Received: 6 March 2023 | Revised: 21 September 2023 | Accepted: 5 November 2023
DOI: 10.1111/jcal.12907

ARTICLE

Journal of Computer Assisted Learning WILEY

Analysing students' self-assessment practice in a distance education environment: Student behaviour, accuracy, and task-related characteristics

Slaviša Radović[1] | Niels Seidel[1] | Joerg M. Haake[2] | Regina Kasakowskij[1]

[1]Center of Advanced Technology for Assisted Learning and Predictive Analytics (CATALPA), FernUniversität in Hagen, Hagen, Germany

[2]Chair of Cooperative Systems, FernUniversität in Hagen, Hagen, Germany

Correspondence
Slaviša Radović, FernUniversität in Hagen, Universitätsstr. 27 – IZ/Building 5, Room A 116, 58097 Hagen, Germany.
Email: slavisa.radovic@fernuni-hagen.de

Funding information
Center of Advanced Technology for Assisted Learning and Predictive Analytics (CATALPA) of the FernUniversität in Hagen

Abstract

Background: Self-assessment serves to improve learning through timely feedback on one's solution and iterative refinement as a way to improve one's competence. However, the complexity of the self-assessment process is widely recognized, as well as that students can benefit from it only if their assessment is accurate enough.

Objectives: In order to gain more insight into the self-assessment process we analysed students' behaviour, accuracy, and question-related characteristics that influence the capability of self-assessment in two studies.

Methods: The initial study examined 131 undergraduate students using voluntary self-assessment questions in an online course in a B.Sc. Computer Science program while a year later a replication study with the same research settings was applied to a different cohort of 264 undergraduate students with minor modifications to the question design, in the light of the original findings.

Results and Conclusions: Results from both studies show that similar patterns could be observed for usage and of accuracy and score distribution for almost all questions. Item difficulty and comprehensiveness of the sample solution were identified as features of self-assessment questions affecting student's self-assessment capability. The replication study showed that task design can be modified to affect students' accuracy. Recommendations to make self-assessment tasks effective and efficient for learning are provided.

KEYWORDS
criteria-based self-assessment, computer-assisted feedback, self-assessment, self-assessment accuracy

1 | INTRODUCTION

The concept of self-assessment is not new and has been used in higher education practice for a long time. As defined by Klenowski (1995, p. 146), self-assessment is 'the evaluation or appraisal of the "value" of one's own performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes'. Typically, in practice, it is operationalized as a cyclic process of improving students' knowledge by reviewing their own performance and identifying the gap between current and desired understanding. Learners themselves become stakeholders in the assessment process (Ćukušić et al., 2014). In the theory of learning assessment, self-assessment is recognized as one of the valuable pedagogical tools to enable students to play an active role in learning process and self-regulation. The other literature in the field also recognizes that this type of assessment encourages learners to 'become more aware of their own process by self-regulating their own learning and consequently promoting individual responsibility' (Duque Micán & Cuesta Medina, 2015, p. 3).

The full potential and pedagogical importance of self-assessment has increased over the past decade (Ozarslan & Ozan, 2017). Advances in technology-enhanced learning and information processing techniques (e.g., data mining and learning analytics) have facilitated the greater development, exploration, and application of various self-assessment methods in the context of blended learning, online learning, and distance education. Ćukušić et al. (2014) noted that most modern Learning Management Systems (LMS) and Integrative Learning Environments (ILE) consider tools for developing and processing online self-assessments as an indispensable learning and assessment feature.

Indeed, the integration of self-assessment into distance education management systems has transformed the learning process, allowing learners to measure their knowledge and receive immediate feedback without regard to teacher availability, location, or time of learning (Ozarslan & Ozan, 2017; Yan, 2020; Yan & Brown, 2017). However, challenges exist in the design of self-assessment itself, as well as in the utilization of such assessments by students (Brown & Harris, 2013; Panadero et al., 2017). For example, certain task characteristics can carry negative connotations and hinder students' usage of self-assessment (Chen & Zhang, 2023). Additionally, Panadero et al. (2016) highlighted that task design characteristics such as task complexity, scope and transfer requirements, distractors, sample solution text length and number of evaluation criteria, could be altered to influence students' accuracy. However, while prior research indicates that self-assessment generally enhances students' motivation to engage with learning (Brown & Harris, 2013; Ozarslan & Ozan, 2017), and fosters both academic achievement and self-efficacy (Panadero et al., 2017), relatively little attention has been paid to altering task design characteristics to impact self-assessment behaviour and accuracy of students. Therefore, this research aims to answer three research questions:

RQ1. What behaviour students show when using voluntary self-assessment tasks? Here, we examine the extent of usage over time of self-assessment tasks.

RQ2. How well do learners self-assess their solutions? Here, we examine accuracy of self-assessment by comparing student self-assessment with instructor assessment and whether accuracy is related to the quality of the student's solution.

RQ3. What self-assessment task-related characteristics influence the self-assessment? Here, we look at the impact of item difficulty, discrimination index, number of self-assessment criteria, use of distractors, and length of the sample solution on self-assessment accuracy.

To gain a comprehensive understanding, the next section of this paper will delve into a literature review. It will begin by briefly exploring the benefits of self-assessment. Subsequently, it will thoroughly examine strategies that bolster students' self-assessment, as well as the associated challenges. This exploration aims to uncover the intricacies of self-assessment accuracy and students' behaviours, thus establishing a solid theoretical and research foundation for our study's research questions. Section 4 will introduce the research methodology, detailing the study's context, outlining the research design, and describing the learning environment. Sections 5 and 6 will present the data analysis and research findings, covering the initial and the replication study. In the final section, the paper will conclude with discussions, limitations, and conclusions drawn from studies.

2 | LITERATURE REVIEW

Numerous studies have demonstrated the benefits of self-assessments, mainly on three sets of dependent variables: (1) self-regulation and learning process; (2) cognition and academic outcomes; and (3) motivation and perceptions of self-efficacy (Brown & Harris, 2013; Panadero et al., 2017; Yan, 2020). First, researchers have found that self-assessments in online learning environments provide learners with a variety of practice opportunities to address learning objectives, requirements, and assessment criteria, as well as the opportunity to use feedback to increase learning (Hartung, 2017). As a result, students indicate that they are able to understand how learning occurs, as a function of their own abilities and personal effort (Race, 2001; Yan, 2020).

In addition to the benefits to the learning process, one of the most significant claims is that it improves learning outcomes and enables developing better understanding. Students acknowledged that after being able to self-assess themselves, they perform better on final exams (Ćukušić et al., 2014), and develop better metacognitive engagement with learning content (Andrade & Du, 2007). Other researchers also noted students' academic benefits such as improved metacognitive competencies, development of a variety of skills and improved performance (Duque Micán & Cuesta Medina, 2015; Ozarslan & Ozan, 2017). In contrast, some researchers expressed less optimism about the effects of self-assessment on student learning (Lew et al., 2010).

Another benefit described is that self-assessment positively affects students' motivation and responsibility for learning (Andrade & Du, 2007), as well as their belief that they can achieve a goal (Castillo-Merino & Serradell-Lopez, 2014). It is also suggested that self-assessment can alleviate student anxiety and can facilitate students acquiring a sense that they have control over their own evaluation. In other studies, students reported feeling less stressed when completing the online self-assessment during learning (Andrade & Du, 2007; Ozarslan & Ozan, 2017). When students feel comfortable taking the assessment, their engagement in the learning process can be increased, thereby promoting their self-efficacy and satisfaction (Andrade & Du, 2007; Thawabieh, 2017).

2.1 | Strategies to support students' self-assessment

Several strategies for addressing the perils of self-assessment can be distilled from research studies. It has become clear that for effective and efficient self-assessment during learning, students should be supported to evaluate their own work using appropriate criteria (Panadero et al., 2016; Yan & Brown, 2017). Andrade and Du (2007) described a pedagogical three-step process: (1) the assessment criteria, desired performance, or expectations are communicated to students, (2) students complete the assignment and review their work using the rubric, and (3) students revise and improve their work based on feedback from the self-assessments. Many other researchers (e.g., Duque Micán & Cuesta Medina, 2015; Yan & Brown, 2017) argued, with minor variations, the stages of the evaluation process similar to those presented in the work of Andrade and Du (2007).

A recent update in this area is the iterative process model of self-assessment for online learning proposed by Haake et al. (2022), as shown in Figure 1. In this model, students begin the self-assessment process by selecting a relevant learning task to work on. Then the task including instructions is shown and students are asked to create and submit a solution. Thereafter, a list of assessment criteria set by the instructor is presented, a link to a sample solution is provided, and students are prompted to rate the submitted solution. After students rate their solution against the relevant assessment criteria, feedback is automatically selected from an instructor-defined feedback database based on their self-assessment and presented to students. With the help of the feedback, students can then reflect on the quality of their learning products. The feedback texts help students improve their solution, for example, by providing feedback on learning objectives (e.g., misunderstood concepts) and the learning process (e.g., links to relevant resources, things to improve, or activities to complete) until they self-assess their solution as correct or good enough (Haake et al., 2020, 2022). After receiving feedback, students can either improve their solution by performing a new iteration of the process (create, submit, self-assess the improved solution again, receive feedback, and accept or reject another iteration) or complete the exercise. From a pedagogical perspective, such iterative self-assessment serves to improve, rather than to provide a comprehensive summative assessment of one's knowledge (Ćukušić et al., 2014; Radović et al., 2019). The reader should note that this approach differs from Andrade and Du (2007) with respect to the automatic selection of instructor-defined feedback on students' self-assessment and the possibility of iterative improvement of the task solution.

2.2 | Challenges of self-assessment

Despite strong empirical support for a positive association between self-assessment and satisfaction with learning process and learning outcomes, the design of self-assessment itself presents some challenges (Brown & Harris, 2013; Panadero et al., 2017). Several prominent perils have been recognized in the literature in the relation to learning accuracy, learners' behaviour, and task characteristics (Boud et al., 2015; Lew et al., 2010).

Regarding RQ1, a strand of research delves into the diverse behaviours exhibited by students (Ifenthaler et al., 2023). Yang et al. (2022) proposed a categorization of students into distinct groups based on their engagement levels, frequency of repetition, ability to recall information, and dependence on hints. Studies have revealed that students who engage in regular self-assessment and repeat tasks multiple times tend to achieve superior scores compared to those who only attempt tasks once (Boud et al., 2015; Ifenthaler et al., 2023; Yang et al., 2022). However, it is worth noting that the mere act of repeatedly reviewing and answering questions during assessments does not inherently facilitate effective learning. The impact of spacing intervals emerges as a potent factor for enhancing the retention of information over extended durations (Yang et al., 2022). Although the utilization of self-assessments by students is often confined to specific periods within the semester (Ifenthaler et al., 2023), the study conducted by Ozarslan and Ozan (2017) suggests that the regularity of self-assessment exhibits a positive correlation with course outcomes, whereas frequency alone does not. Furthermore, a notable challenge posed by online assessments lies in permitting students to iteratively submit responses and access hints at their discretion. Yang et al. (2022) reported that while this approach grants students the flexibility to manage their learning pace, it tends to give rise to unconventional behaviours such as 'repetitive practice without spacing, misappropriation of feedback, and excessive reliance on hints'.

To appropriately serve its function, students' self-assessments need to be accurate (Ernst et al., 2023). *Regarding RQ2*, numerous students often exhibit a tendency to inaccurately evaluate their own abilities (Max et al., 2022). Ernst et al. (2023) conducted a comprehensive literature review, revealing that students' relative accuracy remained notably low and exhibited a negative correlation with students' overconfidence. A parallel observation emerged from the study conducted by Chen and Zhang (2023), wherein participants who self-reported as 'highly accurate' in surveys predominantly achieved scores below the mean, while conversely, those who reported being 'not very accurate'

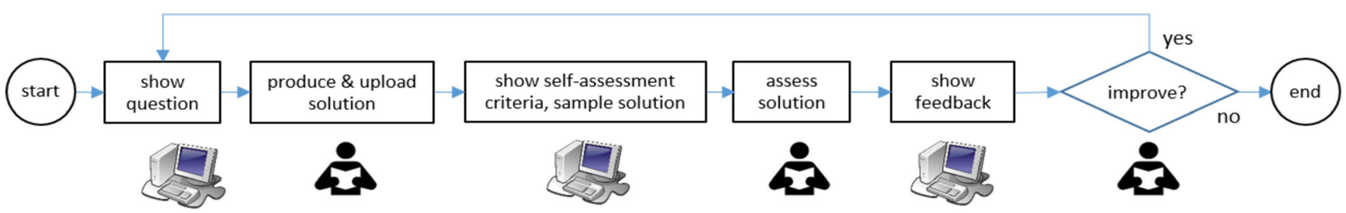

FIGURE 1 Conceptual model of iterative process model of self-assessment for online learning.

often attained above-average scores. Furthermore, it was established that students in the early stages of their academic journey (Zhang & Zhang, 2022) or those with poorer academic outcomes tend to overestimate their own performance and level of knowledge (Carroll, 2020; Max et al., 2022). Additionally, Lew et al. (2010) found that students' self-assessment ability does not improve over the course of a semester, and that more academically competent students have more accurate self-assessments of their knowledge than less competent students. However, an important conclusion from Thawabieh's (2017) study recognizes that students can accurately assess themselves when given criteria and feedback on both self-assessment activity and misconceptions in their solution.

Regarding RQ3, different forms of assessment tasks and corresponding characteristics may also influence student behaviour and learning benefits (Boud et al., 2015). Research studies indicate that the type of task (e.g., less cognitively demanding tasks versus more cognitively demanding authentic problems that require knowledge transfer) used for the assessment activity can influence the self-assessment process (Andrade, 2019). Chen and Zhang (2023) observed that certain task characteristics (such as those involving language intricacies, adjectives with elusive quantifiability, terms laden with jargon or technicality, and words carrying negative connotations) were difficult for students to use. Furthermore, Rohr's study (2018) substantiated the connection between self-efficacy and the quality of textual task completion. It has also been demonstrated that the self-assessment accuracy tends to diminish with increasing task complexity (Andrade, 2019; Rohr, 2018). Additionally, Panadero et al. (2016) highlighted that task design (characteristics, e.g., task complexity, scope and transfer requirements, distractors, solution text length and number of evaluation criteria) can be altered to influence students' accuracy. Yet, the specific adjustments required for such modification remain indistinct.

Recent studies (Chen & Zhang, 2023; Panadero et al., 2016; Zhang & Zhang, 2022) distinctly underscore that only a limited corpus of studies has dedicated attention to dissecting the interplay between self-assessment behaviour and task-related attributes, alongside students' accuracy.

3 | METHOD

3.1 | Research design

To investigate our research questions and to assure that results are reliable and generalizable to different cohorts of students, this research consists of an initial and replication study. Both studies were conducted under the same circumstances using the same research conditions and measurement scales, situated in the compulsory course 'Operating Systems and Computer Networks' of a distance learning B.Sc. Computer Science (CS) course of a German distance teaching university. The initial study took place in winter semester 2020/2021, the replication study in winter semester 2021/2022. For the enrolled students a supplementary course was set up in a Moodle learning environment including the SelfAssess-plugin (cf. Section 3.2). The use of the learning environment was voluntary, but conditional on a two-step consent to use the platform and to participate in the study.

3.2 | Learning environment and content

In order to support iterative improvement, the SelfAssess plugin (Steinkohl et al., 2021), that allows students to create and assess solutions unlimited times (see Figure 1), was implemented in Moodle. Self-assessment tasks were aimed at training constructive or analytical skills. Therefore, criteria for self-assessment were tailored to test properties of a correct solution—often implicating steps of a correct argumentation or computation. Thus, a binary scale was sufficient for letting students assess whether their solution fulfils the respective criterion. Figure 2 shows the plugin user interface as well as the process of working on a

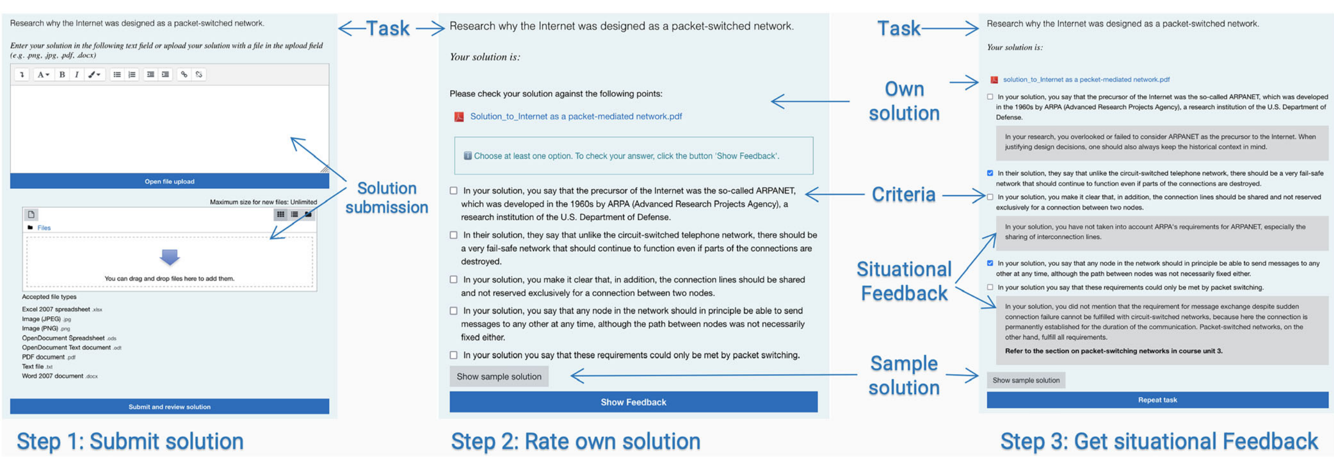

FIGURE 2 Example of a student's interaction with the SelfAssess plugin.

self-assessment task. In the first step, a student works on the task and submits his solution. In a second step, the student evaluates his solution by rating whether the indicated criterion is fulfilled by the submitted solution (criterion checked) or not (criterion unchecked). In the third step, the student receives feedback appropriate to his or her self-assessed solution. Now the first step (i.e., editing process of the solution) can be repeated in order to improve the solution based on the feedback or the solution can be finally handed in when one is done with working on the solution. In addition, the student has the possibility to access the sample solution in step 2 as well as in step 3.

For both studies, the learning environment contained four course units with course texts, 43 self-assessment tasks (cf. dataset Haake et al., 2021) assigned to the course units, complementing 23 multiple choice questions, and 30 exercises corrected by a tutor. To be able to evaluate the self-assessment of the valid sample solutions by an expert, it was necessary to limit the number of tasks to restrict the number of solutions to be evaluated by the expert. From each of the four course units we selected one task that was completed by the most students (using the same tasks for Study 1 and Study 2). In course Unit 4 two tasks have been processed by the same number of participants, which is why we have chosen both. So, from a total of 43 tasks (Haake et al., 2021) five tasks were selected for the analysis. The design characteristics of these self-assessment tasks, presented in Table 1, are described with respect to the number of words used in task description and sample solution, number of self-assessment criteria, number of distractors and pedagogical goals for including tasks in learning material. In the light of the results of Study 1 the low accuracy of Task 5 was likely caused by the comprehensiveness of the sample solution. Therefore, we changed this task in Study 2 in a way to reduce the length of the sample solution from 289 to 31 words. At the same time the number evaluation criteria could be reduced from 7 to 4.

3.3 | Data collection and analysis

The course and thus the field study 1 began on 1st October 2020. Study 2 began on 1st October. Students were free to choose when to start with the course and when to engage with which tasks and exercises. The course and research period ended after 6 months.

User interactions and user inputs within the Moodle environment have been captured in the Moodle database, especially in the standard log store. Compared to other question type plugins, the SelfAssess plugin (Steinkohl et al., 2021) allows for additional logging data such as the number of times a task was retried, solution submitted, and self-assessment. Additionally, the submitted students' solutions were stored in the Moodle data folder. These files have been used for expert rating by the course instructor. The validity of the student solution, the file type and the writing style (by hand or typewritten) was classified along with the correctness of valid submitted solutions.

3.4 | Measures

As shown in Section 3.2, each self-assessment task defines a number N of criteria with (1) an associated text expressing the condition that the student needs to evaluate as being fulfilled by his/her solution and (2) a Boolean indicator, whether this criterion should hold for a correct solution or not (i.e., indicating a distractor). Students could

TABLE 1 Overview of self-assessment task-related characteristics in Study 1 and Study 2.

Task	CU	Study	Wt	Ws	SAC	Nd	Task description
Task 1	1	Study 1 and 2	42	97	6	5	Compare memory access speed w/o caching (recall facts/ mechanisms/impact on speed, compare two alternatives)
Task 2	2	Study 1 and 2	34	68	3	0	Compute number of pages needed for given memory request (recall paging method and fragmentation concept, compute needed pages and unused space)
Task 3	3	Study 1 and 2	77	278	6	0	Compute end-to-end delay for given network (ecall properties of nodes and links in a network, delay and route concept, compute delay along a route)
Task 4	4	Study 1 and 2	51	90	3	0	Provide a rationale for given cheapest path in a given network (recall path and cost concepts, find all possible paths, compare costs, provide argumentation)
Task 5	4	Study 1	38	289	7	1	Apply distance vector routing algorithm for a given situation (requiring 8 iterations) (recall distance vector routing algorithm, apply algorithm for given starting state, perform needed number of iterations and produce routing tables for all nodes in the network)
Task 5	4	Study 2	38	31	4	1	Apply distance vector routing algorithm for a given situation (1 iteration) (recall distance vector routing algorithm, apply algorithm for given starting state, perform only one iteration and produce routing tables for all nodes in the network)

Abbreviations: CU, course unit; Nd, number of distractors; s, solution; SAC, self-assessment criteria; t, task; W, text length in words.

open the task in the plugin, read the question text, create a solution, submit it, read the assessment criteria texts and select those that they deem fulfilled by their solution. The resulting self-assessment can be represented as a N-tuple of zeros or ones, indicating whether the ith criterion has been selected by the student.

In order to measure the accuracy of a student's self-assessment for a given self-assessment task, we asked the instructor who defined the self-assessment task to provide a coding schema that defines when a solution fulfils a given criterion or not. Using this coding schema, the instructor rated the submitted student solutions on the given criteria, resulting in instructor assessments represented by similar N-tuples.

The Hamming distance (Hamming, 1950) denotes the number of differences between the two N-tuples. Zero differences denote identical assessments and thus optimal accuracy. A value of N indicates completely different assessments and thus a completely inaccurate self-assessment. Student self-assessment accuracy is then defined as the Hamming similarity between the instructor assessment x and the student's self-assessment y (Hamming, 1950):

$$\begin{aligned}\text{Accuracy}(x,y) &= \text{Hamming Similarity}(x,y)\\ &= 1-\frac{\text{Hamming Distance}(x,y)}{N} \in [0\ldots1] \subset \mathbb{R}.\end{aligned}$$

A value of zero denotes completely inaccurate self-assessment while a value of one denotes completely accurate self-assessment. This definition allows us to compare the self-assessment accuracy of tasks having a different number of criteria. While the previous definitions apply to individual students on an individual self-assessment we can extend the analysis to *task accuracy* representing the mean of all accuracies of student self-assessments of a task.

In addition, the instructor ratings enabled an evaluation of the task items using classical test theory (CTT) (De Champlain, 2010). In CTT, the item difficulty and the item discrimination index are commonly used. Item difficulty is the mean score for an item within a population of participants in a range of zero and one. For ease of interpretation, we transpose item difficulty to the difference between one and the mean score. This means that tasks with a low mean score are considered difficult, and vice versa. Since tasks that are too difficult (>0.7) impair motivation, the majority of tasks should have a difficulty between 0.3 and 0.7. The discrimination index is the Point Biserial Correlation Coefficient of the total scores and the achieved scores for the particular task. As a rule of thumb, tasks with values below 0.2 should be reviewed by content experts to determine if there is a valid reason to retain or exclude the task (De Champlain, 2010).

4 | RESULTS OF STUDY 1

Participants of this study were 180 of the 457 CS course students agreed to take part in the study. By the end of the semester, the same number of active participants had been recorded. The participating students were between 19 and 65 years old ($M = 37.21$, $SD = 9.03$). Participants consisted of 128 males and 52 females.

4.1 | Behaviour shown when using voluntary self-assessment tasks

A total of 131 participants performed at least on self-assessment, of which 97 participants submitted only appropriate solutions regarding the tasks, while 7 participants consistently submitted unrelated files and 27 participants submitted both. In 472 of the 1472 submitted solutions obviously no content-related connection with the respective task could be determined. Most of the valid submitted solutions were typed on a digital device, 144 solutions consisted of photos or screenshots of handwritings. As the semester progresses, a decreasing number of active participants and responses could be observed regarding all course activities (Figure 3). Only 33 participants repeated individual self-assessment tasks. Most repetitions took place within a time range of less than 2 days, only 15 participants repeated a task after more than 2 days.

As detailed in research methodology, from a total of 43 tasks (Haake et al., 2021) five tasks were selected for the analysis. A total of 202 out of 1000 solutions were submitted for these five tasks (by 46 participants), but 56 of them were not related to the task and solutions of 38 more self-assessments were accidentally not recorded or cancelled. Consequently, the accuracy of the self-assessment was evaluated for 108 valid submitted solutions.

4.2 | Accuracy of self-assessment solutions

Figure 4 shows the relationship between accuracy (cf. Section 3.4) and the ratio of the number of fulfilled criteria, as assessed by the

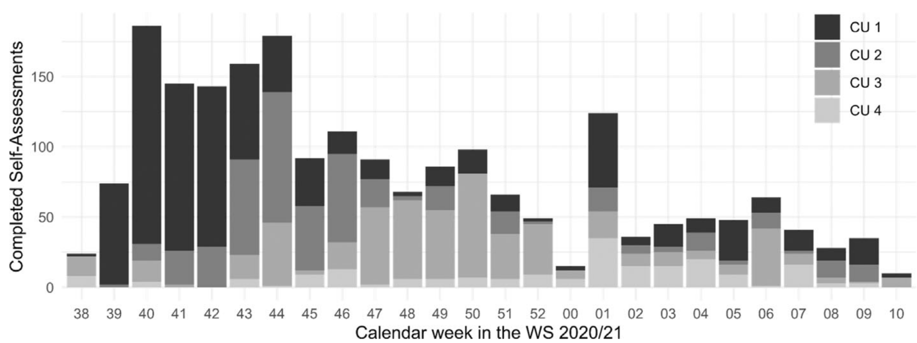

FIGURE 3 Completed self-assessments for each course unit (CU) over the semester.

instructor, divided by the number of criteria *N*. Thus, zero achieved points denote a completely wrong answer while a one indicates a completely correct solution. Values in between indicate partially correct answers. The size of the circles indicates the number of occurrences of this combination. The colour indicates the self-assessment task.

Obviously, students on different correctness levels showed different accuracy for a given question, whereas this relationship differs between questions. Task 1 shows that the majority of students achieve poor solutions (below 0.5 correctness) in combination with a relatively low accuracy (below 0.3). However, accuracy increases with the solution scores. Task 2 shows a diverse pattern, but the majority submitted a correct solution with good accuracy. Task 3 shows more correct solutions (above 0.5 correctness) with good accuracy (upper right quadrant) and a minority of weaker solutions that have been recognized as deficits by the respective participants (upper left quadrant). Task 4 shows that all students were achieving passable solutions with higher accuracy (upper right quadrant). Task 5 shows that most students achieved few points and were also bad at assessing their work. The few solutions with higher scores also showed a higher accuracy.

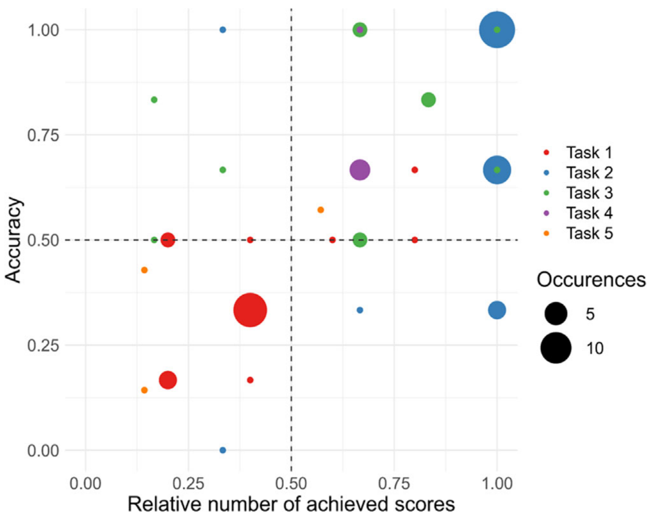

FIGURE 4 Self-assessment accuracy per student and task over the instructor-rated achieved points of a total of 108 edited/self-rated solutions to that task.

A more detailed analysis of the relative shares of accuracy and scores, with respect to each self-assessment task, is illustrated in Figure 5. The results are presented in the form of a matrix, with elements in the upper right area indicating students' high scores and accuracy, while low values for both scores and accuracy are shown in the lower left area. Our analysis shows that the distribution of accuracy and scores for three tasks (2, 3, and 4) predominantly achieved moderate to high values. A higher proportion of lower accuracy (0.17 and 0.33), but still reasonably good scores (0.33 and 0.26) could be noted for Task 1. In contrast, the highest proportion of low accuracy and low scores (0.8) was found for Task 5, indicating that students not only had difficulties in providing the correct solution, but also in self-assessing their solutions using provided evaluation criteria.

Figures 4 and 5 do not yet show whether tasks were processed more than once and to what extent the solution was improved as a result. In 26 cases users iterated a task one more time. In 35 cases, a task was processed twice or even up to 5 times. The solution improved in 11 cases, while it worsened in two cases. In 22 cases the total score did not change, but often the self-assessment did. At the same time the self-assessment accuracy improved in 19 cases ($M = 0.47$, $SD = 0.25$), decreased in 9 cases ($M = -0.35$, $SD = 0.24$) and remained on the same level in 7 cases.

4.3 | Self-assessment task-related characteristics influencing the self-assessment

Following the iterative problem-solving process in Figure 1, self-assessment tasks can be characterized by three analytical dimensions: (i) question, (ii) sample solution, and (iii) self-assessment. The first dimension refers to the question itself as a starting point for the problem-solving process. The tasks considered here differ qualitatively in the task design (e.g., calculation, short answers, and multiple choice), proposed method complexity, scope and transfer requirements and quantitative measures like the text length (Table 2) as well as indicators derived from CTT. The Tasks 1 and 3 have a reasonable degree of difficulty (0.4–0.8), while Tasks 2 and 4 are considered easy and Task 5 seems to have a very high degree of difficulty. The Tasks 1, 2, and 3 have a good discrimination index, thus enabling discrimination between high-performing and low-performing students. In

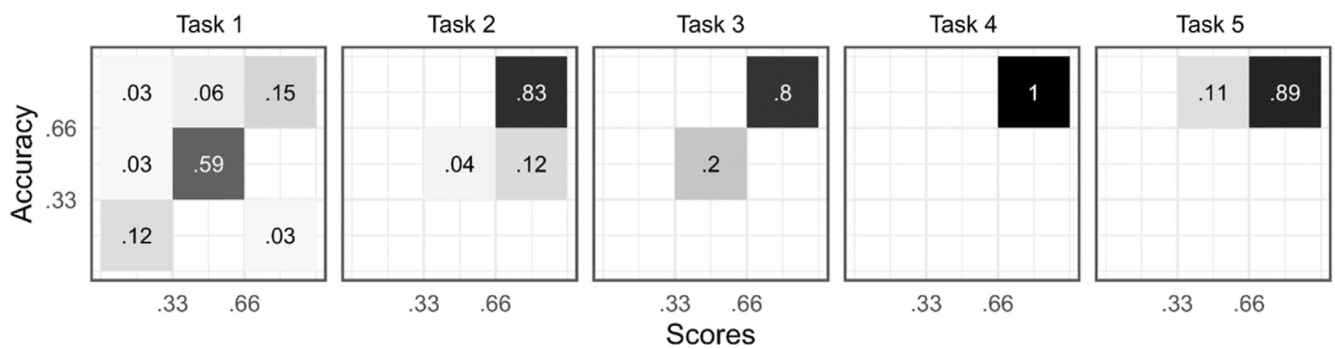

FIGURE 5 Relative share of occurrences for scores and accuracy per third.

contrast, the Tasks 4 and 5 are not suitable to distinguish the performance, not least due to the low number of responses.

As a second dimension the expected solution can be described by the length of the sample solution as shown in Table 2. The sample solutions contain between 68 and 289 words. The solution for Task 5 stands out, because students are required to create 20 tables as part of their solution.

As a third dimension, characteristics related to the self-assessment include the number of evaluation criteria as shown in Table 2 and the correspondence between students' self-evaluation and the expert rating. The evaluation of self-assessment accuracy indicated task-related differences. To better understand these differences, we computed the mean and standard deviation as well as the item difficulty and discrimination index per tasks (Table 2). While the self-assessment tasks in the first course unit were completed by 46 people, the two tasks in course Unit 4 were completed by only 5 people each (see column '*N*' in Table 2). Thus, the results become less reliable with increasing course unit number.

The mean accuracy for the Tasks 2, 3, and 4 show a considerable good level. Task 1 in the first course unit had most student participants, including those that did not submit solutions to the later tasks. Thus, the deviation from the mean accuracy may be larger as more heterogeneous students worked on this task. The comparatively low accuracy of Task 5 is related to the low number of participants and the expected extensive solution. As stated in Table 2 a positive correlation (Kendall's τ) between the self-assessment accuracy and the achieved points could be found.

5 | RESULTS OF STUDY 2: REPLICATION AND CONTENT REFINEMENT

Study 2 was conducted as a replication study to improve the quality and trustworthiness of the analysis of students' self-assessment practice. For this replication study in the winter term 2021/2022 the same research settings was applied to a new cohort of students with minor modifications to the task design of Task 5 (cf. Section 3.2). Participants of this study were 264 of the 560 CS course students who agreed to use the Moodle learning environment and to take part in the study. By the end of the semester, the same number of active participants had been recorded. The participating students were between 19 and 67 years old ($M = 36.38$, $SD = 9.31$). The majority of the participants were male (212), while 52 participants were female.

5.1 | Behaviour shown when using voluntary self-assessment tasks

While 210 participants performed at least on self-assessment, only 78 participants submitted appropriate solutions regarding the tasks. A total of 1203 solutions were submitted. The majority of submitted solutions were typed on a digital device or consisted of photos or screenshots of handwritings. As the semester progresses, a decreasing number of active participants and responses could be observed regarding all course activities (Figure 6). Total number of 48 participants repeated individual self-assessment tasks.

TABLE 2 Overview of self-assessment task-related characteristics in Study 1.

		W				Accuracy		Scores				
Task	CU	t	s	SAC	*N*	*M*	SD	*M*	SD	τ	Idif	DI
Task 1	1	42	97	6	46	0.47	0.34	0.41	0.38	0.70***	0.59	0.37
Task 2	2	34	68	3	35	0.72	0.30	0.82	0.33	0.43*	0.18	0.31
Task 3	3	77	278	6	13	0.81	0.18	0.54	0.36	0.34	0.46	0.74
Task 4	4	51	90	3	5	0.73	0.15	0.67	0.00	n.a.	0.33	n.a.
Task 5	4	38	289	7	5	0.23	0.26	0.17	0.23	0.94·	0.83	0.00

Note: Self-assessment accuracy per task. Significance levels for *p*-values: <0.001 (***); 0.001 (**); 0.01 (*); 0.05 (·)).
Abbreviations: CU, course unit; DI, discrimination index; Idiff, item difficulty; s, solution; SAC, self-assessment criteria; t, task; W, text length in words; τ, Kendall's τ.

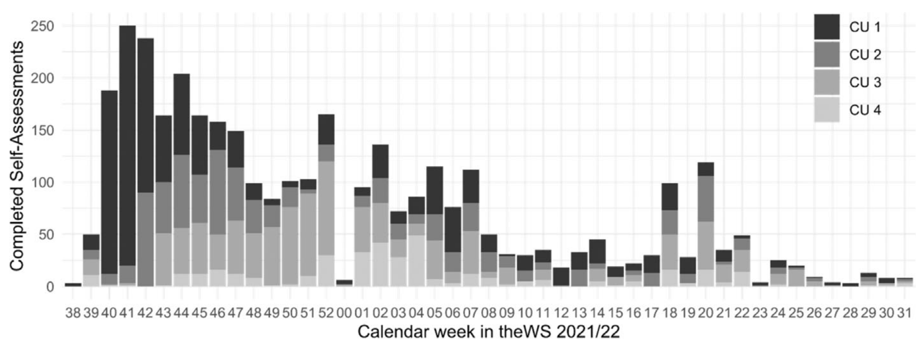

FIGURE 6 Completed self-assessments for each course unit (CU) over the semester.

For reasons of consistency, we selected the same five tasks for expert evaluation as in Study 1. These tasks coincidentally were again the most frequently used in each course unit. A total of 219 out of 1484 solutions were submitted for these five tasks (by 40 participants), but 135 submissions (61.64%) were not related to the task and solutions. Consequently, the accuracy of the self-assessment was evaluated for 84 valid submitted solutions.

5.2 | Accuracy of self-assessments

Figure 7 shows the relationship between accuracy and the ratio of the number of fulfilled criteria, as assessed by the instructor, divided by the number of criteria *N*. Thus, zero achieved points denote a completely wrong answer while a one indicates a completely correct solution. Values in between indicate partially correct answers. The size of the circles indicates the number of occurrences of this combination. The colour indicates the self-assessment task.

Again, students on different correctness levels showed different accuracy for a given task, whereas this relationship differs between

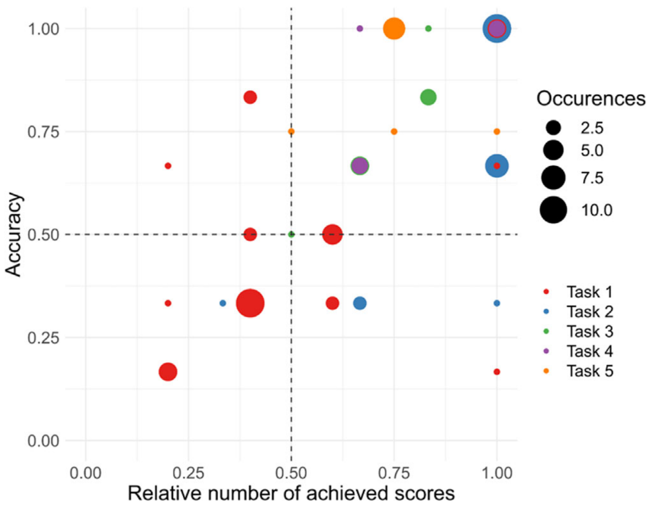

FIGURE 7 Self-assessment accuracy per student and task over the instructor-rated achieved points of a total of 84 edited/self-rated solutions to that task.

tasks. Task 1 shows a wide spread of circles with a cluster of students achieving weaker solutions (around 0.5 correctness) in combination with a middle amount of accuracy (between 0.33 and 0.66). Accuracy mainly increases with the solution scores. Task 2 shows that most students achieved solutions with scores above 0.6 with good accuracy (above 0.6). Task 3 shows a vast majority of more correct solutions (above 0.5 correctness) with good accuracy (upper right quadrant) and a small minority of weaker solutions that have been recognized as deficits by the respective participants (centre of the diagram). Task 4 shows that all students were achieving passable solutions with higher accuracy (upper right quadrant). Task 5 shows that most students achieved good solutions and were also good at assessing their work.

Figure 8 illustrates a more detailed analysis of the relative share of accuracy and scores related to each self-assessment task. Again, the results are presented in a matrix form, with elements in the upper right area indicating high scores and accuracy, while low values for both scores and accuracy, are plotted in the lower left area. Our analysis shows that the distribution of accuracy and scores for four tasks (Tasks 2, 3, 4, and 5) were predominantly located in the upper and right matrix areas, with shares >0.8. This indicates that the vast majority of students were able to give correct answers to these tasks, but also to correctly self-assess their solutions using evaluation criteria. A very small proportion of low (0.12 and 0.03) and moderate (0.03 and 0.59) accuracy shares can be noted for the Task 1. Still, this was accompanied with a quite good score distribution (totalling 0.82 for moderate and high shares).

5.3 | Self-assessment task-related characteristics influencing the self-assessment

The task characteristics are compared across the three analytical dimensions: (i) task, (ii) sample solution, and (iii) self-assessment. The first dimension refers to the task itself as a starting point for the problem-solving process. The tasks considered here differ qualitatively in the task design (e.g., calculation, short answers, and multiple choice), proposed method complexity, scope and transfer requirements and quantitative measures like the text length (Table 3) as well as indicators derived from CTT. Task 1 has a reasonable degree of

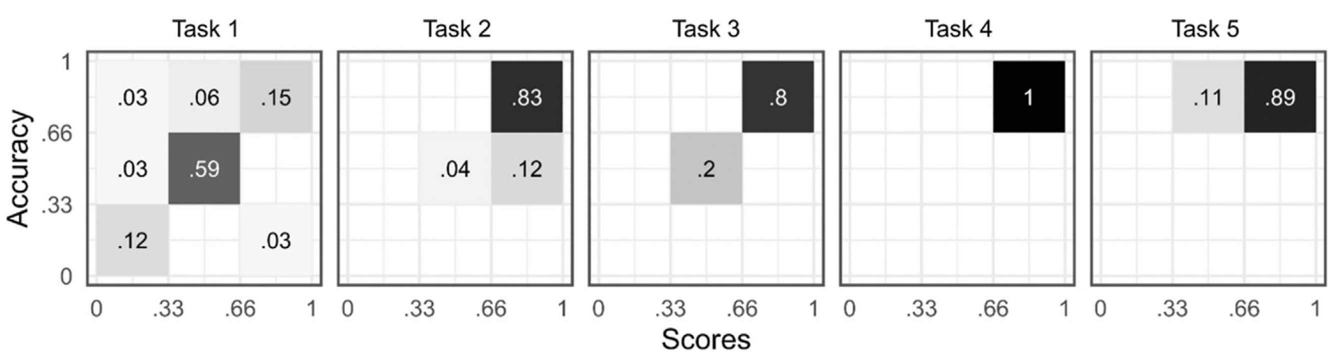

FIGURE 8 Relative share of occurrences for scores and accuracy per third.

TABLE 3 Overview of self-assessment task-related characteristics in Study 2.

		W				Accuracy		Scores				
Task	CU	t	s	SAC	*N*	*M*	*SD*	*M*	*SD*	τ	Idif	DI
Task 1	1	42	97	6	34	0.47	0.26	0.51	0.26	0.54***	0.49	0.49
Task 2	2	34	68	3	24	0.76	0.25	0.92	0.18	0.57*	0.08	0.44
Task 3	3	77	278	6	10	0.70	0.19	0.68	0.17	0.96***	0.32	−0.02
Task 4	4	51	90	3	7	0.86	0.18	0.81	0.18	0.75*	0.19	−0.10
Task 5	4	38	31	4	9	0.92	0.12	0.75	0.12	1.00	0.25	0.18

Note: Self-assessment accuracy per task. Significance levels for *p*-values: <0.001 (***); 0.001 (**); 0.01 (*); 0.05 (·)).
Abbreviations: CU, course unit; DI, discrimination index; Idiff, item difficulty; s, solution; SAC, self-assessment criteria; t, task; W, text length in words; **τ**, Kendall's **τ**.

difficulty (0.4–0.8), while the remaining tasks are considered easy. Tasks 1 and 2 have a good discrimination index, thus enabling discrimination between high-performing and low-performing students. The discrimination index for Task 5 is almost acceptable, that is, suitable for performance differentiation. In contrast, the Tasks 3 and 4 show a negative discrimination index.

The sample solution for the task was considered as a second dimension. The length of the task description and sample solutions presented in Table 3 is identical to the numbers in Table 2, except for the revised Task 5. The latter task has been reformulated and refined so that its sample solution now comprises only two routing tables with about 31 words instead of the former 20 tables containing 289 words.

As a third dimension, characteristics related to the self-assessment include the number of evaluation criteria as shown in Table 3 and the correspondence between students' self-evaluation and the expert rating. The evaluation of self-assessment accuracy indicated task-related differences. To better understand these differences, we computed the mean and standard deviation as well as the item difficulty and discrimination index per task (Table 3). While the self-assessment task in the first course unit was completed by 34 people, the two tasks in course Unit 4 were completed by only 7 and 9 people (see column '*N*' in Table 3). The mean accuracy for the Tasks 2, 3, 4, and 5 shows a considerable good level. Task 1 had moderate accuracy. As stated in Table 3 a strong positive correlation (Kendall's **τ**) between the self-assessment accuracy and the achieved points could be found for all tasks, although the correlation for Task 5 is not significant due to the low number of observations.

6 | DISCUSSION

Following the extensive research literature by Brown and Harris (2013), Brown et al. (2015), Ozarslan and Ozan (2017), Panadero et al. (2017) and many others, students should be given the opportunity to self-assess their knowledge during the learning process. Although it is widely recognized that students can only use self-assessment if their self-assessment is accurate enough, relatively little research attention has been paid to analysing in students' behaviour and accuracy in relation to the task design characteristics (Panadero et al., 2016). Therefore, this study was designed to provide more insight into the complexity of self-assessment. The findings of both the initial and replication studies yield a number of important discussion points.

Regarding the first research question on students' use of voluntary self-assessment tasks, the research findings of both studies are similar. Students engaged in self-assessment from the beginning of the semester, but this activity decreased as the semester progressed and the course content increased. This general pattern applies to all course-related activities in all computer science courses at our university and is often found in research literature on distance education (Geri et al., 2014). Furthermore, we found that students submitted both computer-edited and handwritten solutions, indicating the need to further support both traditional and digital processes of student work in modern learning management systems. Finally, although students were involved in an iterative process of self-assessment (submitting solutions, reviewing their work against the criteria, and revising their work if necessary), our data suggest that some students misbehave and tried to game the system by submitting nonsense solutions to collect sample solutions and feedback for self-assessment. These emerging behaviours are consistent with previous research on the topic warning that students exabit 'unconventional behaviour' (Yang et al., 2022), are 'not honest in self-assessments' (Brown et al., 2015, p. 13), and clearly reveal usability barriers caused by the mandatory submission of solutions and the one-way test sequence imposed by Moodle (Haake et al., 2020).

Regarding the second research question, on how well students self-assess their solutions, we investigated the accuracy of self-assessment by comparing students' self-assessment with the teacher's assessment and examined whether accuracy was related to the quality of the student's answer. The initial findings of Study 1 indicated that for three of the five tasks, most students were able to provide correct answers and correctly self-assess their solutions using assessment criteria. However, very low accuracy and low scores were observed for Task 5, so it was decided for Study 2 to adjust the task design characteristics. The data from the replication study, compared with the initial study, suggest very similar patterns of accuracy and score distribution for almost all tasks, except for the modified task (see Figure 9). A significant improvement of both dimensions (accuracy and score) was

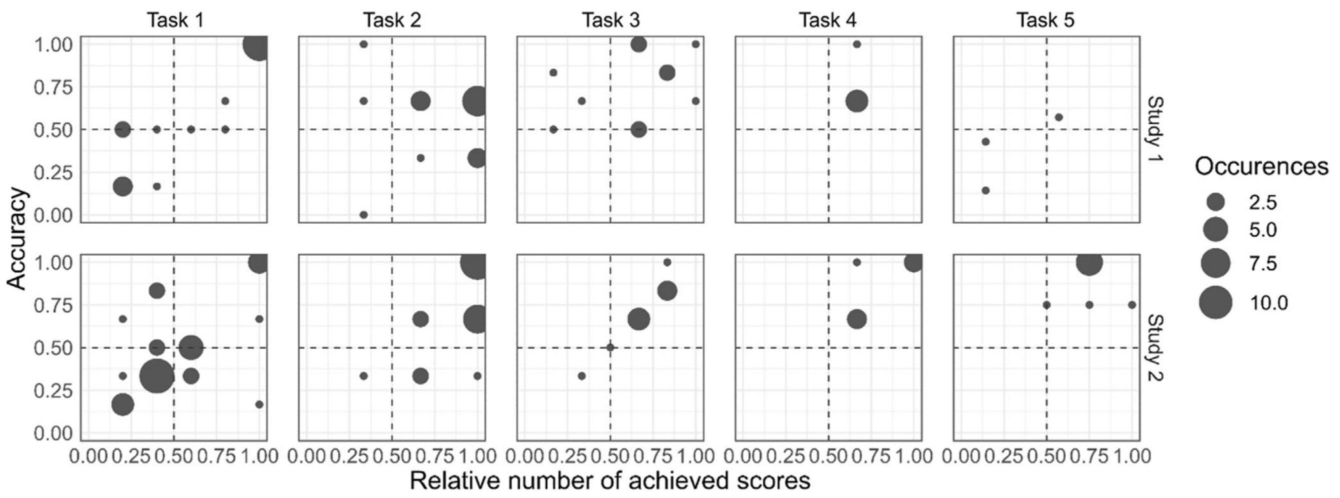

FIGURE 9 Self-assessment accuracy per student and task over the instructor-rated achieved points of a total of 108 edited/self-rated solutions to that task of Study 1 (top) and 88 Study 2 (bottom).

evident after changing the original task design. This is consistent with the seminal work of Bandura (1997), who emphasized that self-efficacy in general is task-specific. Taken together, task design (task complexity, scope, and transfer requirements as well as quantitative measures such as text length and number of evaluation criteria) can be modified to affect students' accuracy. To some extent, this provides an answer to the long-standing issue: 'to date, however, it is unclear what adjustments (e.g., criteria, instruction, etc.) teachers can make in light of the possibility of inaccurate student self-assessments' (Panadero et al., 2016, p. 15). Therefore, the following conclusion will provide recommendations distilled from the current two studies to help practitioners, designers and educators successfully facilitate an effective and efficient self-assessment process during learning.

Regarding the third research question, which characteristics of the self-assessment task affect self-assessment, we were able to identify the difficulty of the item and, relatedly, the comprehensiveness of the sample solution. Although a distinction should be made between the complexity and comprehensiveness of the sample solution, the complexity of sample solutions remains a challenge to be explored. Concerning self-assessment accuracy in Study 1 and 2 we observed similar results, except for the Task 5 that has been changed. Our results suggest that tasks that require a sample solution of small or moderate size are easier to self-assess than tasks requiring longer sample solutions. Moreover, fine-grained assessment criteria and distractors may help self-assessment. Regarding item difficulty, the research results of both studies indicate fairly similar difficulty values for the same tasks, except for the modified task. While Task 5 in Study 1 appears to have a very high difficulty level, the result of Study 2 showed a significant reduction in difficulty. Similar results are echoed in other studies, where it is noted that self-assessment accuracy decreases as task complexity increases (Andrade, 2019; Rohr, 2018). This effect may be a consequence of modifying the initial complexity of the task design, scope requirements and quantitative measures such as text length and number of evaluation criteria. The discrimination index as a measure for task characteristics was only partly helpful because the selected tasks did not represent a complete test and were not corresponding to the same topics (e.g., operation systems in CU 1 and 2 vs. computer networks in CU 3 and 4).

7 | LIMITATIONS

Several limitations of this study should be considered. First, our participants were students of a university providing distance education, so our findings hold limited generalizability to other levels of education and other face-to-face settings. Second, only 39.39% and 47.14% of the students enrolled in the courses of Study 1 and Study 2 respectively, participated in this research. The participation rate of both studies is similar to the typical participation rate in other online courses in computer science at the university in question. However, it is possible that we were able to target particularly motivated and therefore also high-performing students. Third, we were only able to include five of 43 self-assessment tasks in the study. By considering a greater number of different self-assessment tasks it would be possible to further examine correlations between item difficulty and the accuracy of self-assessment. Fourth, we tried to stimulate students' iterative self-assessment activity, but this was to be done on a mere voluntary basis, and students could not expect to receive any direct individual benefit or achieve any common purpose in doing so. We also noted that students tried to secure tasks' feedback and sample solutions for themselves, by submitting blank answers. Both factors and the mentioned usability barriers of the system may explain limited participation in the iterative process of self-assessment revision. Finally, while the instructor used a predefined binary rating schema for scoring the solutions, a second rate could be used to ensure the reliability of the scores and thus of the instructor accuracy.

8 | CONCLUSIONS

Taken together, the findings of this study suggest that self-assessment practice should be encouraged and move beyond simply providing a solution to a task to an iterative practice of improving the answer against the evaluation criteria and by presenting feedback for improving deficiencies. The results of both the initial and replication studies show the potential for scalable learning support. The subsequent set of five practical recommendations has been thoughtfully developed to aid educators in effectively and efficiently facilitating the process of self-evaluation within a learning context:

1. *Facilitate iterative self-assessment*: Ensure that learners are afforded the opportunity not only to assess their solutions against established criteria but also to revise and resubmit their solutions for further evaluation. This iterative process fosters a deeper engagement with the material and encourages a continuous refinement of understanding.
2. *Employ comprehensive criteria*: When constructing self-assessment tasks, consider integrating clear, specific, and well-defined criteria. Utilize finely grained criteria to enable students to discern even subtle advancements in their knowledge acquisition. This meticulous approach enhances the precision of self-evaluation.
3. *Incorporate distractors for enhanced understanding*: Incorporate distractors, which encompass plausible, yet incorrect responses rooted in common errors and misconceptions. This practice prompts learners to critically assess their understanding by differentiating between accurate and misleading choices.
4. *Balance task complexity through varied formats*: Embrace diversity in response formats, ranging from calculations and short answers to multiple-choice options. Additionally, account for the anticipated length of solution text, recognizing that tasks requiring more extensive explanations can pose increased challenges for accurate self-evaluation.
5. *Cultivate self-regulation through incentives*: Contemplate the integration of incentive mechanisms to promote active engagement in self-assessment activities. By fostering a sense of reward and accomplishment (e.g., activity badges, extra points for the exam for accurate solutions), students are encouraged to consistently engage in self-assessment and reflect upon their learning journey as the course progresses.

By implementing these practical strategies, educators can create a supportive environment that empowers learners to use self-assessment. Since supporting students learning is not a one-off activity, but part of a curriculum improvement and lifelong learning initiative, we recommend that self-assessment tasks are continuously monitored and, if necessary, improved to suit the needs of learners in a specific context. In this way, self-assessment can become a habit of the learning process rather than only a superficial and mandatory part of course design.

AUTHOR CONTRIBUTIONS

Slaviša Radović: Conceptualization; writing – original draft; writing – review and editing. **Niels Seidel:** Conceptualization; data curation; formal analysis; funding acquisition; methodology; project administration; resources; software; visualization; writing – original draft; writing – review and editing. **Joerg M. Haake:** Conceptualization; data curation; funding acquisition; methodology; resources; writing – original draft; writing – review and editing. **Regina Kasakowskij:** Conceptualization; software; writing – original draft; writing – review and editing.

FUNDING INFORMATION

This work was funded by the Center of Advanced Technology for Assisted Learning and Predictive Analytics (CATALPA) of the FernUniversität in Hagen.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/jcal.12907.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The use of the learning environment in this study was voluntary, but conditional on a two-step consent to use the platform and to participate in the study. The second informed consent for participation in the study could be withdrawn or granted again at any time, while the first consent was required for GDPR compliance. As an incentive for students' participation, additional exercises such as self-tests, self-assessments, and assignments were offered, as well as additional tools for semester planning and for reading the digital course texts. These differences in the learning offer are comparable to different didactic offers of tutors in face-to-face teaching. Students not participating in the study had no disadvantages regarding the examination, since the course texts provided to all enrolled students form the basis of the examination.

ORCID

Slaviša Radović https://orcid.org/0000-0001-8840-6053
Niels Seidel https://orcid.org/0000-0003-1209-5038
Joerg M. Haake https://orcid.org/0000-0001-9720-3100
Regina Kasakowskij https://orcid.org/0000-0003-2368-9851

REFERENCES

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, *32*(2), 159–181. https://doi.org/10.1080/02602930600801928

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*(87), 1–13. https://doi.org/10.3389/feduc.2019.00087

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: Disruptive effects of assessment patterns. *Higher Education Research & Development*, *34*(1), 45–59. https://doi.org/10.1080/07294360.2014.934328

Brown, G., Andrade, H., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*, 444–457. https://doi.org/10.1080/0969594x.2014.996523

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). SAGE.

Carroll, D. (2020). Observations of student accuracy in criteria-based self-assessment. *Assessment & Evaluation in Higher Education*, *45*(8), 1088–1105.

Castillo-Merino, D., & Serradell-López, E. (2014). An analysis of the determinants of students' performance in e-learning. *Computers in Human Behavior*, *30*, 476–484. https://doi.org/10.1016/j.chb.2013.06.020

Chen, I., & Zhang, J. (2023). The accuracy of students' self-assessment-how useful is a checklist in EAP writing. *English Language Teaching*, *16*(3), 23–31.

Ćukušić, M., Garača, Ž., & Jadrić, M. (2014). Online self-assessment and students' success in higher education institutions. *Computers & Education*, *72*, 100–109. https://doi.org/10.1016/j.compedu.2013.10.018

De Champlain, A. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, *44*(1), 109–117. https://doi.org/10.1111/j.1365-2923.2009.03425.x

Duque Micán, A., & Cuesta Medina, L. (2015). Boosting vocabulary learning through self-assessment in an English language teaching context. *Assessment & Evaluation in Higher Education*, *42*(3), 398–414. https://doi.org/10.1080/02602938.2015.1118433

Ernst, H. M., Wittwer, J., & Voss, T. (2023). Do they know what they know? Accuracy in teacher candidates' self-assessments and its influencing factors. *British Educational Research Journal*, *49*, 649–673. https://doi.org/10.1002/berj.3860

Geri, N., Gafni, R., & Winer, A. (2014). The u-curve of e-learning: Course website and online video use in blended and distance learning. *Interdisciplinary Journal of E-Learning and Learning Objects*, *10*, 1–16.

Haake, J. M., Ma, L., & Seidel, N. (2021). *Self-assessment questions—Operating systems and computer networks*. https://doi.org/10.5281/zenodo.5021350

Haake, J. M., Seidel, N., Buchart, M., Karolyi, H., & Kasakowskij, R. (2022). *Accuracy of self-assessments in higher education*. DELFI 2021–Die 19 (pp. 15–26). Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.

Haake, J. M., Seidel, N., Karolyi, H., & Ma, L. (2020). *Self-assessment mit high-information feedback*. DELFI 2020–Die 18 (pp. 145–150). Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.

Hamming, R. W. (1950). Error-detecting and error-correcting codes. *Bell System Technical Journal*, *29*(2), 147–160.

Hartung, S. (2017). Lernförderliches Feedback in der Online-Lehre gestalten. In H. R. Griese-hop & E. Bauer (Eds.), *Lehren und Lernen online* (pp. 199–217). Springer.

Ifenthaler, D., Schumacher, C., & Kuzilek, J. (2023). Investigating students' use of self-assessments in higher education using learning analytics. *Journal of Computer Assisted Learning*, *39*(1), 255–268. https://doi.org/10.1111/jcal.12744

Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice*, *2*(2), 145–163.

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, *35*(2), 135–156. https://doi.org/10.1080/02602930802687737

Max, A.-L., Lukas, S., & Weitzel, H. (2022). The relationship between self-assessment and performance in learning TPACK: Are self-assessments a good way to support preservice teachers' learning? *Journal of Computer Assisted Learning*, *38*(4), 1160–1172. https://doi.org/10.1111/jcal.12674

Ozarslan, Y., & Ozan, O. (2017). Self-assessment quiz taking behaviour analysis in an online course. *European Journal of Open, Distance and E-Learning*, *19*(2), 15–31. https://doi.org/10.1515/eurodl-2016-0005

Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*, 803–830. https://doi.org/10.1007/s10648-015-9350-2

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

Race, P. (2001). A briefing on self, peer and group assessment. *LTSN Generic Centre*. Assessment Series No. 9. https://scirp.org/reference/referencespapers.aspx?referenceid=2384289

Radović, S., Marić, M., & Passey, D. (2019). Technology enhancing mathematics learning behaviours: Shifting learning goals from "producing the right answer" to "understanding how to address current and future mathematical challenges". *Education and Information Technologies*, *24*(1), 103–126.

Rohr, V. (2018). *Der Zusammenhang zwischen akkurater Selbsteinschätzung und Performanz im wirtschaftswissenschaftlichen Kontext (Dissertation Wirtschaftswissenschaften)*. Universität Mannheim.

Steinkohl, K., Burchart, M., Haake, J. M., & Seidel, N. (2021). *Self-assess question type plugin for moodle*. https://github.com/D2L2/qtype_selfassess

Thawabieh, A. M. (2017). A comparison between students' self-assessment and teachers' assessment. *Journal of Curriculum and Teaching*, *6*(1), 14–20.

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, *45*(2), 224–238. https://doi.org/10.1080/02602938.2019.1629390

Yan, Z., & Brown, G. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education*, *42*(8), 1247–1262. https://doi.org/10.1080/02602938.2016.1260091

Yang, A., Chen, I., Flanagan, B., & Ogata, H. (2022). How students' self-assessment behavior affects their online learning performance. *Computers and Education: Artificial Intelligence*, *3*, 100058. https://doi.org/10.1016/j.caeai.2022.100058

Zhang, X. S., & Zhang, L. J. (2022). Sustaining learners' writing development: Effects of using self-assessment on their foreign language writing performance and rating accuracy. *Sustainability*, *14*, 14686. https://doi.org/10.3390/su142214686

How to cite this article: Radović, S., Seidel, N., Haake, J. M., & Kasakowskij, R. (2024). Analysing students' self-assessment practice in a distance education environment: Student behaviour, accuracy, and task-related characteristics. *Journal of Computer Assisted Learning*, *40*(2), 654–666. https://doi.org/10.1111/jcal.12907