The Use of Digital Peer Assessment in Higher Education
An Umbrella Review of Literature

van Helden, G.; van der Werf, V.; Saunders-Smits, Gillian; Specht, M.M.

DOI
10.1109/ACCESS.2023.3252914

Publication date
2023

Document Version
Final published version

Published in
IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE EDUCATION SOCIETY SECTION

Received 11 January 2023, accepted 1 March 2023, date of publication 6 March 2023, date of current version 10 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3252914

TOPICAL REVIEW

The Use of Digital Peer Assessment in Higher Education–An Umbrella Review of Literature

GITTE VAN HELDEN[1,2,3], VIVIAN VAN DER WERF[1,4], GILLIAN N. SAUNDERS-SMITS[1,5], AND MARCUS M. SPECHT[1,3,4,6]

[1]Leiden-Delft-Erasmus Centre for Education and Learning, 2628 XE Delft, The Netherlands
[2]Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands
[3]4TU Centre for Engineering Education, Delft University of Technology, 2628 XE Delft, The Netherlands
[4]Leiden Institute of Advanced Computer Science, Leiden University, 2333 CA Leiden, The Netherlands
[5]Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands
[6]Faculty of Electrical, Engineering, Mathematics, and Computer Science, Delft University of Technology, 2628 XE Delft, The Netherlands

Corresponding author: Gitte van Helden (G.vanHelden@tudelft.nl)

This work was supported in part by the Erasmus+ Program—Key Action 2 (KA2)—Cooperation for Innovation and the Exchange of Good Practice and Key Action 226 (KA226)—Partnerships for Digital Education Readiness as part of the Relevant Assessment and Pedagogies for Inclusive Digital Education (RAPIDE) Project, Relevant Assessment and Pedagogies for Inclusive Digital Education.

ABSTRACT Increasing student numbers in higher education, particularly in engineering and computer science, make it difficult for motivated lecturers to continue engaging in active teaching methods such as Flipped Classrooms and Work-Based Learning. In these settings, digital Peer Assessment can be one approach to provide effective and scalable feedback. In Peer Assessment, students assess each other's performance whilst gaining useful reflection and judgment skills at the same time. This umbrella review of 14 review papers on the use of (digital) Peer Assessment in education provides a comprehensive overview of design choices and their consequences open to educational practitioners wishing to implement digital Peer Assessment in their courses, the type of tooling available and the possible effects of these choices on the learning outcomes as well as potential pitfalls and challenges when implementing Peer Assessment. The paper will inform and assist educators in finding or developing a tool that fits their needs.

INDEX TERMS Digital education, educational technology, engineering education, peer assessment.

I. INTRODUCTION

Assessment, and in particular formative assessment for learning, plays a growing role in higher education. Its crucial role for learning skills and competencies is widely recognized by scholars in the field of education [1]. Within the field of engineering and computer science education this trend is also very visible [2]. However, there are challenges for lecturers when it comes to providing high-quality, effective, and scalable assessments, even more so within the fields of engineering and computer science education. Engineering and computer science student populations across the world are growing fast [3]. Educational institutions need to accommodate the increasing number of students, often without an equal increase in resources, such as teaching staff. Also, the more complex learning tasks, such as integrated systems design and research projects that integrate different competencies in engineering education and are often carried out in groups, make it challenging for lecturers to provide students with appropriate assessment and feedback. Furthermore, as the recent COVID-19 pandemic has shown, lecturers must also be able to switch delivery from face-to-face to online and hybrid learning and back again at short notice, ideally without any loss of quality in the learning experience for students [4], [5]. As a result, time-intensive processes, such as assessing students' progress and providing students with much-needed feedback, must be reshaped to maintain a manageable workload. One possible solution to aid with the assessment of students is to engage the students themselves in this process [6]. This concept is widely known as Peer Assessment (PA) and, in one form or another, has been in use in education for a long time, although it may

The associate editor coordinating the review of this manuscript and approving it for publication was Rebecca Strachan.

not have been described as such in literature. The first person to conduct a formal literature review on the use of peers for assessment purposes in higher education was Topping [7]. His definition of PA as ''an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status'' [7, p.250] has been widely accepted and is still being used today.

PA has proven itself to be a useful strategy for lowering lecturers' burden [8] while simultaneously having a positive impact on students' learning, allowing students to not only deepen their understanding of the material studied but also to develop transversal skills such as receiving feedback, forming a judgment, and self-reflection [9]. PA holds the potential to be used on a large scale, but on the condition that lecturers have access to effective and reliable tools and procedures [10]. Developments over the past 20 years in the wide-spread implementation of digital tools in education have facilitated the possibility to engage in peer assessment activities in larger and remote or online groups, because these tools provide (better) opportunities to collect, organize, and analyze information on a large scale and make digital PA accessible to larger, potentially less digitally literate audiences [11], [12]. Moreover, as opposed to face-to-face activities, digital PA allows for students to engage a-synchronous from different places [13].

This work is part of the Erasmus+ project Relevant Assessment and Pedagogies for Inclusive Digital Education (RAPIDE). The aim of the project is to co-create, implement, and share innovative pedagogies and aligned assessments for relevant and inclusive digital education to deal with the COVID-19 induced and similar crises, and to support the meaningful digital transformation of Higher Education Institutions (HEIs). Involved partners are the Faculty of Organization and Informatics, University of Zagreb, Croatia (coordinator); Delft University of Technology, the Netherlands; Goethe University, Germany; the School of Medicine, University of Zagreb, Croatia; the Open University, United Kingdom; and the University of Rijeka, Croatia.

The focus of the project is on the development of tool-kits and digital pedagogies for assessment approaches, including PA, for practitioners. For this, it is essential to collect an overview of insights from existing literature that are both evidence-based and useful for educational practice. A quick initial online search on literature reviews on PA resulted in a multitude of literature reviews on PA and digital PA. Many of these reviews appeared to focus on a different aspects within the field of PA, such as the implementation of PA in a specific domain (for example, teacher education [14] or nurse education [15], the effect of a design variation (for example, anonymity [16]) the combination of PA with specific educational practices (for example, gamification [17]). These literature reviews derive insights from relevant empirical papers and move from interventions that work in a specific context to more generalized conclusions. Whereas such results are relevant for the scientific community, it often does not provide comprehensive, practical guidelines for practitioners who want to implement digital PA in their education, which, given the earlier mentioned required flexibility and capacity issues, is a must. The lack of comprehensive, practical guidelines is even more noticeable in the field of engineering and computer science education, where courses vary greatly in delivery: from theoretical lectures with accompanying coursework to hands-on computer programming assignments, and from writing essays to team-based design and research projects. Even though student numbers in engineering and computer science education are rising, especially in fields related to big data and artificial intelligence, lecturers still want to be able to give quality feedback and assessment to all their students. Therefore, in the current study, it is intended to overcome this gap between scientific results and practical implications regarding digital PA by presenting an overview of key learning points on digital PA from existing literature reviews. As such, it is hoped that this article can inform educational practice both within the field of engineering and computer science education but also the wider higher education field.

II. THEORETICAL FRAMEWORK

A. THE USE OF PA IN EDUCATION

PA is widely used in a variety of formats, and in almost every academic domain. Although often all formats are referred to as PA, it is important to realize that PA is an umbrella term, which captures a variety of types of PA [16]. In this paper, three distinctly different types of PA are proposed. These distinctions are made as each type has a different function in terms of educational output.

Type 1 is peer review, which entails students reviewing each other's (written) output and giving each other feedback. The recipient of the feedback may or may not have to account for the received feedback in the next iteration of the output created. Examples of outputs are essays, reports but also prototypes or computer code.

Type 2 is peer grading, where students grade each other's work, in a formative or summative way, against a set of given criteria. Examples are grading each other's homework assignments or essays. This type of feedback does not necessarily require students to give detailed feedback, rather the feedback is limited to whether the answer is correct or to what extent the student has delivered what was asked based on the given criteria.

Type 3 is peer evaluation, which means that students evaluate each other in the context of a group process and reflect and give feedback on for instance transversal skills within this process, such as the ability to work together in teams but also on their effort or their intellectual contributions within the context of their assignment.

When it comes to PA, regardless of the type, one might argue that students are taking over the role of the lecturer. However, there are also benefits for students to

be engaged with peer input. When giving or receiving peer assessment, students are practicing making judgments of quality, calibrating their judgment, and improving their self-evaluation skills [1], [18]. Since 2000, ABET has included transversal skills in its accreditation criteria for engineering degrees. Among these transferable skills, the ability of students to reflect on their own performance, the ability to give constructive feedback, and the ability to form judgment are high on the list of preferred skills [19], [20]. Reflection is important and helps students to look back on their experiences and their work and get an understanding of what that means for the future [2], [21]. Reflection in this sense is not only meant on a personal level but also on an output level. Is the deliverable that I created as a student fit for purpose, how could it be done differently next time, and why? All these types of questions belong to reflection.

B. USE OF DIGITAL TOOLS IN PA

The increased use of digital tools in the educational environment have made implementing PA much more feasible, especially when it comes to large courses (>50 students) and courses with remote participants, as is often the case in Work based Learning (WBL) and Flipped Classroom approaches. Even before the COVID-19 pandemic, universities across the Higher Education Landscape have implemented Virtual Learning Environments (VLEs) in their education [22], [23]. Although the intensity of their use by lecturers varies greatly even within a degree program, VLEs often already supply a variety of online tools enabling lecturers to set up a form of PA [24]. Next to that, several digital reflections and assessment tools have become available on the market such as *Spark* [25] *FeedbackFruits* [26], *Scorion* [27], and *CATME* [28]. These tools offer different functionalities that are suited for different types of PA. CATME and Spark support peer evaluation, whereas FeedbackFruits targets peer review. Scorion can be used to facilitate peer review as well as peer grading. Within the context of the RAPIDE project, the focus lies on the use of digital tools for the PA process to enable students from across the Higher Education Spectrum ranging from Remote students, Students in WBL courses, Blended or Flipped Classroom students, as well as the traditional on-campus students and across all types of PA. As the COVID-19 pandemic has proven, having digital tools available that work across all these contexts provides much-needed flexibility when the switching of educational format is still a realistic prospect. It may also encourage staff to create future-proof courses. This article will therefore limit itself to discussing findings, tools, and applications relevant within a digital context.

C. RESEARCH QUESTIONS

In this paper, the aim is to present a set of comprehensive and practical guidelines for lecturers in higher education to inform their decisions on which type of PA is suitable for their chosen educational application and what type of digital tool they can use. These guidelines will be created based on the following four research questions, all pertaining to the higher education domain.

1) What design variations are relevant for designing (digital) PA?
2) What kind of tooling has been used in digital PA?
3) What are the reported effects of PA on the learning outcomes?
4) What are the challenges for the implementation of PA?

III. METHODOLOGY

A. REVIEW METHOD

To answer the research questions, a systematic umbrella review was conducted. An umbrella review is defined as a ''review compiling evidence from multiple reviews into one accessible and usable document'' [29, p.95]. The approach of an umbrella review was chosen as this type of review is intended for looking at broad problems, in this case, the use of peer assessment in education, for which there are multiple approaches and solutions. In an umbrella review, evidence is collected from multiple review papers on the topic and organized in an accessible and practical way, highlighting the different solutions from the different reviews, ready for field use. In the context of this article ''ready for field use'' means usable for lecturers and for future research and development in this area. The systematic review of literature that formed the foundation of this umbrella review was conducted using the recommendations in [29, p.95] by systematically searching for, appraising and synthesizing research evidence, ''often adhering to guidelines on the conduct of a review.'' The protocol used to guide this systematic review is Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [30], which provides researchers with a checklist and process flow chart that encompasses four stages: identification, screening, eligibility and inclusion (Fig.1).

B. SELECTION OF STUDIES

To identify a relevant body of literature, first criteria for exclusion and inclusion were articulated as suggested by [31]. Documents were excluded when 1) no full text was available, 2) no English text was available, 3) the document was not peer-reviewed, 4) the document did not contain a stand-alone literature review, 5) the focus of the document was not on PA, 6) the document did not address higher education settings, and 7) the document was published before 2005. The literature search for inclusion in the umbrella review was limited to literature review articles in journals and conferences that were published over the past 17 years. It was decided to limit the search to 17 years as over the past 25 years the use of PA has become more mainstream in part due to the increased availability and accessibility of digital tools to enable the use of these techniques. Review articles published before 2005 will likely not include articles that discuss these possibilities and will likely not be as relevant in today's higher education practice given the developments

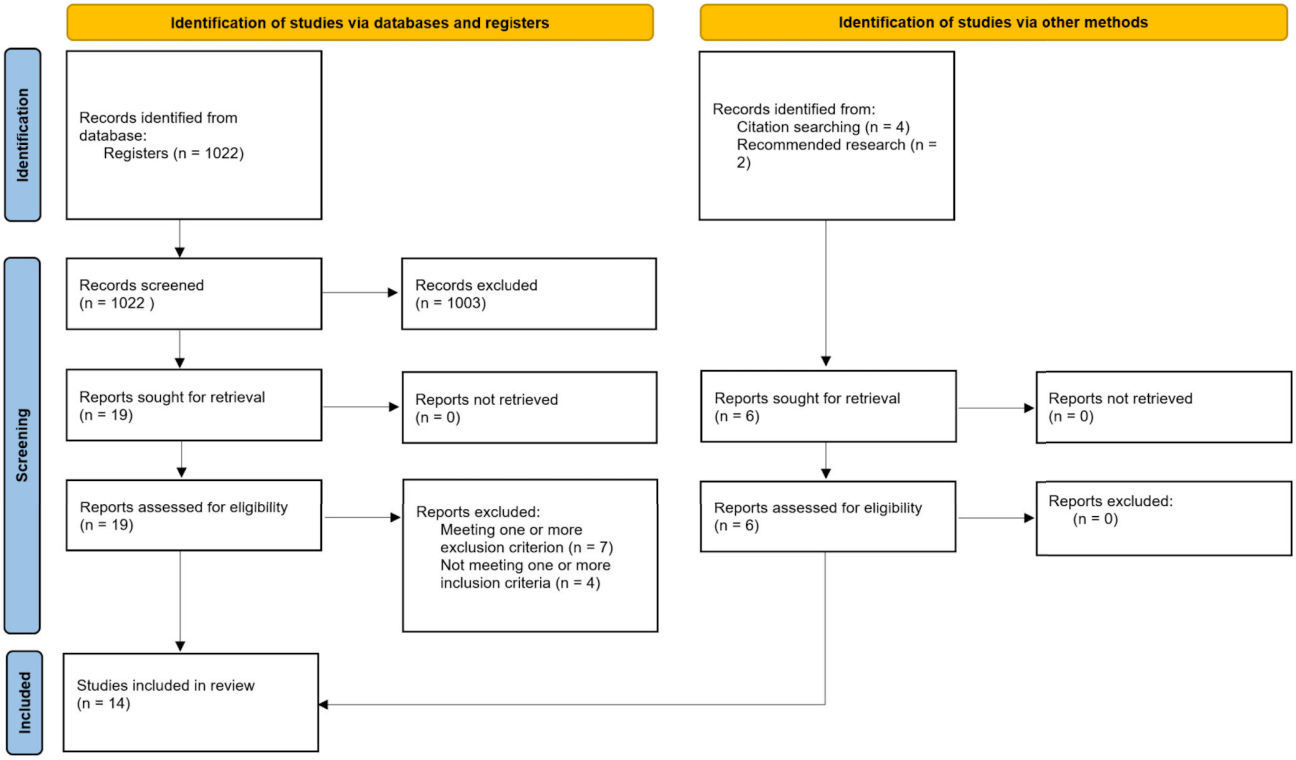

FIGURE 1. PRISMA flow diagram showing all systematic review steps.

in educational technology. To illustrate this point, in [15], the authors searched as far back as 1988 but found no articles on peer assessment matching their inclusion criteria until 1995. It was noted by [7] that literature reviews on PA are still a relatively new phenomenon. In his 1998 article as motivation for writing his widely credited review he states that he cannot find any previous literature reviews on PA. It is therefore unlikely that relevant review articles prior to 2005 will exist. For a review article to be included, it should contribute to answering at least one of the research questions. As such the inclusion criteria that were formulated concern that a document addresses: 1) design variations in the implementation of PA, 2) tools that were used in the implementation of PA, 3) effects of the implementation of PA on learning outcomes, or 4) practical considerations and challenges for the implementation of PA.

The systematic review was carried out using an approach is known as ''citation searching'' [30] or ''snowball sampling'', analogue to human subject research [32]. This approach was chosen as research has shown that just consulting databases such as Scopus yield incomplete results and that in most systematic reviews up to 51% of references in a systematic review are identified by snowballing [33]. Hence, as a starting point, a systematic search was conducted on 28 of September 2022 in a single database, SCOPUS, using the search terms ''peer review'' or ''peer assessment'' and ''literature review'' or ''systematic review'' and education. The search was limited by including only document types ''review'', ''research'' and ''conference paper'' and setting date range starting at 2005. In addition to this, the search was limited to journals with titles that included ''education'', ''educational'', ''learning'', ''teaching'', ''teacher'', ''pedagogy'', ''pedagogical'', or ''instruction''. This decision was made, to increase the likeliness that included research was peer reviewed by experts in the field of education. The identification stage led to an initial body of literature of 1022 articles. For the screening phase, the title, abstract and keywords of all selected articles were screened by the first, second, and fourth authors of this paper. An article was excluded when it met one or more exclusion criteria or did not meet one of the inclusion criteria. This led to a selection of 19 articles that were included in the third phase, in which the full text was screened for eligibility. All articles were discussed by the four authors of this paper. It was decided to exclude 11 articles for not meeting at least one of the criteria for inclusion. This led to a remaining body of 8 articles. In line with the snowballing approach, to reduce the risk of missing articles, the reference lists of the eligible studies were screened to identify additional relevant articles. An additional 4 articles were identified. Finally, the ''recommended research'' section of the websites used to retrieve full-text articles, also led to the identification of 2 relevant articles. The 14 articles included in this umbrella review are listed in Table 1.

TABLE 1. Articles included in umbrella review.

Author	Year	Article type	Review type	Incl. studies	Topic	Country
Tenório *et al.* [8]	2016	Journal	Systematic	44	Technology supported PA	Brazil
Topping [14]	2021	Journal	Systematic	43	Technology supported PA in teacher education	United Kingdom
Tornwall [15]	2018	Journal	Integrative	24	PA in nurse education	USA
Panadero and Alqassab [16]	2019	Journal	Systematic + narrative	14	Anonymity in PA	Spain Hong Kong
Indriasari *et al.* [17]	2020	Journal	Systematic	39	Gamification of PA	New Zealand Indonesia
Topping [34]	2017	Journal	Systematic	27	PA	United Kingdom
Indriasari *et al.* [35]	2020	Journal	Systematic	51	Peer code review	New Zealand Indonesia
Santos-Rosa *et al.* [36]	2016	Conf.	Systematic	12	Technology supported PA	Brazil Portugal
Zheng *et al.* [37]	2019	Journal	Literature	134	Technology supported PA	China Taiwan
Triantafyllou and Timcenko [38]	2014	Conf.	Narrative	Not listed	PA in engineering group project	Denmark
Luxton-Reilly [39]	2009	Journal	Systematic	18	Tools for digital PA	New Zealand
Fu *et al.* [40]	2019	Journal	Systematic	70	Trends in technology supported PA	China Taiwan
van Zundert *et al.* [41]	2010	Journal	Systematic	26	PA	The Netherlands
Lerchenfeldt *et al.* [42]	2019	Journal	Systematic	31	PA in medical education	USA

These 14 papers were all published between 2009 and 2021. Of the 14 articles, 12 are journal articles and two are conference papers. The articles stem from researchers from all over the world covering all continents except for Africa, and a total of 12 unique sets of authors with two authors, [14], [34] and [17], [35] contributing two articles each.

C. QUALITATIVE SYNTHESIS

From an initial analysis in ATLAS.ti, four overarching themes were identified that guided the formulation of the research questions: ''design variations of PA activities'', ''use of tooling in PA activities'', ''effects of PA activities,'' and ''challenges of implication of PA activities.'' All articles were first coded top-down using these four overarching themes by all authors separately. Next, a bottom-up analysis was performed for each of the overarching codes by the first two authors, in which emergent topics were clustered thematically. For example, for design of PA activities, three topics emerged sometimes with subtopics: 1) assessed product or process, 2) assessment criteria and feedback, and 3) assessment procedure. During a final coding round, codes were checked by authors one and two separately, and any issues and uncertainties were jointly resolved and decided upon. See Table 2 for a full overview of themes, topics, and subtopics.

IV. RESULTS

A. DESIGN VARIATIONS FOR PA ACTIVITIES

When designing a PA activity, a number of design choices should be considered [14], [34]. For example, the extent to which design choices will work largely depends on the learning objectives and context of a course [36]. In this section (IV-A), the design choices that were distilled from literature combined with considerations that can be made are outlined.

1) ASSESSED PRODUCT OR PROCESS

A variety of student work can be assessed by peers. PA can be performed on different types of outputs, including written texts (e.g., essays, research proposals, lesson plans), artifacts (e.g., images, audio, video, source code, design product), or an action (e.g., performance, delivery of a lesson for peers, presentations) [14], [17], [34], [37]. However, PA is not only about these products of learning, and can also be focused on the behavior associated with learning processes [34]. For example, students can provide feedback on each other's behavior in a group during collaborative learning situations [38], [42]. Finally, also the assessment itself can be assessed by its receiver. This is called backward feedback or backward evaluation. In this case, the assessee provides feedback to the assessor on the quality of their assessment [17], [39].

What is being assessed highly depends on the domain in which the PA activity is conducted, and the learning outcomes that are associated with this domain. In social sciences, courses tend to be oriented toward writing texts and as a result, students most often assess each other's essays [37]. In engineering and technological sciences, there is an emphasis on design and creation, as such products being assessed are often artifacts such as prototypes, designs (of cars, aircraft, machines, etc.), but also computer code, websites, or videos. Similarly, [17] found that although in most domains text is used as the assessment product, this is different in the fields of physical sciences, mathematics, and computer sciences. Here, as assessed product, text was often combined with an object such as source code. On the other hand, in teacher education, an example of WBL, students

TABLE 2. Themes in PA review papers.

Themes	Topics	Subtopics
A. Design Variations for PA activities	1) Assessed product or process	-
	2) Assessment criteria and feedback	a. Assessment criteria b. Feedback type c. Grading
	3) Assessment procedure	a. Organization and distribution b. Individual vs. collaborative c. Amount of feedback d. Rewards e. Anonymity f. Training & support g. Validity & reliability
B. Digital tools for PA	1) Types of digital tools	-
	2) Functionalities of digital tools	-
	3) Choosing or creating a tool	-
C. Effect on learning outcomes	1) Cognitive	a. Student performance b. Critical thinking & reflection
	2) Affective	a. Group work b. Interest, engagement & motivation c. Attitude
	3) Lecturers	-
D. Challenges	1) Motivation	-
	2) Assessment criteria	-
	3) PA skill	-
	4) Negative social effects	-
	5) Validity & reliability	-
	6) Organizational difficulties	-

engage more often in simulations of teaching practice where PA activities often revolve around a specific action or demonstration of a skill, such as practicing a lesson in front of peers or having peers watch each other's recorded class [14], which is also seen in the field of medical education [42]. Finally, there are PA activities that focus on group processes. This often occurs in fields such as engineering education, in which team-based pedagogies are widely applied [38].

2) ASSESSMENT CRITERIA AND FEEDBACK

Another type of design choice that lecturers must make when implementing PA in their courses concerns creating or providing assessment criteria, the type of feedback desired, and whether students (also) have to grade each other.

a: ASSESSMENT CRITERIA

Lecturers have the choice to provide students with assessment criteria, let students (co-)create the assessment criteria, or provide no assessment criteria at all. Multiple review papers highlight that in almost all cases of PA, assessment criteria were predefined by the lecturer [14], [34], [35], [37], [39], [40]. It is notable that students are rarely invited to contribute to the development of assessment criteria, since it is acknowledged that co-creation of these criteria could lead to more engagement, reduce anxiety, and create a sense of ownership [14], [34]. Moreover, in the few studies that did involve assessors in the creation of assessment criteria, there are indications that this co-creation improves consistency and mutual understanding of assessment criteria [40]. Finally, although it is often said that it is essential to clarify the assessment criteria before letting students engage in PA, there are also some studies in which the advantages of not providing any assessment criteria are stressed [40]. It could create more freedom for students to express their opinion without being restricted by, for example, a rubric. Having no assessment criteria may lead to richer perspectives.

b: FEEDBACK TYPE

In the reviews, three types of feedback can be found that are used in PA: qualitative feedback, quantitative feedback, and a mix of these two [36], [37], [40]. Quantitative feedback can take the form of Boolean criteria (e.g., check box to indicate if certain criteria are present), discrete choices (e.g., choice between a number of specified criteria), or numeric scales (e.g., rating on a 1-10 scale) [37], [39]. Qualitative feedback can be given in a written or oral form [36], [39]. Mixed feedback is usually a combination of both.

Regarding the usefulness of these feedback types, [37] suggest that quantitative feedback might be best suited for summative purposes (grading), while qualitative feedback should be applied for formative purposes (learning). However, findings indicate that a combination of qualitative and quantitative feedback has the greatest positive impact on product quality [37]. The number of studies that have adopted mixed feedback types is increasing over the years [40], suggesting that mixing feedback types also serves a variety of learning objectives and could be applied in many domains.

c: GRADING

Inherent to PA is that students evaluate each other's work. Sometimes these judgments play a solely formative role

and do not affect the official course grade of a student. However, lecturers can also choose to let PA results account for (a percentage of the) course grade. When doing so, it is important to prepare students sufficiently for the role of assessor and take into account ethical factors [36].

3) ASSESSMENT PROCEDURE

A third type of design choice concerns the different aspects of the assessment procedure. This includes choices and considerations about the organization and distribution of feedback, individual vs. collaborative assessment, the amount of feedback (rounds), possible rewards, the anonymity of the assessor or assessee, the presence of prior training or continuous support, and, finally, the validity and reliability of the finalized assessment.

a: ORGANIZATION AND DISTRIBUTION

To perform PA, assessees need to be linked to one or more assessors. This distribution of peers can be done in several ways: students can be randomly allocated, choose their peer feedback partner(s), or be allocated based on student features, such as skill, interest, or quality of feedback [14], [36], [39]. There is no consensus on what type of allocation is the best, as they all hold advantages and disadvantages. Random distribution of students is most used. An advantage is that it can decrease the assessment bias that can emerge [37]. However, random allocation in a group context could lead to the formation of groups that have too many differences between groups [34]. To avoid this, it could be beneficial to allocate students based on their features. Next to this, [34] argues that it is recommended to create pairs with similar abilities so that they can provide feedback on a similar level. This also creates the opportunity for the lecturer to focus on the pairs who are less proficient and need more guidance.

b: INDIVIDUAL VS. COLLABORATIVE

Creating a product or engaging in a process that needs to be assessed as well as assessing a product or a process, can be done by individuals or by groups. Reference [36] suggest that letting assessees work in groups can decrease the workload for lecturers. When lecturers provide an alternative judgment of student work, it might be needed to form groups in large classes of more than 30 students to provide feedback of sufficient quality. Alternatively, when working with pairs in large groups, the lecturer could choose PA as the primary mode of assessment and fulfill a mediating role in the assessment process.

Whereas [36] implies that group work can be used to lower lecturers' burden, we know that working in groups itself can also be a valuable learning opportunity. PA can be leveraged in collaborative courses to lower negative group behaviors, such as free riding [38]. Additionally, it can provide lecturers with useful information about individual group members that otherwise would not be visible. Also, the process of assessing can be performed in a group [35], [37]. This type of collaborative PA can improve the reliability and validity of PA, as students have the opportunity to discuss feedback [37].

c: AMOUNT OF FEEDBACK

Assessees can receive multiple reviews on their work. First, this can be done by letting one product or process be assessed by two or more assessors. Reference [37] recommends the use of multiple assessors, but at the same time warns that the load for students might become too high. Second, multiple rounds of PA can be initiated [14], [37]. Most studies have implemented only one round of PA [34] or did not specifically address the impact of the number of rounds [14]. This is notable, as the number of rounds likely has an influence on the effectiveness of PA [14] and there are some results that indicate that implementing multiple rounds positively affects PA quality [37].

d: REWARDS

Assessors and assessees can be rewarded in various ways. Reference [37] recommends providing rewards to students who are carefully engaged in PA activities. Students can be rewarded for the quality of their product or feedback, the quantity of their product or feedback, or the time they invested in PA [17]. As such, rewards are not only given based on the quality of the work, but also for participation [35]. Examples of rewards can be course credits, points for participation, bonus grades, books, and excursions [37].

One specific form of rewarding students is through gamification, which is successfully used to enhance student engagement [17]. Elements that were used to gamify PA were points, leader boards, badges, progress bars, levels, virtual gifts, and prizes. It was found that gamification was mostly implemented during the act of the assessor providing feedback and the assessee implementing the feedback [17]. However, there are more interactions in the PA process that could potentially be gamified, for example, providing feedback on the assessor or interaction between the assessor and the assessee.

e: ANONYMITY

In PA different levels of anonymity can be used. In an open review process, the identity of both the assessor and the assessee are known. In a single-blind review process, the identity of the assessor is unknown. In a double-blind review process, both the assessor and assessee remain anonymous. It is not always possible to perform a double-blind review process [15]. When assessing a performance, for example, a presentation, it is often not feasible to hide the identity of the assessee.

In research, there is no agreement on whether it is desirable to implement anonymity in PA [15], [16]. Arguments for using anonymity are that it can protect learners' privacy and may reduce a scoring bias [37]. Reference [16] conducted a literature review that was focused on the effect of anonymity in peer review. Although they found that

anonymity (including both single- and double-blind reviews) led to slight advantages in learning performance and quality of the feedback given by peers, they warn educators to be careful when implementing anonymity. Results were based on a very small number of papers. Also, the effects of anonymity on the accuracy of peer grading, social effects, and students' perspectives on the peer assessment activity itself were mixed. In addition to this, there are some unintended negative effects that can be caused by anonymous PA. For example, anonymity could hamper connections in a community and thus decrease opportunities to learn from each other. Especially in digital environments, anonymity could lead to anti-social online behavior. Furthermore, anonymity can reduce students' accountability which can negatively affect the quality of their feedback. It is noteworthy that the number of studies addressing non-anonymous PA is growing, compared to anonymous PA [40], which might reflect a higher awareness of unintended side-effects of using PA anonymously.

f: TRAINING & SUPPORT

It is not apparent that students are able to perform PA. Rather, judging the quality of work and providing useful feedback are skills that should be gradually developed. This can be done by providing training before implementing PA [15], [34], [35], [42]. Training can increase PA quality and students' positive attitudes toward PA [34], [41]. Training can include talking about expectations, role-playing, examples, and moments for practicing [34], [41], [42]. A specific method that can be used is calibrated peer review, in which students first mark exemplar papers before assessing an actual paper to calibrate their marks [38]. Furthermore, students can also be supported during the PA process. Scaffolds can be implemented in the PA process to guide students in the right direction [34], [37]. Scaffolds can for example be shaped as guidelines or checklists [34]. However, studies that implement scaffolds or other forms of guidance are rare [14], [37].

g: VALIDITY & RELIABILITY

Important questions are to what extent students' judgment reflects the quality of the work of their peers, and whether these judgments are consistent. It is known from research that PA generally is a valid and reliable way of assessment [34], [39]. Social relations among students can be a threat to validity, as enmity or friendship could influence the judgment of an assessor [34]. Generally, an approach that is used towards checking the validity of PA is to compare the judgment of a peer assessor with the judgment of a lecturer [39]. Reliability of PA tends to be lower in courses where the practice is assessed when compared to academic products [34], [42]. A way to increase reliability is by letting students co-create or discuss assessment criteria [34]. Also, training before PA activities and support during the PA process can increase reliability.

B. DIGITAL TOOLS FOR PA

There is an increasing amount of research on the use of technology to support the PA process [40]. The use of digital tools for PA can yield benefits, such as lowering the administrative burden for lecturers [38], [39]. Moreover, their functionalities can support educational design choices that are more difficult to facilitate in a traditional classroom setting, such as setting up a double-blind PA process [39]. In this paragraph, different types of digital tools for PA will first be outlined. Next, an overview of functionalities that can be performed by digital PA systems will be presented.

1) TYPES OF DIGITAL TOOLS

There is a multitude of digital technologies that can be used to support PA, in which an important distinction can be made [36]. First, there is the use of digital technologies that already exist and were developed for other purposes. Examples from the literature are the use of Facebook as a discussion forum [14] or the use of GitHub for code review [35]. Research on the use of mobile devices for PA is scarce [14], [37], [40]. This is notable because mobile devices are widely adopted in educational settings and beyond, are relatively low cost, and can be easily combined with other tools such as apps or social media [14], [37]. Reference [40] does report on an increase in research on the use of mobile devices. When using existing tools, it is important to keep in mind the appropriateness of the tool with regards to context. For example, not in each context information can, or should, be shared on social media such as Facebook [14].

Second, there are tools that are specifically designed for the purpose of facilitating PA, which mostly are shaped as integrated learning and learning management systems. Reference [39] introduced a distinction between three types of PA systems. First, there are generic systems that are designed to support PA activities over a range of disciplines and contexts. The characteristic of these tools is thus that they are highly flexible. The rubrics that are used for these types of tools can be designed by a lecturer, support multiple types of feedback, and often the workflow can be adapted. Second, there are domain-specific systems that are designed to support PA activities in specific domains. Examples are PA systems dedicated to programming or writing research reports. As the scope of these PA systems is narrower, this type of system is less flexible. Often the rubrics and the workflow are already integrated into the system and cannot be adapted by an educator. Third, there are context-specific systems that are developed and tested in one specific context and have the potential to be applied in a broader range of contexts. However, for this purpose, the context-specific systems would have to be rewritten and further developed. Again, this type of system is less flexible, for example, by having a fixed rubric and a fixed workflow. Moreover, this type of tool generally supports fewer types of feedback.

2) FUNCTIONALITIES OF DIGITAL TOOLS

Digital tools enable educators to design the PA process in a way that is not possible or more difficult in a traditional face-to-face classroom setting. First, digital PA tools can provide a wide range of design choices with regard to anonymity [37], [39]. Whereas it is almost impossible to organize a double-blind reviewing process in a traditional classroom, many PA tools can ensure anonymity during the entire process. Some tools provide different modes of anonymity, including double-blind, single-blind, pseudonym, or open reviewing.

Second, PA tools can facilitate the process of distributing student work to peer assessors in many ways [37], [39]. Most tools have the opportunity to randomly allocate student work. However, tools can also have additional features, such as distributing student work between and within groups [39]. Additionally, some tools can distribute student work based on the quality of the work. Often, this happens in review processes where one assessor provides feedback on multiple products. Based on earlier rounds of review the tool allocates a judgment of quality to students' work, for example good, intermediate, and poor. Next, it is made sure that each assessor receives work within all categories, so students can see examples of good and bad work and make a comparison. Finally, there are cases in which unrestricted viewing is supported. This means students can make choices on which and how many products they would want to assess.

Third, PA tools can support the design of assessment [39]. Assessment criteria can be predefined in the PA tool, or the option can be provided for an educator to adapt the criteria. Although PA tools could also be leveraged to let students (co-)design assessment criteria, this feature is rarely embedded in PA systems [39].

Fourth, PA tools can support one or more forms of feedback. Examples include Boolean criteria, discrete choice, numeric scales, open-ended feedback, and color-coded highlights in the text [36], [39].

Fifth, multiple ways to assess or increase the quality of the feedback can be implemented in PA tools [39]. This can be done by assessing the validity, for example by comparing the assessment of peers and the instructor or other peers. Another way can be to implement backwards feedback so that the recipient of feedback assesses the quality of the feedback that he or she was given.

Sixth, PA tools can enable a dialogue between the assessee and the assessor. This can be done through the exchange of private messages or the use of an open discussion forum [39]. An advantage of digital tools is that dialogue is not limited by time or space [36] and that a dialogue can take place while the identity of the assessor and assessee does not have to be revealed.

Finally, PA tools can be used to facilitate a workflow. This workflow can automate processes that would require significant efforts in a traditional classroom setting, such as embedding multiple rounds of revision or the implementation of scripts that assist students in improving their work [39].

3) CHOOSING OR CREATING A TOOL

When lecturers want to use a tool to facilitate the PA process, they can choose to use existing technologies that are not designed for PA, an existing dedicated PA tool, or develop a PA tool themselves. When making this decision, the focus should be on the functionalities that the tools to be implemented should bring. Although PA tools open up many possibilities, most of them stick to basic functionalities such as the opportunity for grading or giving comments. PA tools with more advanced functionalities, such as enabling a dialogue, are rarer [37]. Additionally, [17] notes that gamification is mostly focused on two parts of the PA process: creating a product and providing feedback. The reason that other actions are not gamified, is that lectures adopt existing tools that do not necessarily support those actions. This can be solved by adopting or creating a dedicated PA tool that supports a broad range of functionalities. It is notable that most PA tools developed so far originate from the field of computer sciences [39]. This is not a surprise, as this field has the technical expertise to develop digital PA tools.

C. EFFECT ON LEARNING OUTCOMES

PA is widely implemented and tested. In this section, an overview will be provided of the effects that are associated with PA. Although effects are generally positive, difficulties that can emerge during PA. The authors observed that not all claims on the effects of learning outcomes were backed up by solid evidence. This remains an area for further development.

1) COGNITIVE

a: STUDENT PERFORMANCE

Multiple reviews report on positive effects of PA on student performance. This includes improvements in general performance [8], [34], domain-related skills [41], academic writing [8], skills related to programming [35], and professional competencies [34], [42]. Moreover, the quality of products that were produced by students increased [35].

b: CRITICAL THINKING & REFLECTION

Evidence was found for PA on improvement in judgment skills [8], [15] and a better understanding of assessment criteria [38]. Also, there is evidence that the quality of feedback that was given by peer assessors improved [35], [42]. However, [41] found that PA skills mostly improved when training was involved.

2) AFFECTIVE

a: GROUP WORK

There are indications that PA promotes collaboration and communication skills, promotes team development, leads to higher group satisfaction, decreases conflict within a group, reduces tensions during intercultural collaboration, and elicits

more positive attitudes about collaboration [8], [42]. Also, it is efficiently used to decrease the number of free riders [38]. However, some negative effects are associated with PA, as it has been found that it can lead to a more negative and judgmental atmosphere during group projects [15].

b: INTEREST, ENGAGEMENT & MOTIVATION

It was found that students usually are engaged in PA activities, perceive PA as a valuable learning experience and feel part of a learning community [35]. Additionally, there are indications that PA can increase interest in reading [8].

c: ATTITUDE

Furthermore, PA was associated with positive effects on student attitude and responsibility [34], involvement and responsibility in learning [36], and ownership of the learning process [38].

3) LECTURERS

There is evidence that PA holds benefits for lecturers as well. This includes a decrease in workload, more efficient evaluation processes and better manageability of administrative tasks [8], [35], [38].

D. CHALLENGES

Implementation of digital PA does not always emerge without difficulties. In the literature reviews included in this umbrella review, several challenges were reported.

1) MOTIVATION

There were challenges reported that related to motivational aspects of the students that were involved [8], [35], [41]. Low motivation can for example translate into an inability to stay on task. The implementation of approaches that stimulate students' engagement and motivation could solve or prevent such problems [8]. An example is the implementation of gamification in PA, which has been shown to lead to higher student engagement [17]. Another approach is letting students discuss or engage in the creation of assessment criteria, as this could increase engagement and a sense of ownership [14], [34].

2) ASSESSMENT CRITERIA

Other difficulties were related to the creation of good assessment criteria. For example, the evaluation of specific assessment criteria can be difficult [8]. Again, involving students in the creation of assessment criteria can be a possible solution, as it can lead to a better mutual understanding of the criteria [40].

3) PA SKILL

Next to this, there can be difficulties with students' ability to provide feedback [8], [35], [38] as well as to learn from feedback [40]. For example, it is possible that peers provide each other with feedback that is incorrect and possibly misleading [8] or feel under-equipped to provide others with feedback [38]. On the other hand, there are examples of students that did not conduct sufficient self-reflection to improve their work based on the feedback they received [40]. To overcome this, attention should be paid to developing skills on how to deliver good feedback and how to use feedback to improve. This can be done by providing training [34], [41] or other forms of guidance such as scaffolds [34], [37].

4) NEGATIVE SOCIAL EFFECTS

In addition, technological assistance is not necessarily sufficient to decrease negative social effects, such as anxiety, which often exist in traditional PA [40]. Negative social effects are potentially even more dangerous in team-based projects, in which a constructive and non-judgmental atmosphere is required [38]. A possible approach to reduce these social challenges is to implement non-anonymous PA with face to face interactions combined with appropriate training and practice [16].

5) VALIDITY & RELIABILITY

Social processes can also negatively influence reliability and validity [34]. For example, when friendships, enmity, or other power bonds influence the feedback that is being given. Training and co-creation of assessment criteria were mentioned as approaches to increase reliability [34].

6) ORGANIZATIONAL DIFFICULTIES

Finally, there can be organizational difficulties. The online PA process can be time-consuming and is more complex to orchestrate than on-paper PA in a traditional classroom [8]. Also, specific factors in the PA process can cause impracticalities, such as the high number of student products that are too excessive to review and insufficient allocated time to perform PA [35]. Additionally, it was reported that lecturers could not always monitor and assist learners due to the high workload of digital PA tools [40]. A solution can be for lecturers to carefully select a tool with functionalities that suit the type of PA they plan to use.

V. CONCLUSION FOR IMPLEMENTATION OF PA IN AN EDUCATIONAL CONTEXT

The design variations listed in the previous section are a summary of the prevalent options listed in the reviews considered in this paper. There are many more design features that can be taken into account. See, for example, [7], [14] for an extensive overview of design features of PA. In this section, the consequences of the preceding results of the umbrella review for educational practice are discussed by answering each of the four research questions posed within the stated purpose of informing the practitioner, the lecturer.

A. RELEVANT DESIGN VARIATIONS FOR PA DESIGN

The first step that must be taken is to decide, based on one's learning outcomes, what will be peer-assessed: text,

artifact, action/skill or process. Lecturers will also have to decide on the type of PA that is fitting: Peer Review, Peer Grading or Peer Evaluation. The context in which the course is taught is relevant in making this choice. Next, decide on whether criteria are needed and if yes, if these are created by the educator or by the assessors and assessees themselves. Also, attention must be paid to the type of feedback that is desired. This is often linked to how PA is used in a course: as summative or formative feedback. When it comes to the organization of PA, it is important to decide the amount of feedback whilst keeping in mind that having multiple reviews (more than one assessor) and multiple reviews rounds leads to better and more reliable results. In addition, consider training your participants in doing PA, as this will also enhance the overall quality. Decide on the need for anonymity, also considering the overall objective of the PA activity, the context of the course and the teaching culture. Finally, to ensure the validity and reliability compare some of the PA outcomes with the instructor's assessment.

B. CHOICE OF TOOLING FOR PA

There is a plethora of digital tooling options available to assist in PA. When making a choice, it is important that the choice is in line with the comfort level of the lecturer and the provided support within the HEI. The more digitally literate they (and their students) are, and the more support is available, the more advanced the tool can be. In addition, when making the choice, it is good to refer to the previous section on design choices. Is the tool able to support the type of PA selected, the desired assessment criteria, etc.? Is the available functionality of the tool in line with the preferred PA design?

Finally, especially in the engineering and computer science domain, Lecturers weigh up before starting to create their own dedicated program, whether effort and usage weigh up against each other, as well as the legacy of the tool with a view to the future. An important principle to keep in mind here is the KISS principle: Keep It Simple and Straightforward. An off-the-shelf tool may offer very similar functionality and save time in the long run.

C. EFFECTS OF PA ON THE LEARNING OUTCOMES

When it comes to the effects of PA on the learning outcomes, only positive outcomes were reported when it comes to student performance and the development of students' critical thinking and reflection skills. The latter skills can even be increased further if students are first offered training in PA.

When used in group work PA has a positive effect on the group process and as a general whole PA also has a positive effect on community building, and students' attitude and responsibility which all are especially important in online or blended educational settings. However, some negative social effects, such as a judgmental atmosphere during group projects, were found.

D. CHALLENGES TO IMPLEMENTATION OF PA

Implementing PA in a digital environment is not without its challenges and is still very much a work-in-progress. Although the principle of PA is widely used, there is little reported research into the effect of specific design choices on learning outcomes. Such a challenge could be resolved by involving educational researchers when using PA in courses thus extending the body of knowledge.

Next, the legacy of the many digital tools, especially the dedicated tools that have been developed, can pose a risk for continuity. This may be a consideration when weighing up whether to develop one's own tool or use existing (commercial or open source) tools.

PA is only as good as the people using it. It is advised that lecturers outweigh possible positive effects, such as increased quality of feedback and motivation, with possible negative effects of, for example, students fearing they are not graded fairly by a ''lazy'' student and the effort of monitoring the quality of feedback to guarantee a safe educational space when doing PA. Especially in group work, be mindful of the influence of social processes and monitor for those effects.

Finally, although hardly any negative effects on the learning outcomes have been reported, do keep in mind that there can be negative social effects. PA does not replace the lecturer and students must accept the chosen tool as a fair tool. Also, especially in large groups, consider whether additional support is available to deal with the administrative workload of PA tools so that lecturers can be unburdened and monitor and assist the students.

REFERENCES

[1] D. Carless, ''Scaling up assessment for learning: Progress and prospects,'' in *Scaling Up Assessment for Learning in Higher Education*, D. Carless, S. Bridges, C. Chan, and R. Glofcheski, Eds., 1st ed. Singapore: Springer, 2017, pp. 3–17, doi: 10.1007/978-981-10-3045-1_11.

[2] J. Turns, B. Sattler, K. Yasuhara, J. Borgford-Parnell, and C. Atman, ''Integrating reflection into engineering education,'' in *Proc. ASEE Annu. Conf. Expo.*, Indianapolis, IN, USA, 2014, pp. 24–776, doi: 10.18260/1-2–20668.

[3] R. Graham, *The Global State of the Art in Engineering Education*. Cambridge, MA, USA: MIT Press. [Online]. Available: https://jwel.mit.edu/assets/document/global-state-art-engineering-education

[4] S. Asgari, J. Trajkovic, M. Rahmani, W. Zhang, R. C. Lo, and A. Sciortino, ''An observational study of engineering online education during the COVID-19 pandemic,'' *PLoS ONE*, vol. 16, no. 4, pp. 1–17, 2021, doi: 10.1371/journal.pone.0250041.

[5] O. B. Adedoyin and E. Soykan, ''COVID-19 pandemic and online learning: The challenges and opportunities,'' *Interact. Learn. Envir.*, vol. 8, no. 3, pp. 54–72, 2020, doi: 10.1080/10494820.2020.1813180.

[6] O. Luaces, J. Díez, and A. Bahamonde, ''A peer assessment method to provide feedback, consistent grading and reduce students' burden in massive teaching settings,'' *Comput. Educ.*, vol. 126, pp. 283–295, Nov. 2018, doi: 10.1016/j.compedu.2018.07.016.

[7] K. Topping, ''Peer assessment between students in colleges and universities,'' *Rev. Educ. Res.*, vol. 68, no. 3, pp. 249–276, Sep. 1998, doi: 10.3102/00346543068003249.

[8] T. Tenório, I. I. Bittencourt, S. Isotani, and A. P. Silva, ''Does peer assessment in on-line learning environments work? A systematic review of the literature,'' *Comput. Hum. Behav.*, vol. 64, pp. 94–107, Nov. 2016, doi: 10.1016/j.chb.2016.06.020.

[9] N.-F. Liu and D. Carless, ''Peer feedback: The learning element of peer assessment,'' *Teach. Higher Educ.*, vol. 11, no. 3, pp. 279–290, 2006, doi: 10.1080/13562510600680582.

[10] M. Keppell, E. Au, A. Ma, and C. Chan, "Peer learning and learning-oriented assessment in technology-enhanced environments," *Assessment Eval. Higher Educ.*, vol. 31, no. 4, pp. 453–464, Aug. 2006, doi: 10.1080/02602930600679159.

[11] L. Nagel and T. G. Kotzé, "Supersizing E-learning: What a CoI survey reveals about teaching presence in a large online class," *Internet Higher Educ.*, vol. 13, nos. 1–2, pp. 45–51, Jan. 2010, doi: 10.1016/j.iheduc.2009.12.001.

[12] B. Divjak and M. Maretic, "Learning analytics for peer-assessment: (Dis) advantages, reliability and implementation," *J. Inf. Organizational Sci.*, vol. 41, no. 1, pp. 21–34, Jun. 2017, doi: 10.31341/jios.41.1.2.

[13] J. Gillett-Swan, "The challenges of online learning: Supporting and engaging the isolated learner," *J. Learn. Des.*, vol. 10, no. 1, pp. 20–30, 2017, doi: 10.5204/jld.v9i3.293.

[14] K. J. Topping, "Digital peer assessment in schoolteacher education and development: A systematic review," *Res. Papers Educ.*, pp. 1–27, 2021, doi: 10.1080/02671522.2021.1961301.

[15] J. Tornwall, "Peer assessment practices in nurse education: An integrative review," *Nurse Educ. Today*, vol. 71, pp. 266–275, Jan. 2018, doi: 10.1016/j.nedt.2018.09.017.

[16] E. Panadero and M. Alqassab, "An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading," *Assessment Eval. Higher Educ.*, vol. 44, no. 8, pp. 1253–1278, Nov. 2019, doi: 10.1080/02602938.2019.1600186.

[17] T. D. Indriasari, A. Luxton-Reilly, and P. Denny, "Gamification of student peer review in education: A systematic literature review," *Educ. Inf. Technol.*, vol. 25, pp. 5202–5234, Nov. 2020, doi: 10.1007/s10639-020-10228-x.

[18] D. Reinholz, "The assessment cycle: A model for learning through peer assessment," *Assessment Eval. Higher Educ.*, vol. 41, no. 2, pp. 301–315, Feb. 2016, doi: 10.1080/02602938.2015.1008982.

[19] M. L. Cruz, G. N. Saunders-Smits, and P. Groen, "Evaluation of competency methods in engineering education: A systematic review," *Eur. J. Eng. Educ.*, vol. 45, no. 5, pp. 729–757, Sep. 2020, doi: 10.1080/03043797.2019.1671810.

[20] M. L. Cruz, M. E. D. Van Den Bogaard, G. N. Saunders-Smits, and P. Groen, "Testing the validity and reliability of an instrument measuring engineering students' perceptions of transversal competency levels," *IEEE Trans. Educ.*, vol. 64, no. 2, pp. 180–186, May 2021, doi: 10.1109/TE.2020.3025378.

[21] M. Mina, J. Cowan, and J. Heywood, "Case for reflection in engineering education- and an alternative," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, El Paso, TX, USA, Oct. 2015, pp. 1–6, doi: 10.1109/FIE.2015.7344252.

[22] D. Mueller and S. Strohmeier, "Design characteristics of virtual learning environments: State of research," *Comput. Educ.*, vol. 57, pp. 2505–2516, Dec. 2011, doi: 10.1016/j.compedu.2011.06.017.

[23] C. Heaton-Shrestha, S. May, and L. Burke, "Student retention in higher education: What role for virtual learning environments?" *J. Further Higher Educ.*, vol. 33, no. 1, pp. 83–92, Feb. 2009, doi: 10.1080/03098770802645189.

[24] D. Bri, M. Garcia, H. Coll, and J. Lloret, "A study of virtual learning environments," *Wseas Trans. Adv. Eng. Educ.*, vol. 6, no. 1, pp. 33–43, 2009.

[25] D. Thompson and I. McGregor, "Self and peer assessment for group work in large classes," in *Proc. Eval. Assess. Conf.*, 2005, pp. 41–52.

[26] C. Campbell, L. Borer, and S. McCarthy, "Using feedbackfruits to enhance student learning: Scaling for transformative implementation," in *Proc. 37th Int. Conf. Innov.*, 2020, pp. 322–327, doi: 10.14742/ascilite2020.0148.

[27] G. N. Saunders-Smits, P. Roling, V. Brugemann, N. Timmer, and J. Melkert, "Using the engineering design cycle to develop integrated project-based learning in aerospace engineering," in *Proc. Int. Conf. Innov., Practice Res. Eng. Educ.*, Coventry, U.K., 2012, pp. 1–13. [Online]. Available: http://resolver.tudelft.nl/uuid:eea9a2e8-f9c3-4b0c-b0d4-a44cde3bf6b1

[28] A. C. Loignon, D. J. Woehr, J. S. Thomas, M. L. Loughry, M. W. Ohland, and D. M. Ferguson, "Facilitating peer evaluation in team contexts: The impact of frame-of-reference rater training," *Acad. Manag. Learn. Educ.*, vol. 16, no. 4, pp. 562–578, Dec. 2017, doi: 10.5465/amle.2016.0163.

[29] M. J. Grant and A. Booth, "A typology of reviews: An analysis of 14 review types and associated methodologies," *Health Inf. Libraries J.*, vol. 26, no. 2, pp. 91–108, Jun. 2009, doi: 10.1111/j.1471-1842.2009.00848.x.

[30] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, and R. Chou, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Int. J. Surg.*, vol. 88, Mar. 2021, Art. no. 105906, doi: 10.1136/bmj.n71.

[31] D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart, "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," *Syst. Rev.*, vol. 4, no. 1, pp. 1–9, 2015, doi: 10.1186/2046-4053-4-1.

[32] W. Trochim and J. Donnelly, *The Research Method Knowledge Base*, 3rd ed. Mason, OH, USA: Cengage, 2008, pp. 33–52.

[33] T. R. Greenhalgh, "Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources," *BMJ*, vol. 331, pp. 1064–1065, Nov. 2005, doi: 10.1136/bmj.38636.593461.68.

[34] K. J. Topping, "Peer assessment: Learning by judging and discussing the work of other learners," *Interdiscipl. Educ. Psychol.*, vol. 1, no. 1, pp. 1–7, Oct. 2017, doi: 10.31532/InterdiscipEducPsychol.1.1.007.

[35] T. D. Indriasari, A. Luxton-Reilly, and P. Denny, "A review of peer code review in higher education," *ACM Trans. Comput. Educ.*, vol. 20, no. 3, pp. 1–25, Sep. 2020, doi: 10.1145/3403935.

[36] S. S. Rosa, C. P. Coutinho, and M. A. Flores, "Online peer assessment: Method and digital technologies," *Proc.-Social Behav. Sci.*, vol. 228, pp. 418–423, Jul. 2016, doi: 10.1016/j.sbspro.2016.07.064.

[37] L. Zheng, N.-S. Chen, P. Cui, and X. Zhang, "A systematic review of technology-supported peer assessment research," *Int. Rev. Res. Open Distrib. Learn.*, vol. 20, no. 5, pp. 168–190, Aug. 2019, doi: 10.19173/irrodl.v20i5.4333.

[38] E. Triantafyllou and O. Timcenko, "Peer assessment in engineering group projects: A literature survey," in *Proc. ICWL*, Tallinn, Estonia, vol. 8699, 2014, pp. 66–71, doi: 10.1007/978-3-319-13296-9_7.

[39] A. Luxton-Reilly, "A systematic review of tools that support peer assessment," *Comput. Sci. Educ.*, vol. 19, no. 4, pp. 209–232, Dec. 2009, doi: 10.1080/08993400903384844.

[40] Q.-K. Fu, C.-J. Lin, and G.-J. Hwang, "Research trends and applications of technology-supported peer assessment: A review of selected journal publications from 2007 to 2016," *J. Comput. Educ.*, vol. 6, no. 2, pp. 191–213, Jun. 2019, doi: 10.1007/s40692-019-00131-x.

[41] M. van Zundert, D. Sluijsmans, and J. van Merrionboer, "Effective peer assessment processes: Research findings and future directions," *Learn. Instruct.*, vol. 20, no. 4, pp. 270–279, Aug. 2010, doi: 10.1016/j.learninstruc.2009.08.004.

[42] S. Lerchenfeldt, M. Mi, and M. Eng, "The utilization of peer feedback during collaborative learning in undergraduate medical education: A systematic review," *BMC Med. Educ.*, vol. 19, no. 1, pp. 1–10, 2019, doi: 10.1186/s12909-019-1755-z.

GITTE VAN HELDEN received the B.S. and M.S. degrees in educational sciences from Utrecht University, Utrecht, The Netherlands, in 2018 and 2019, respectively. She is currently pursuing the Ph.D. degree in aerospace engineering with Delft University of Technology, Delft, The Netherlands, funded by the 4TU Centre for Engineering Education.

After receiving her M.S. degree, she worked as a Research Assistant with the Freudenthal Institute, Utrecht University, with a focus on embodied cognition in mathematics education. She was a Junior Lecturer with the Institute of Education and Child Studies, Leiden University, The Netherlands. Her Ph.D. research project is focused on the development of Collaborative Design Laboratories for educational purposes. She aims to provide insights in how learning activities in engineering design education can be orchestrated in order to promote productive collaboration.

VIVIAN VAN DER WERF received the B.A. and M.A. degrees in archaeology and the M.S. degree in educational sciences from Leiden University, Leiden, The Netherlands, in 2013, 2018, and 2018, respectively. She is currently pursuing the joint Ph.D. degree with Leiden-Delft-Erasmus Centre for Education and Learning and the Leiden Institute of Advanced Computer Science, Leiden University.

During and after her studies, she acquired several years of experience in the field of higher education as a research assistant, a policy analyst, and a data analyst, however, she never lost interest in how people learn new skills and how learning can be facilitated. Her research interest includes programming education. She is keen to apply and combine her knowledge from these disciplines in her research projects in order to understand how people learn and how this translates to teaching and education.

GILLIAN N. SAUNDERS-SMITS received the M.S. and Ph.D. degrees in aerospace engineering from Delft University of Technology, Delft, The Netherlands, in 1998 and 2008, respectively.

After a short stint in industry, she started her academic career as a Lecturer in aerospace engineering with Delft University of Technology, in 1999. She was promoted to a Senior Lecturer/an Associate Professor, in 2016. From 2020 to 2022, she was seconded to the Leiden-Delft-Erasmus Centre for Education and Learning. She is currently a Senior Lecturer/an Associate Professor with the Cognitive Robotics Department and the Mechanical Engineering Department, Delft University of Technology. She has over 20 years of experience in teaching a variety of engineering courses using a plethora of methods as well as much experience in managing and organizing education and curricula. She is an invited speaker with over 60 publications to her name. Her main research interests include skills and competencies, curriculum development, and online education in the field of engineering education.

Dr. Saunders-Smits is a member of ASEE, KIvI, SEFI, and UK&IE EERN. Her awards and honors include the ''Universiteitsfonds—Ritsema van Eck Prize'' for excellent teamwork, in 2003, the SEFI Best Annual Conference Paper Award, in 2011, and the SEFI Best Annual Conference Research Paper Award, in 2020. In 2021, she was one of the finalists for the edX prize. In 2022, she was awarded the Open Education Ambassador Award from TU Delft.

MARCUS M. SPECHT received the Diploma degree in psychology and the Ph.D. degree in adaptive information technology from the University of Trier, Trier, Germany, in 1995 and 1998, respectively.

Since 2001, he has been the Head of the Department of Mobile Knowledge at Fraunhofer Institute for Applied Information Technology (FIT), Sankt Augustin, Germany. From 2005 to 2018, he was a Professor of learning technologies with the Open University of the Netherlands, Heerlen, The Netherlands, and the Head of the Learning Innovation Laboratory. He is currently a Professor of digital education with the Delft University of Technology, Delft, The Netherlands, and the Director of the Leiden-Delft-Erasmus Centre for Education and Learning. His research interests include mobile and contextualized learning technologies and social and immersive media for learning.

Prof. Specht is a member of ACM and SIKS and ICO Research Schools, The Netherlands. He is currently an Apple Distinguished Educator. He was the President of the International Association of Mobile Learning, from 2013 to 2015.