generate these processes, the following embodiments can be adapted or added: the *Required Self-Assessment* should not be done after the deadline, but
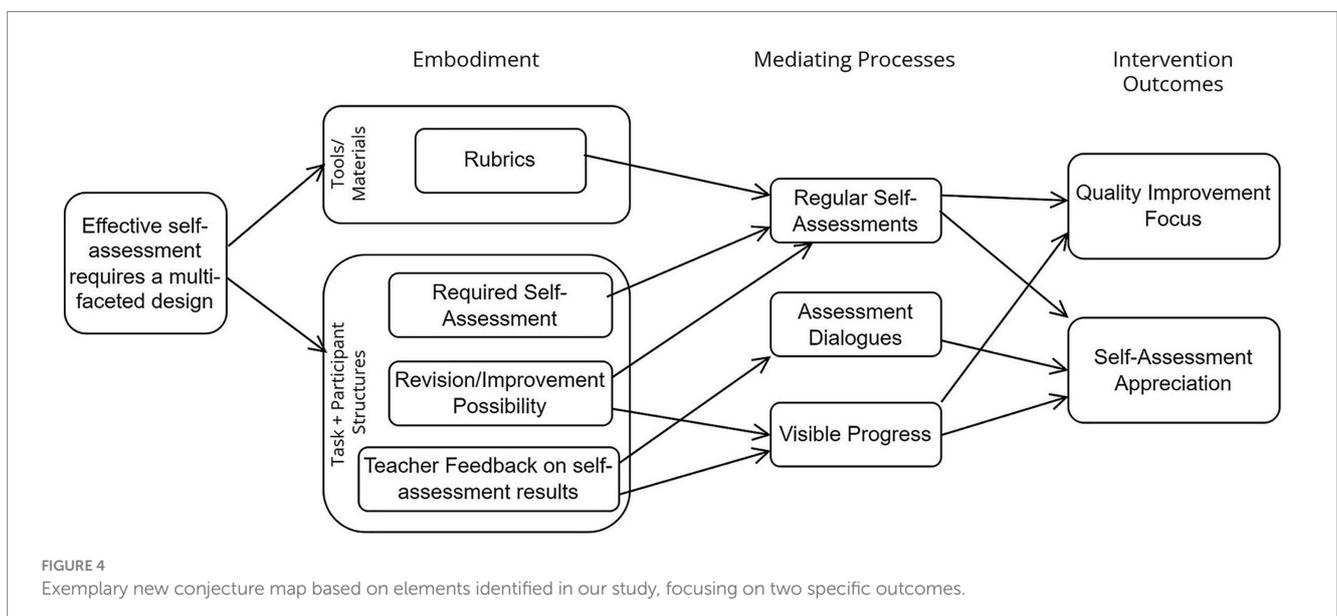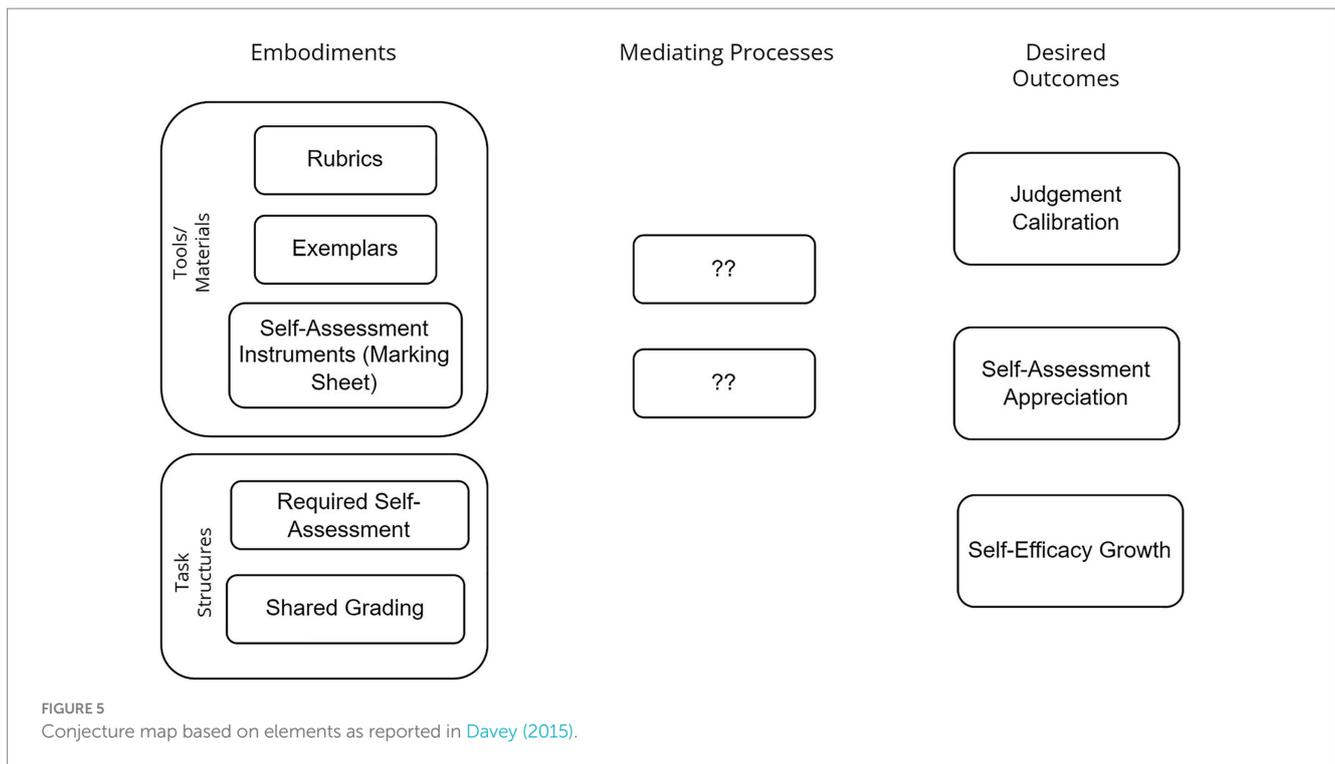

FIGURE 4
Exemplary new conjecture map based on elements identified in our study, focusing on two specific outcomes.

FIGURE 5
Conjecture map based on elements as reported in Davey (2015).

also required earlier (and optionally more than once) so that the results can be discussed and used for improvement (*Revision/Improvement Possibility*). Students should also be encouraged to perform self-assessments even if they are not required. *Assessment Training* would help students learn how to interpret and apply the rubrics, ideally using multiple *Exemplars* of varying quality (not only an idealized solution). There also should be *Teacher Feedback on Self-Assessment Results*.

4.2.3 Research direction 3: exploration of existing educational designs

A third research direction that we propose is more explorative than design-based. One could look for educational designs in which multiple embodiments were implemented and use our framework for analyzing the effects of these embodiments: which of the mediating processes were generated and which outcomes were produced by them. An example could be a software engineering project in which students work on several artifacts, such as software requirements documents, software design documents, program source code, and test reports. In such projects, rubrics often describe the quality criteria for all dimensions of the various artifacts. Such educational designs offer a rich research context for studying self-assessment: Students could be asked to regularly perform self-assessments of their work-in-progress, and data could be collected and analyzed to address the effects of implemented self-assessment elements.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1213108/full#supplementary-material

References

Adachi, C., Tai, J. H.-M., and Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assess. Eval. High. Educ.* 43, 294–306. doi: 10.1080/02602938.2017.1339775

Agost, M.-J., Company, P., Contero, M., and Camba, J. D. (2021). CAD training for digital product quality: a formative approach with computer-based adaptable resources for self-assessment. *Int. J. Technol. Des. Educ.* 32, 1393–1411. doi: 10.1007/s10798-020-09651-5

Andrade, H. (2019). A critical review of research on student self-assessment. *Front. Educ.* 4:87. doi: 10.3389/feduc.2019.00087

Andrade, H., and Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assess. Eval. High. Educ.* 32, 159–181. doi: 10.1080/02602930600801928

Andrade, H., and Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Pract.* 48, 12–19. doi: 10.1080/00405840802577544

Atchoarena, D. (2021) 'Universities as lifelong learning institutions: a new frontier for higher education?', in H van't Land, A. Corcoran and D.-C. Iancu. (eds) *The Promise of Higher Education*. Cham: Springer International Publishing, pp. 311–319.

Biggs, J., and Tang, C. (2011). *Teaching for quality learning at university*. 4th Edn Open University Press.

Bird, F. L., and Yucel, R. (2015). Feedback codes and action plans: building the capacity of first-year students to apply feedback to a scientific report. *Assess. Eval. High. Educ.* 40, 508–527. doi: 10.1080/02602938.2014.924476

Birjandi, P., and Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assess. Eval. High. Educ.* 37, 513–533. doi: 10.1080/02602938.2010.549204

Boud, D., and Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *High. Educ.* 18, 529–549. doi: 10.1007/BF00138746

Boud, D., Lawson, R., and Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assess. Eval. High. Educ.* 38, 941–956. doi: 10.1080/02602938.2013.769198

Bourke, R. (2018). Self-assessment to incite learning in higher education: developing ontological awareness. *Assess. Eval. High. Educ.* 43, 827–839. doi: 10.1080/02602938.2017.1411881

Bozzkurt, F. (2020). Teacher candidates' views on self and peer assessment as a tool for student development. *Aust. J. Teach. Educ.* 45, 47–60. doi: 10.14221/ajte.2020v45n1.4

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706QP063OA

Brown, G. T. L., Andrade, H. L., and Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assess. Educ. Princ. Pol. Pract.* 22, 444–457. doi: 10.1080/0969594X.2014.996523

Buckelew, S. P., Byrd, N., Key, C. W., Thornton, J., and Merwin, M. M. (2013). Illusions of a good grade. *Teach. Psychol.* 40, 134–138. doi: 10.1177/0098628312475034

Burgess, H., Baldwin, M., Dalrymple, J., and Thomas, J. (1999). Developing self-assessment in social work education. *Soc. Work Educ.* 18, 133–146. doi: 10.1080/02615479911220141

Cabedo, J. D., and Maset-Llaudes, A. (2020). How a formative self-assessment programme positively influenced examination performance in financial mathematics. *Innov. Educ. Teach. Int.* 57, 680–690. doi: 10.1080/14703297.2019.1647267

Carless, D., and Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assess. Eval. High. Educ.* 42, 930–941. doi: 10.1080/02602938.2016.1211246

Carroll, D. (2020). Observations of student accuracy in criteria-based self-assessment. *Assess. Eval. High. Educ.* 45, 1088–1105. doi: 10.1080/02602938.2020.1727411

Cassidy, S. (2007). Assessing "inexperienced" students' ability to self-assess: exploring links with learning style and academic personal control. *Assess. Eval. High. Educ.* 32, 313–330. doi: 10.1080/02602930600896704

Chen, Y.-M. (2008). Learning to self-assess oral performance in English: a longitudinal case study. *Lang. Teach. Res.* 12, 235–262. doi: 10.1177/1362168807086293

Cherner, T. S., and Kokopeli, E. M. (2018). 'Using Web 2.0 tools to start a webquest renaissance', in A. A. Khan and S. Umair (eds) *Handbook of Research on Mobile Devises and Smart Gadgets in K-12 Education*. Hershey, PA, pp. 134–148.

Cockett, A., and Jackson, C. (2018). The use of assessment rubrics to enhance feedback in higher education: an integrative literature review. *Nurse Educ. Today* 69, 8–13. doi: 10.1016/j.nedt.2018.06.022

Cowan, J. (2010). Developing the ability for making evaluative judgements. *Teach. High. Educ.* 15, 323–334. doi: 10.1080/13562510903560036

Davey, K. R. (2015). Student self-assessment: results from a research study in a level IV elective course in an accredited bachelor of chemical engineering. *Educ. Chem. Eng.* 10, 20–32. doi: 10.1016/j.ece.2014.10.001

Deeley, S. J., and Bovill, C. (2017). Staff student partnership in assessment: enhancing assessment literacy through democratic practices. *Assess. Eval. High. Educ.* 42, 463–477. doi: 10.1080/02602938.2015.1126551

Diefes-Dux, H. A. (2019). Student self-reported use of standards-based grading resources and feedback. *Eur. J. Eng. Educ.* 44, 838–849. doi: 10.1080/03043797.2018.1483896

Dochy, F., Segers, M., and Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Stud. High. Educ.* 24, 331–350. doi: 10.1080/03075079912331379935

Edwards, N. M. (2007). Student self-grading in social statistics. *Coll. Teach.* 55, 72–76. doi: 10.3200/CTCH.55.2.72-76

Eva, K. W., and Regehr, G. (2008). I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*, 28, 14–19. doi: 10.1002/chp.150

Evans, C. (2013). Making sense of assessment feedback in higher education. *Rev. Educ. Res.* 83, 70–120. doi: 10.3102/0034654312474350

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., and van Merriënboer, J. J. G. (2012). Drawing students' attention to relevant assessment criteria: effects on self-assessment skills and performance. *J. Vocat. Educ. Train.* 64, 185–198. doi: 10.1080/13636820.2011.630537

González-Betancor, S. M., Bolívar-Cruz, A., and Verano-Tacoronte, D. (2019). Self-assessment accuracy in higher education: the influence of gender and performance of university students. *Act. Learn. High. Educ.* 20, 101–114. doi: 10.1177/1469787417735604

Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf. Libr. J.* 26, 91–108. doi: 10.1111/j.1471-1842.2009.00848.x

Han, C., and Riazi, M. (2018). The accuracy of student self-assessments of English-Chinese bidirectional interpretation: a longitudinal quantitative study. *Assessment and Evaluation in Higher Education*, 43, 386–398. doi: 10.1080/02602938.2017.1353062

Henderson, M., Ryan, T., and Phillips, M. (2019). The challenges of feedback in higher education. *Assess. Eval. High. Educ.* 44, 1237–1252. doi: 10.1080/02602938.2019.1599815

Hung, Y. (2019). Bridging assessment and achievement: repeated practice of self-assessment in college English classes in Taiwan. *Assess. Eval. High. Educ.* 44, 1191–1208. doi: 10.1080/02602938.2019.1584783

Jackson, S. C., and Murff, E. J. T. (2011). Effectively teaching self-assessment: preparing the dental hygiene student to provide quality care. *J. Dent. Educ.* 75, 169–179. doi: 10.1002/j.0022-0337.2011.75.2.tb05034.x

Jansen, J. J. M., Grol, R. P. T. M., Crebolder, H. F. J. M., Rethans, J. J., and van der Vleuten, C. P. M. (1998). Failure of feedback to enhance self-assessment skills of general practitioners. *Teach. Learn. Med.* 10, 145–151. doi: 10.1207/S15328015TLM1003_4

Jones, L., Allen, B., Dunn, P., and Brooker, L. (2017). Demystifying the rubric: a five-step pedagogy to improve student understanding and utilisation of marking criteria. *High. Educ. Res. Dev.* 36, 129–142. doi: 10.1080/07294360.2016.1177000

Kearney, S. (2013). Improving engagement: the use of "authentic self-and peer-assessment for learning" to enhance the student learning experience. *Assess. Eval. High. Educ.* 38, 875–891. doi: 10.1080/02602938.2012.751963

Knight, S., Leigh, A., Davila, Y. C., Martin, L. J., and Krix, D. W. (2019). Calibrating assessment literacy through benchmarking tasks. *Assess. Eval. High. Educ.* 44, 1121–1132. doi: 10.1080/02602938.2019.1570483

Lavrysh, Y. (2016). Peer and self-assessment at ESP classes: CASE study. *Adv. Educ.*, 60–68. doi: 10.20535/2410-8286.85351

Lawson, R. J., Taylor, T. L., Thompson, D. G., Simpson, L., Freeman, M., Treleaven, L., et al. (2012). Engaging with graduate attributes through encouraging accurate student self-assessment. *Asian Soc. Sci.* 8, 291–305. doi: 10.5539/ass.v8n4p3

Li, M., and Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Lang. Test.* 38, 189–218. doi: 10.1177/0265532220932481

Mannion, J. (2021). Beyond the grade: the planning, formative and summative (PFS) model of self-assessment for higher education. *Assess. Eval. High. Educ.* 47, 411–423. doi: 10.1080/02602938.2021.1922874

McDonnell, J., and Curtis, W. (2014). Making space for democracy through assessment and feedback in higher education: thoughts from an action research project in education studies. *Assess. Eval. High. Educ.* 39, 932–948. doi: 10.1080/02602938.2013.879284

Nicol, D., and MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199–218. doi: 10.1080/03075070600572090

Nielsen, K. (2014). Self-assessment methods in writing instruction: a conceptual framework, successful practices and essential strategies. *J. Res. Read.* 37, 1–16. doi: 10.1111/j.1467-9817.2012.01533.x

Nieminen, J. H., and Tuohilampi, L. (2020). "Finally studying for myself" – examining student agency in summative and formative self-assessment models', *Assessment and Evaluation in Higher Education*, 45, 1031–1045. doi: 10.1080/02602938.2020.1720595

Panadero, E., Alonso-Tapia, J., and Reche, E. (2013). 'Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers', *Studies in Educational Evaluation*, 39, 125–132. doi: 10.1016/j.stueduc.2013.04.001

Panadero, E., Brown, G. T. L., and Strijbos, J.-W. (2016). The future of student self-assessment: a review of known unknowns and potential directions. *Educ. Psychol. Rev.* 28, 803–830. doi: 10.1007/s10648-015-9350-2

Panadero, E., García-Pérez, D., Ruiz, J. F., Fraile, J., Sánchez-Iglesias, I., and Brown, G. T. L. (2023). 'Feedback and year level effects on university students' self-efficacy and emotions during self-assessment: positive impact of rubrics vs. instructor feedback', *Educational Psychology*, 43, 756–779. doi: 10.1080/01443410.2023.2254015

Panadero, E., Jonsson, A., and Botella, J. (2017). 'Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses', *Educational Research Review*, 22, 74–98. doi: 10.1016/j.edurev.2017.08.004

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instruc. Sci.* 18, 119–144. doi: 10.1007/BF00117714

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assess. Eval. High. Educ.* 34, 159–179. doi: 10.1080/02602930801956059

Sandoval, W. (2014). Conjecture mapping: an approach to systematic educational design research. *J. Learn. Sci.* 23, 18–36. doi: 10.1080/10508406.2013.778204

Scott, G. W. (2017). Active engagement with assessment and feedback can improve group-work outcomes and boost student confidence. *High. Educ. Pedag.* 2, 1–13. doi: 10.1080/23752696.2017.1307692

Seifert, T., and Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assess. Eval. High. Educ.* 44, 169–185. doi: 10.1080/02602938.2018.1487023

Sitzmann, T., Ely, K., Brown, K. G., and Bauer, K. N. (2010). Self-assessment of knowledge: a cognitive learning or affective measure? *Acad. Manag. Learn. Educ.* 9, 169–191. doi: 10.5465/amle.9.2.zqr169

Smyth, P., and Carless, D. (2021). Theorising how teachers manage the use of exemplars: towards mediated learning from exemplars. *Assess. Eval. High. Educ.* 46, 393–406. doi: 10.1080/02602938.2020.1781785

Tai, J., Ajjawi, R., Boud, D., Dawson, P., and Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *High. Educ.* 76, 467–481. doi: 10.1007/s10734-017-0220-3

Taranto, D., and Buchanan, M. T. (2020). 'Sustaining lifelong learning: a self-regulated learning (SRL) approach', *Discourse and Communication for Sustainable Education*, 11, 5–15. doi: 10.2478/dcse-2020-0002

Taras, M. (2015). Autoevaluación del estudiante: ¿qué hemos aprendido y cuáles son los desafíos? *Relieve-Revista Electrónica de Investigación y Evaluación Educativa* 21. doi: 10.7203/relieve.21.1.6394

The Design-Based Research Collective (2003). 'Design-based research: an emerging paradigm for educational inquiry', *Educational Researcher*, 32, 5–8. doi: 10.3102/0013189X032001005

To, J., and Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assess. Eval. High. Educ.* 44, 920–932. doi: 10.1080/02602938.2018.1548559

Wanner, T., and Palmer, E. (2018). Formative self-and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback. *Assess. Eval. High. Educ.* 43, 1032–1047. doi: 10.1080/02602938.2018.1427698

Xiang, X., Yuan, R., and Yu, B. (2021). Implementing assessment as learning in the L2 writing classroom: a Chinese case. *Assess. Eval. High. Educ.* 47, 727–741. doi: 10.1080/02602938.2021.1965539

Yan, Z., and Brown, G. T. L. (2017). A cyclical self-assessment process: towards a model of how students engage in self-assessment. *Assess. Eval. High. Educ.* 42, 1247–1262. doi: 10.1080/02602938.2016.1260091

Yan, Z., Lao, H., Panadero, E., and Fernández-Castilla, B. (2022). 'Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis', *Educational Research Review*, 37, 1–15. doi: 10.1016/j.edurev.2022.100484