International Journal of Assessment Tools in Education
2024, Vol. 11, No. 1, 1–19
https://doi.org/10.21449/ijate.1428025
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Student engagement with a novel online assessment strategy

Sally E. Jordan [1*], **John P.R. Bolton** [1]

[1]School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, United Kingdom

ARTICLE HISTORY

Received: Aug. 21, 2021

Accepted: Dec. 22, 2023

KEYWORDS:

Assessment strategy,
Computer-marked assessment,
Formative assessment,
Online assessment,
Student engagement.

Abstract: The study investigated the impact on student engagement and achievement of a "formative thresholded" continuous assessment strategy in which students had to meet a modest threshold, but their continuous assessment marks did not contribute to their final grade. Students were free to choose their own blend of tutor-marked and computer-marked assignments provided they met the threshold in a minimum number of assignments. For students reaching this threshold, the final grade was determined solely by an end-of-year examination. This assessment strategy was compared with one in which summative assignments contributed to the final grade alongside the final examination. The methodology took the form of data analysis of assessment records from more than 3500 anonymized students over a period of 7 years. Evidence was found for improved student engagement under the formative thresholded assessment strategy, with students choosing to complete many more computer-marked assignments, balanced by slightly fewer tutor-marked assignments. There was no change in overall retention. Many students made use of the capability to repeat different versions of the computer-marked questions. There was some evidence of students gaming the system (repeating questions until they obtained a version they had seen before), though most students did not do this at all and only a small number did so regularly. Students appeared to make better use of model answers, provided in response to three incorrect tries at a question, than of hints provided after each try. It was noted that the formative thresholded assessment strategy has potential for wider use.

1. INTRODUCTION

During the Covid-19 pandemic, much university teaching and assessment that previously took place in face-to-face settings was moved online. Despite some concerns, students were generally satisfied with the quality of the assessment they received (Şenel & Şenel, 2021). The adjustments made during the pandemic were remarkable, and it is now recognized that, beyond the immediate crisis response, there are longer-term lessons for sustainable online learning (Adedoyin & Soykan, 2023; Yang & Xin, 2022). While the early focus was on the immediate affordances brought by online assessment, it is now appropriate to consider longer term implications (St-Onge et al., 2022). In order to do this, it is necessary to move beyond a consideration of which assessment tools to use and instead to start from evidence-based assessment design and strategy (Morris et al., 2021).

*CONTACT: Sally Jordan sally.jordan@open.ac.uk School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, United Kingdom

e-ISSN: 2148-7456 /

This paper draws lessons from earlier (pre-Covid) data obtained at a university which has taught at a distance for more than 50 years and has made increasing use of online teaching and assessment through recent decades. In this environment, which much of the sector found itself engaging with seriously for the first time in 2020, it is particularly important that students are supported and well engaged in their learning. It has been argued that regular assignments, in particular computer-marked assignments with instantaneous and targeted feedback, might provide distance-learning students with the equivalent of a tutor sitting right beside them (Ross et al., 2006). For this reason, modules at the university in the study are usually assessed by way of both continuous assessment through the year and an end-of-module assessment, frequently an examination. The continuous assessment element has learning and motivation as its primary purpose, and comprehensive feedback is provided, but historically it usually also contributed towards a student's grade for the module and hence to their overall result. However, as discussed further in Section 2, when assessment has both formative and summative functions, the relationships between the two are often complex and sometimes confusing, so it was decided to trial a different assessment strategy.

The new "formative thresholded" assessment strategy studied here required students to reach a modest threshold on continuous assessment tasks, but the marks for these did not contribute to the final grade. The strategy was designed to enable students and their tutors to focus on the formative function of the assignments, freed from anxiety over the minutiae of the score received. The formative focus meant that concerns about plagiarism were reduced, so assignments could be re-used from year to year and computer-marked assignments could be repeated as many times as a student wished. The detailed feedback and indicative scores enabled students to monitor their progress and self-regulate. Thus, in several senses of the word, the assessment could be considered to be sustainable (Boud & Soler, 2016).

As Holmes (2018) suggests, assessment can be used to increase student engagement and motivation and thus to contribute to the quality of the overall student experience, but concern has been expressed that student engagement with assessment might decrease without the perceived incentive of a contribution towards a student's final module grade (Kibble, 2007). Thus, the key research question to be addressed was: What is the impact on student engagement and achievement of a modest threshold on assignments, compared with a previous situation in which some assignments had been summative and some had been purely formative?

The assignments under consideration were electronically submitted tutor-marked assignments and interactive computer-marked assignments. A variety of feedback was provided on both and, for the computer-marked assessment, the feedback was instantaneous, and students had an opportunity to try the questions again, bringing alignment with Gibbs and Simpson's (2005) conditions under which assessment supports learning. The secondary research question was thus: What use do students make of different types of feedback on computer-marked assignments, varying from hints to model answers, with the opportunity to attempt questions again?

The remainder of the paper is structured as follows. Section 2 reviews relevant literature. Section 3 gives further details of the computer-marked assignments and the formative thresholded assessment architecture in which they were embedded. This section also describes the datasets and outlines the detailed research questions addressed in the paper. Section 4 contains the results of the data analysis and Section 5 discusses the results and draws conclusions.

2. LITERATURE REVIEW

The literature underpinning the work we describe is wide-ranging and extensive. For this reason, our review is necessarily selective, concentrating on significant recent contributions, meta-analyses, and independent literature reviews.

2.1. Student Engagement

In recent years, much has been written about student engagement, especially in an online environment (Bond et al., 2020). The meaning of "student engagement" has evolved over time, with widely accepted components including time on task, quality of effort and involvement in productive learning activities (Kuh, 2009). Student engagement has been recognized as one of the most important drivers of academic success (Kahu, 2013) and is associated with improved achievement and retention (Kuh et al., 2008), with disengagement having negative consequences on a student's outcomes and being a predictor of drop-out (Finn & Zimmer, 2012; Ma et al., 2015). It is widely accepted that digital educational technologies offer the potential to improve engagement (Chen et al., 2010), but the relationships are complex (Rashid & Asghar, 2016) and care must be taken in the design and integration of the technology to ensure that it is effective (Bond & Bedenlier, 2019; Englund et al., 2017; Kirkwood, 2009). Guo et al. (2023) identified eight specific factors which can encourage engagement or act as a barrier to it.

2.2. A Role for Computer-Marked Assessment?

The term "e-assessment" can refer to any use of a computer as part of the assessment process and terms such as "digital assessment", "online assessment", "computer-based assessment" and "technology-enhanced assessment" are similarly broad. More specifically, "computer-marked assessment" refers to situations in which students' responses are automatically marked. It is important to note that, while early computer-marked assessment relied on students indicating their responses to multiple-choice questions on a machine-readable form, for marking at a later stage, by the 21st Century, the focus had become online computer-marked assessment, enabling instantaneous interaction between a student and the system on which the assessment sits (Jordan, 2023). At the same time, the range of question types that are available has extended beyond selected-response question types such as multiple-choice, multiple-response and drag-and-drop to include automatically marked constructed-response question types with answers comprising numbers and algebraic expressions (Sangwin, 2013; Sangwin & Harjula, 2017), words and sentences (Burrows et al., 2015; del Gobbo et al., 2023), and even essays (Ramesh & Sanampudi, 2022; Süzen et al., 2020).

The interactivity of modern computer-marked assessment brings potential for it to motivate and engage students (Holmes, 2015). Riegel & Evans (2021) report on positive emotions experienced by students following a computer-marked assessment, contrasting with the negative emotions experienced following a conventional quiz. Feedback can be provided instantaneously, and, in principle, students can then repeat the question or a similar one, thus enabling them to learn from the feedback while it is still fresh in their minds (Gibbs & Simpson, 2005; Nicol & McFarlane-Dick, 2006). Students report that they prefer receiving feedback from a computer because it is perceived to be unbiased, non-judgmental and impersonal (Sim et al., 2004) and enables them to make mistakes in private (Miller, 2008). However, the mere act of taking online tests has been shown to improve subsequent performance more than additional study of the material, even when the tests are given without feedback. This is the so-called "testing effect"; Roediger and Karpicke's (2006) review of work in this area has been updated in Roediger and Karpicke (2018) and Yang et al. (2019).

2.3. Importance of Assessment Strategy

Although computer-marked assessment brings potential to improve student engagement, it is not uniformly well received by either students or educators (Ghabraie, 2020; Kumar & Sajja, 2021). Factors that have been identified as significant include the choice of an appropriate question type for each item, and the need to take care in writing and checking questions (Benson & Brack, 2010; Jordan, 2023). However, here we concentrate on the importance of assessment design and strategy, which we take to include the relationship with the rest of the assessment on the module and program, whether the focus is formative or summative, and the way in which the computer-marked assessment operates (e.g. whether students can repeat questions and assignments). Nguyen et al. (2017) found that assessment design and strategy can have a significant impact on student engagement and pass rate, while highlighting the dangers inherent when assessment design alters between different but closely associated modules. It is also important that assessment is well aligned with learning activities (Barthakur et al., 2022).

Many authors have highlighted the complex interrelationships between the formative and summative functions of assessment (e.g. Lau, 2016). Where formative and summative functions co-exist, there is a danger that the purpose will be confused (Boud, 2000) and, in particular, that the summative function will dominate (Lipnevich & Smith, 2009). There may be poor alignment of student and staff perceptions, with staff seeing the purpose as primarily formative, while students are primarily concerned with the grade (Winstone & Boud, 2022). When assessment is delivered or marked by electronic means there can be further confusion and lack of trust (Rolim & Isias, 2019).

When assessment has a formative focus, concerns over plagiarism are reduced, which is particularly relevant given the growth in contract cheating (Ahsan et al., 2022) and in the use of generative artificial intelligence (Farrelly & Baker, 2023). In formative use, computer-marked assignments can be repeated multiple times, which means that students can act on the feedback provided by immediately revisiting the same question or a similar one (Lefevre & Cox, 2017). Formative assessment can help students to monitor their progress, encourage further study and increase their learning and understanding (McCallum & Milner, 2021).

3. METHOD

The investigation was based on a detailed analysis of student interactions with assessment collected over a period of seven years. The context in which the study took place is described in Section 3.1 while the data analysis methodology is described in more detail in Section 3.2.

3.1. Context

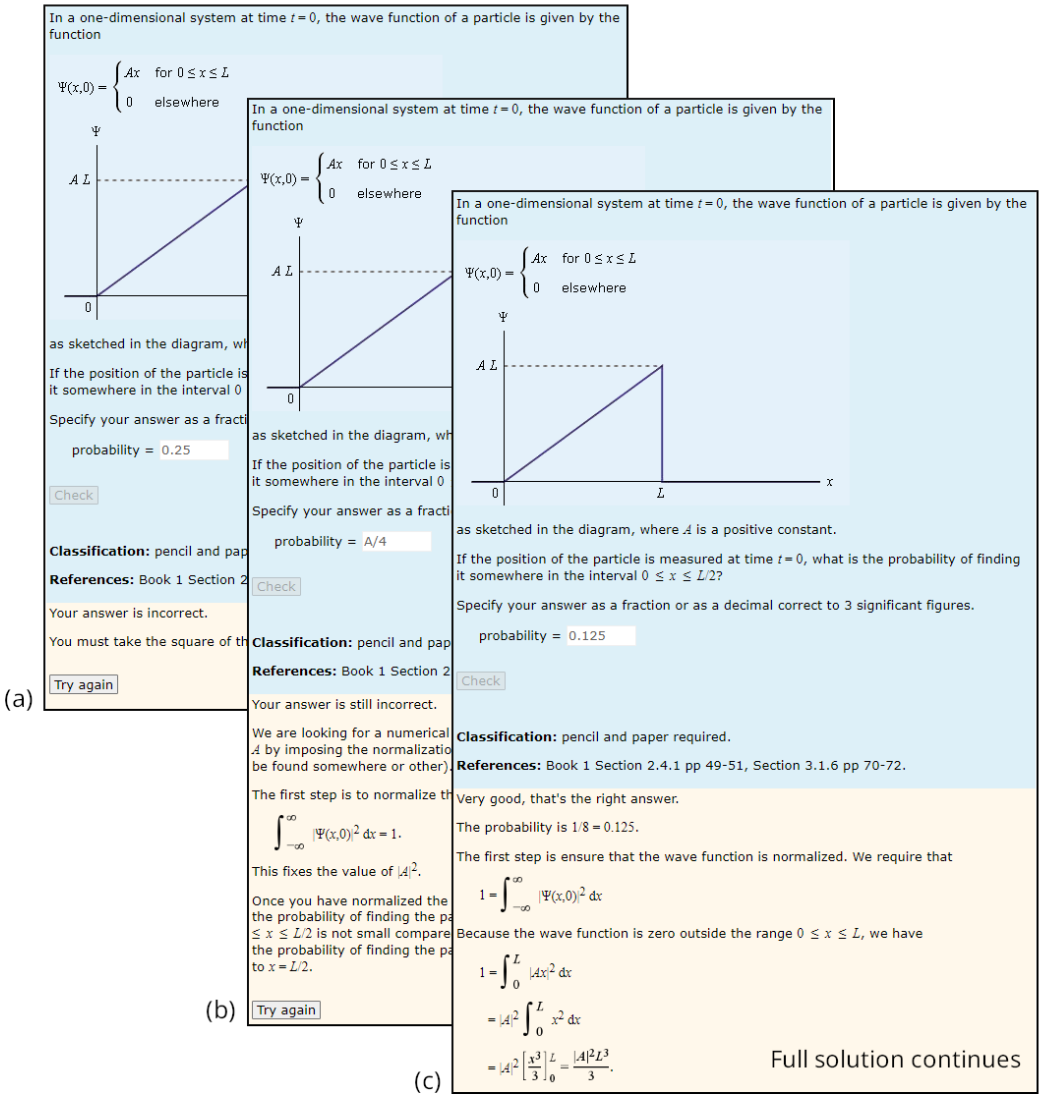
The investigation was conducted at the Open University, which is based in the UK, and focused on two of its upper-stage undergraduate modules: *The Quantum World* and *Electromagnetism*, key components of any physics degree. These modules piloted the formative thresholded assessment strategy prior to Faculty-wide adoption. Each of the modules counted for 30 points in the UK Credit Accumulation and Transfer Scheme (CATS), meaning that each was equivalent to a quarter of a year's study time for a full-time student, spread over a nine-month period. The modules could be studied alone or concurrently with each other or with other modules. Most Open University students study part-time alongside employment and/or caring responsibility and therefore study at an intensity of 30 or 60 CATS points per year. Students on the two modules investigated here had a wide range of ages from 18 to over 65, with the distribution peaking around age 30. The male: female ratio on both modules was about 4:1.

The Open University offers supported distance learning to students who may be studying anywhere in the world. Primary teaching resources are supplied in printed form and online, and students are supported by tutors who offer non-compulsory tutorials (usually online) as well as marking assignments and providing feedback comments electronically. Continuous assessment is completed in students' own homes and at a time of their choosing. Furthermore, although the students on the modules in question were strongly advised to study recommended lower-level mathematics and physics modules, the University's "open" mission extends to it having no entry qualifications and students could not be forced to study prerequisite modules.

The two modules shared the same overall assessment pattern. In each case, the continuous assessment consisted of four tutor-marked assignments and six interactive computer-marked assignments, and in order to pass the threshold, students were required to score more than 30% in seven of the ten assignments, two of which had to be tutor-marked assignments. Students passing the threshold could then sit a final examination, and their overall module grade was determined solely by the result of this.

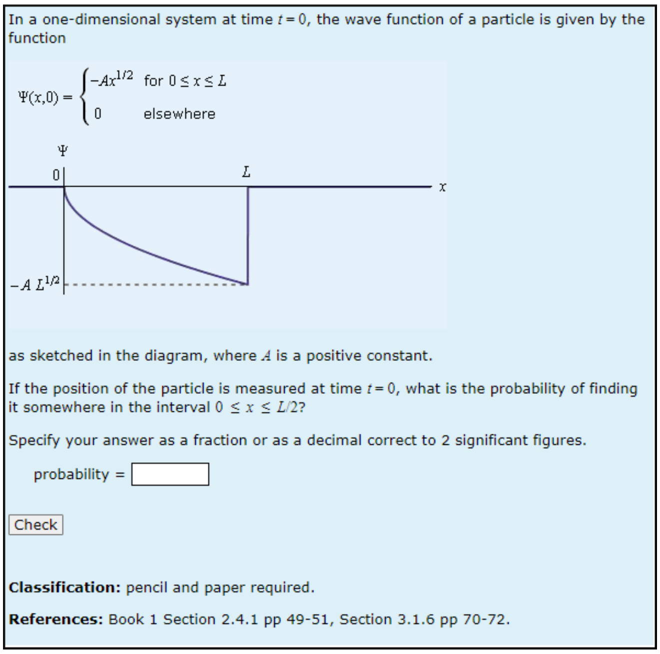
In guidance material, students were encouraged to use the continuous assessment as a resource to help them to develop skills and understanding, with the focus being on learning rather than marks. Each of the tutor-marked assignments included several questions assessing key physics learning outcomes, and required written answers from students, usually with extensive mathematical content. Tutors provided detailed feedback on each student's work as well as a summary of the main points requiring attention. Each of the six computer-marked assignments included eight questions, comprising a variety of question types, with constructed response questions (i.e. free-text entry) being favored over selected response question types such as multiple-choice. In line with their pedagogic function, the computer-marked assignments gave students many opportunities to re-try questions after receiving feedback. Each student received a particular version of each question and had three opportunities to get the correct answer. The opportunities were interspersed with increasingly detailed hints, wherever possible tailored to the errors made, and a fully worked solution was provided whenever a student completed a question, either by getting it right or by having had three tries. The functioning of a sample question is illustrated in Figure 1.

Figure 1. *A computer-marked question as attempted by a student, showing the tailored feedback generated after incorrect student responses at (a) the first try and (b) the second try; (c) the complete solution offered after a third incorrect try or a correct response.*



After completing a question, students could immediately make an attempt at a different version of the question. For brevity, we refer to attempting a fresh version as repeating a question. All versions of a question covered the same knowledge and skills, and were of similar difficulty, but they differed sufficiently to make students think afresh, even after seeing a model solution, as illustrated in Figure 2. Each question had between three and eight versions of similar difficulty, arranged in a cyclic order so that a student was presented with fresh versions for as long as is possible, until the cycle repeated itself. The initial version of each question was chosen at random so repeating an assignment generally led to a different set of question versions for a student to tackle. Students were encouraged to make fresh attempts until they were confident of having the relevant skills and knowledge. To promote this, the mark awarded for the question was that for the most recent attempt; the system therefore rewarded eventual success rather than any initial failure.

Figure 2. *A different version of the question illustrated in Figure 1.*


In a one-dimensional system at time $t = 0$, the wave function of a particle is given by the function

$$\Psi(x,0) = \begin{cases} -Ax^{1/2} & \text{for } 0 \le x \le L \\ 0 & \text{elsewhere} \end{cases}$$

as sketched in the diagram, where A is a positive constant.

If the position of the particle is measured at time $t = 0$, what is the probability of finding it somewhere in the interval $0 \le x \le L/2$?

Specify your answer as a fraction or as a decimal correct to 2 significant figures.

probability =

Check

Classification: pencil and paper required.

References: Book 1 Section 2.4.1 pp 49-51, Section 3.1.6 pp 70-72.

At any stage, students could click a "Finish" button, whether or not they had answered all the questions in the assignment. This generated an overall score and a summary report of their strengths and weaknesses. However, students could then tackle the whole assignment again, as many times as they wished and it was their highest overall score that was used to determine whether the threshold had been met. Although the threshold was modest, students were encouraged to aim well above this by submitting all the assignments and repeating individual questions and/or whole assignments until they achieved scores of around 75% or more.

3.2. Data Analysis

In this study, the extent of student engagement was measured by: the number of completed assignments, the number of times questions and entire assignments were repeated and the effects of these repetitions in boosting marks above the modest required threshold. The depth of engagement was evaluated by considering the extent to which hints and model answers were acted on. Student achievement was measured by the marks obtained in the continuous assessment and the examination, and by the overall retention rate.

Data were analyzed from a number of presentations of the two modules which, for *The Quantum World*, bridged the changes in the assessment strategy, as shown in Table 1. Unfortunately, the Open University's student population also altered, in particular as a result of significant changes to higher education funding in England in 2012. To minimize the impact of the changing student population, only presentations of the modules up to those starting in October 2013 and February 2014 were considered in the study. The dates were chosen so as to maximize the dataset while not considering presentations likely to have attracted students who had commenced their studies in October 2012 or later, acknowledging that students usually studied the modules in question towards the end of their degrees. These two modules were selected for the study because they used both tutor-marked and computer-marked assignments and were the initiators and first users of the new assessment strategy. In addition, it was reasonable to assume that by this stage in their study, most students would be familiar with the University's approach and would have taken the recommended prerequisite modules, although the open entry policy meant that this could not be forced. The modules had the same recommended pre-requisites and belonged to the same degree structures. Because of student choice over study paths and study intensity, the modules did not share a single cohort, though many students studied both modules, either concurrently or in different (usually consecutive) presentations.

Most of the reported analysis relates to the module *The Quantum World*, before and after the changes to its assessment strategy, with later presentations included for verification purposes. We have only included comparison with the module *Electromagnetism* where interesting differences were found between the modules. Where outcomes on subsequent presentations were found to be similar, data were combined to increase the size of the dataset.

Table 1. *Key features of the presentations of the modules included in the investigation.*

Module	Presentation abbreviation	Start date	Number of students	Characteristics of continuous assessment
The Quantum World	QW Old 1	Feb 2007	383	Summative tutor-marked assignments, no computer-marked assignments
	QW Old 2	Feb 2009	380	Summative tutor-marked assignments, purely formative computer-marked assignments
	QW New 1	Feb 2010	207	Formative thresholded tutor-marked assignments and computer-marked assignments
	QW New 2	Feb 2011	290	
	QW New 3	Feb 2012	243	
	QW New 4	Feb 2013	323	
	QW New 5	Oct 2013	365	
Electro-magnetism	EM New 1	Feb 2010	286	Formative thresholded tutor-marked assignments and computer-marked assignments
	EM New 2	Feb 2011	182	
	EM New 3	Feb 2012	333	
	EM New 4	Feb 2013	278	
	EM New 5	Feb 2014	313	

Several specific investigations were conducted, as outlined in the following sub-sections.

3.2.1. *Module completion rates*

The overall completion rates for students on *The Quantum World* were compared for the presentations before and after the change in assessment strategy i.e. the presentations abbreviated to QW Old 1 and QW Old 2 (see Table 1) were compared with QW New 1-5.

3.2.2. *Assignment submission rates and grades*

Analyses were conducted, again for *The Quantum World* over a range of presentations, into:

1. The impact of the switch to formative thresholded assessment on overall submission rates.
2. The extent to which students did more than the bare minimum required, as measured both by the number of assignments submitted and the grades achieved relative to the threshold.

3. Whether there was a relationship between the number of assignments completed and the student's final module result, based on examination performance. Note that we would expect there to be a strong correlation between the number of assignments submitted and the final examination mark, but this is not necessarily evidence of a causal link between these two factors; conscientious students are likely to study harder, resulting in both completion of more assignments and a better mark in the final examination. However earlier work (Bolton, 2010a) had shown that the single most important predictor of success in *The Quantum World* was performance on the recommended mathematics prerequisites. This allowed us to explore how completing a given combination of assignments increases or decreases a student's examination mark, on average, relative to other comparable students. The details of this analysis are given in Section 4.2.3.

3.2.3. *Repetition of computer-marked questions and the use of feedback*

Analyses were conducted into the extent to which students repeated computer-marked questions and the use they made of the feedback provided. Because some differences in behavior were noticed between the students on the different modules, *The Quantum World* and *Electromagnetism*, we report and discuss both, in considering:

1. The use made of hints by students prior to a second and/or third try at a question.
2. The extent to which students repeated complete questions (in different versions).
3. The extent to which students were "gaming the system" by repeating a question without making a serious attempt to answer it until they had seen the full answers to all versions of the question, at which point their next attempt was correct at the first try. As a proxy for this, we looked for a pattern of behavior which generated a repeated score of zero within the system for exactly the same number of attempts at the question as there were versions of that question, immediately followed by an answer that obtained full marks.

4. RESULTS

4.1. Module Completion Rates

60.3% of the students who commenced study of QW Old 1 (summative tutor-marked assignments and no computer-marked assignments) completed it, while 70.9% of the students who started the next presentation, QW Old 2 (summative tutor-marked assignments, purely formative computer-marked assignments), completed the module. This is a significant result ($p = 0.0034$ that such an outcome might arise by chance) and there is no reason to suspect that the student population changed significantly between the two presentations. It therefore seems likely that the introduction of the computer-marked assignments had a positive effect.

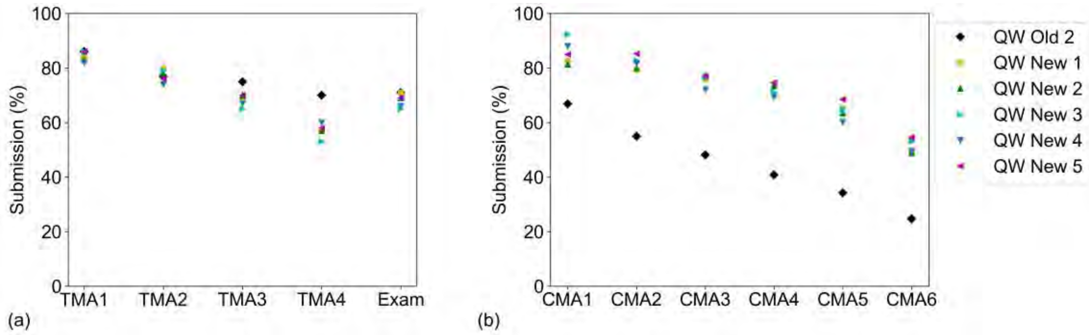
In contrast, the completion rates following the change to formative thresholded assessment (70.5% for QW New 1 and 69.5% for the 1221 students on all remaining presentations in the study) were very close to those for QW Old 2, indicating that the move to formative thresholded assessment had no significant impact on student retention. Completion and pass rates showed almost no correlation with age or gender.

4.2. Assignment Submission Rates and Grades

4.2.1. *Overall submission rates*

Figure 3 shows the assignment submission rates for QW Old 2 (summative tutor-marked assignments, purely formative computer-marked assignments) compared with the remaining five presentations in the study (formative thresholded tutor-marked and computer-marked assignments). The new thresholded assessment strategy gave students more flexibility than the previous strategy of summative tutor-marked assignments and optional computer-marked assignments. Figure 3 shows that students used this flexibility to complete many more computer-marked assignments, balanced by slightly fewer tutor-marked assignments. The number of students attending the examination was unaffected by the change in continuous assessment strategy.

Figure 3. *Submission rates for assignments on the module The Quantum World. Figure 3(a) shows the four tutor-marked assessments (TMA1 – TMA4) and the examination. Figure 3(b) shows the six computer-marked assignments (CMA1 – CMA6). The labeling of the different presentations of the module is as given in Table 1.*

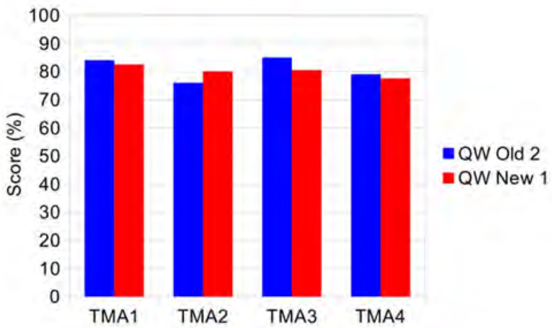


4.2.2. *Are students doing more than the bare minimum?*

Of the 1036 students who submitted and achieved the threshold of 30% in seven assignments, on the presentations of *The Quantum World* from QW New 1 to QW New 5, 865 (83.5%) did more than the minimum i.e. submitted at least eight assignments, and 499 (48.2%) of the students submitted all ten assignments available.

The scores recorded for assignments were also, on average, very much higher than the minimum of 30% required. Figure 4 illustrates that there was no consistent difference between the percentage scores for summative tutor-marked assignments (QW Old 2) and formative thresholded tutor-marked assignments (QW New 1), and the outcomes have been similar for all for subsequent presentations. Since QW New 1, mean scores of 80% or above have been observed on all computer-marked assignments, with a small but consistent increase in mean score in the years following the move to formative thresholded assessment.

Figure 4. *Mean percentage scores for tutor-marked assignments (TMA 1 – TMA 4) before and after the change in assessment strategy on the module The Quantum World.*



4.2.3. *Relationship between the number of assignments submitted and examination score*

Table 2 shows the mean examination marks for the students who were in each assignment submission category for QW New 1 and QW New 2 combined. Cells left blank in the table correspond to combinations that did not meet the continuous assessment threshold (so students could not obtain credit), though it is noteworthy that very few students who studied to the end of the module were in these categories. Non-zero entries along the uppermost diagonal correspond to the minimum allowed number of assignments with scores over 30%. Note, for example, that the 58 students who had submitted just three tutor-marked assignments (TMAs) and four computer-marked assignments (CMAs) had a mean examination mark of 49.4% while the 478 students who had submitted all four tutor-marked assignments and all six computer-marked assignments had a mean examination mark of 74.1%. As expected, students who did more assignments during the year also did better on the examination, but it would be dangerous to attribute this outcome to a simple causal link between engagement with continuous assessment and final examination score.

Table 2. *Mean examination score for students who successfully completed given combinations of assignments in The Quantum World. In each cell n is the number of students. Blank cells indicate combinations that fall short of the threshold requirement for sitting the examination.*

	2 TMAs	3 TMAs	4 TMAs
3 CMAs			55.8% (n = 20)
4 CMAs		49.4% (n = 58)	62.7% (n = 49)
5 CMAs	52.0% (n = 58)	56.2% (n = 54)	66.8% (n = 93)
6 CMAs	44.0% (n = 55)	53.9% (n = 88)	74.1% (n = 478)

Table 3 presents the results of the more subtle investigation based on the difference between actual and reference examination scores, averaged over students in each cell of the table. The reference score for each student is the average examination score for *all* students with the same grade in the mathematics prerequisite module. A previous study (Bolton, 2010a) showed that this is a strong predictor of final outcome. Because the analysis is based on a redistribution of marks, cells can have positive or negative values, averaging to zero across the whole student cohort. Cells with positive (or negative) values indicate combinations of assignments that deliver better (or worse) results than predicted on the basis of prior performance. For example, the 478 students who completed all available assignments had an average uplift in their examination score of 6.6 percentage points compared to the average for all students who matched their achievement in the prerequisite module. By contrast, the 58 students who submitted just three tutor-marked assignments and four computer-marked assignments did worse than their peers by 12.0 percentage points.

Table 3. *Average differences between actual and reference examination scores for students with given combinations of completed assignments. In each cell n is the number of students.*

	2 TMAs	3 TMAs	4 TMAs
3 CMAs			-7.0 (n = 20)
4 CMAs		-12.0 (n = 58)	+0.1 (n = 49)
5 CMAs	-8.7 (n = 58)	-7.8 (n = 54)	+2.4 (n = 93)
6 CMAs	-15.4 (n = 55)	-8.3 (n = 88)	+6.6 (n = 478)

4.3. Repetition of Computer-Marked Questions and Use of Feedback

4.3.1. *Use of hints*

Table 4 shows the percentage of all computer-marked question attempts which were correct at the first, second and third try. Overall, 61.4% of question attempts on *The Quantum World* and 51.7% of question attempts on *Electromagnetism* were correct by the third try. It is noteworthy that most correct question attempts were correct at the first try, without any need for hints. For students who initially got the answer wrong, the first hint was more effective than the second, though many students failed to benefit from either. It is also noteworthy that, on average, students were less likely to get the correct answer for *Electromagnetism* questions, even after three tries.

Table 4. *Percentage of responses that were correct at first, second and third try, for all available questions on The Quantum World (QW) and Electromagnetism (EM).*

	QW (as % of all question attempts)	QW (as % of those making this try)	EM (as % of all question attempts)	EM (as % of those making this try)
Correct at 1st try	43.0	43.0	34.3	34.3
Correct at 2nd try	12.9	22.7	12.2	18.5
Correct at 3rd try	5.5	12.4	5.2	9.8
Correct at 1st, 2nd or 3rd try	61.4		51.7	

Table 5 illustrates the extent to which the feedback provided in the hints appears to have been ignored, as implied by student responses being identical from one try to the next. This behavior is surprisingly common, and indeed a total of 20.8% of all student attempts at questions on *The Quantum World* and 28.7% of all student attempts at *Electromagnetism* questions offered three identical responses. Interestingly, while for *The Quantum World* most of these instances comprised situations where a response had been entered at first try but was not then subsequently altered, for *Electromagnetism* it was more common for students to enter no response at all.

Table 5. *Percentage of responses that were identical at first, second and third try, for all available questions on The Quantum World (QW) and Electromagnetism (EM).*

	QW (as % of all question attempts)	QW (as % of those making this try)	EM (as % of all question attempts)	EM (as % of those making this try)
Identical response at 1st & 2nd try.	21.7	38.1	30.4	46.3
Identical response at 2nd & 3rd try.	24.9	56.5	34.1	63.7
Three blank responses.	6.4	14.4	19.4	36.3
Three identical but non-blank responses.	14.4	32.7	9.3	17.4

4.3.2. *Repetition of questions*

We now report the extent to which students made repeated attempts at different versions of whole questions. For *The Quantum World*, over all questions and all presentations, the mean number of attempts at each question was 2.1, while for *Electromagnetism* the mean number of attempts was 2.5. However, these figures mask huge variation between different questions and different students. Unsurprisingly, questions which had a lower mean first attempt score were more likely to be repeated than higher scoring questions. It is also noteworthy that the mean number of attempts was very slightly higher for constructed-response questions, in which students construct their own answer (2.1 for *The Quantum World* and 2.6 for *Electromagnetism*) than for selected-response questions such as multiple-choice, in which students select from options provided in the question (2.0 for *The Quantum World* and 2.3 for *Electromagnetism*).

Figure 5 shows the overall distribution of question attempts for students on QW New 5 and EM New 5. Most students attempted most questions just once, but a small number of students attempted questions very many times. Note the very different pattern of Figure 5(a) (for *The Quantum World*) and Figure 5(b) (for *Electromagnetism*). For *Electromagnetism*, a larger number of students attempted the questions precisely four times than would be expected from the overall trend. Since all *Electromagnetism* questions have three versions, this may be as a result of students "gaming the system" i.e. writing down the correct answer to the first version they received then waiting for the same version to appear again. This point is further explored in the next section. *The Quantum World* questions had a variable number of versions.

Figure 5. *Distribution of the number of attempts made by all students on all questions for (a) QW New 5 and (b) EM New 5.*

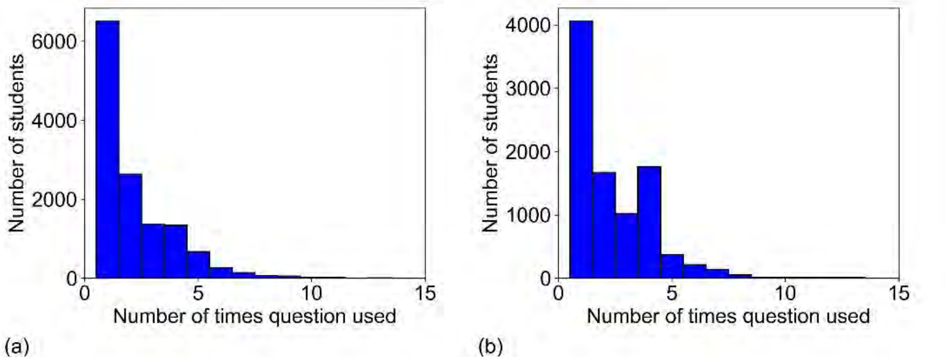
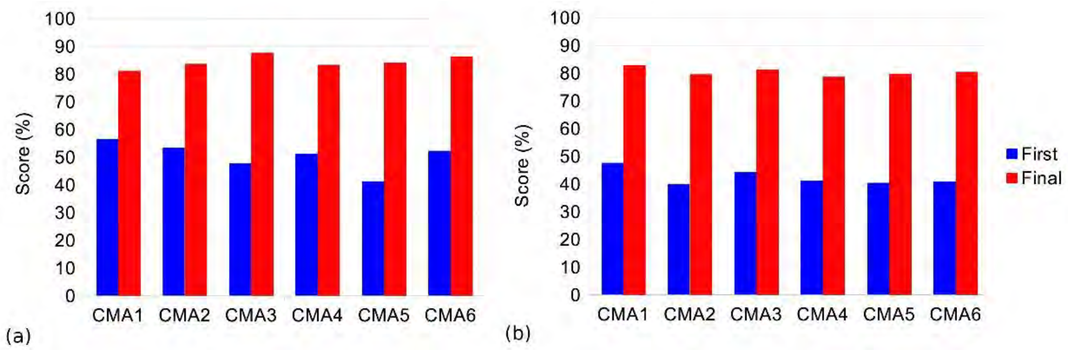


Figure 6 illustrates the impact of repeating questions on the final scores for each computer-marked assignment. Students on both modules appeared to have taken very seriously the advice that they should aim for at least 75%, and they repeated questions to achieve this. The first attempt score for a typical assignment had a mean of around 50% for *The Quantum World* and around 45% for *Electromagnetism* (shown by the blue bars in Figure 6); this increased to around 80% for both modules on the final attempt (shown by the red bars).

Figure 6. *The boost in marks achieved by repeating computer-marked questions in CMA1-CMA6 for (a) The Quantum World (data from QW New 2 to QW New 5 combined) and (b) Electromagnetism (data from EM New 2 to EM New 5 combined.) Blue bars show the first attempt, red bars the last.*



4.3.3. *Gaming the system*

From a baseline of 566 registered students on the QW New 3 and QW New 4 presentations of *The Quantum World*, 65.5% *never* exhibited the behavior of repeating questions for exactly the same number of times as there are versions and then getting the question correct immediately afterwards (a signature of "gaming the system"). The corresponding percentage from a baseline of 611 registered students on the equivalent presentations of *Electromagnetism* was 52.8%. A further 9.8% of the students on *The Quantum World* and 11.2% of those on *Electromagnetism* exhibited this behavior on just one of the 48 questions. A minority of students exhibited the behavior more widely. On *The Quantum World*, 12.9% did so for 5% of the questions, 6.9% for 10% of the questions and 1.8% for more than half of the questions. The corresponding percentages for the *Electromagnetism* module are slightly larger: 19.8%, 9.1% and 1.8%.

5. DISCUSSION and CONCLUSIONS

We were seeking to investigate the impact on student engagement of allowing students freedom in choosing a blend of tutor-marked and computer-marked assignments, subject to modest thresholds on grades. This compared with a more rigid previous system based on summative tutor-marked and formative computer-marked assignments. At the whole module level, a critical measure of student engagement is whether students are retained to the end of the module. If the change in assessment strategy had resulted in a reduction in retention, this would have been a real cause for concern, while an increase in retention would have been pleasing. However, there was no significant change in retention at all.

In contrast, the introduction of computer-marked assignments in the first place coincided with a significant increase in retention for that presentation of *The Quantum World*, and the improved retention rate was maintained through the change in assessment strategy. This is particularly pleasing in the light of the Open University's open entry policy. Care must be taken in attributing causality for the improvement, because other changes were made at the same time as computer-marked assignments were introduced, but a student survey conducted at that time, to which 53 students responded, illustrated strong student satisfaction with the computer-marked assignments (Bolton, 2010b). 83% of respondents agreed that computer-marked assignments helped them to acquire skills and knowledge, while 79% agreed that computer-marked assignments helped them to prepared for the examination, and 75% agreed that computer-marked assignments helped them to understand what they needed to study further. Although feedback from humans, in this case the students' tutors, is inevitably more flexible than the automated hints provided by the computer-marked assignments, only a minority of respondents (42%) agreed that tutor-marked assignments taught them more than computer-marked assignments. The positive reaction to computer-marked assessment is in line with the findings of Holmes (2015) and Ghabraie (2020).

When a modest threshold was introduced, most students were observed to do considerably more than was required, both in terms of the number of assignments submitted and as measured by the marks obtained for each assignment (noting that, provided the threshold had been reached, the mark did not contribute to students' results). Most starkly, the computer-marked assignment submission rate increased markedly when the threshold was introduced, and students repeated computer-marked questions in order to achieve an improvement in their score, leading to a marked increase in the mean score for each assignment. By these measures it appears that the modest threshold, accompanied by advice to aim for a much higher score, was effective in encouraging engagement.

After considering other reasons for the correlation between the number of assignments submitted and examination score, it appears that submitting more assignments had a genuinely positive impact on final outcome for most students. This contrasts with Kibble's (2007) finding that a significant number of students who scored 100% on quizzes in response to incentives (credit of between 0.5% and 2% per quiz) did not subsequently perform well on summative examinations. The quizzes, like ours, had a primarily formative function and Kibble describes the student usage as "inappropriate", with a suspicion that students had copied from their peers, meaning that they were taking approaches in which "quiz points are scored, but which circumvent learning" (Kibble, 2007, p. 259).

Less positively, on both of the modules in our study, but in particular *Electromagnetism*, it seems that many students were not making as much use as had been hoped of the hints provided after one or two unsuccessful tries at computer-marked questions, relying rather more on the full answers provided after three unsuccessful tries. This result is in line with a survey in which students on *The Quantum World* were asked to rank eight features of the computer-marked assignments in terms of their helpfulness: "Being given a full solution" was reported as the most helpful aspect by 50% of the 53 respondents, compared with "Being given hints when my answer was incorrect" which was only considered to be the most helpful aspect by 23% of respondents (Bolton, 2010b, p. 85). These findings contrast with those from another Open University module, where hints were found to be more useful than the full answers (Jordan, 2011). However, that module used a very different assessment strategy to that considered here: the computer-marked assignments were summative and students were not able to repeat complete questions, meaning that students were not able to act on the final answers given (Gibbs

& Simpson, 2005). This serves to illustrate the importance of assessment strategy in determining the outcome of a change in assessment practice (Nguyen et al., 2017).

Some interesting insights can be obtained by comparing the use that was made of feedback on the two modules in the study. *Electromagnetism* students were less likely to get the correct answer after three tries, and considerably more likely to enter a blank response at each try, though students on *The Quantum World* were more likely to enter identical but not blank responses. *Electromagnetism* students were also more likely to make repeated question attempts, more likely to make precisely four attempts at questions, and more likely to repeat questions with zero score until one attempt more than the number of versions available, at which point they got a previously seen version of the question correct. There are several possible explanations for this differing behavior, each supported by some evidence, so all the explanations are worthy of consideration and of further investigation.

Firstly, *Electromagnetism* was generally considered to be a more difficult module than *The Quantum World*. We have already noted that questions with a lower mean first attempt score were more likely to be repeated than higher scoring questions, and students were plausibly repeating the difficult questions in order to learn from the full answers provided after three unsuccessful tries.

It has been noted elsewhere that students leave blank responses when they do not understand the question or the feedback provided (Jordan, 2014). Prior to the study reported here, student responses to the computer-marked assessment questions on *The Quantum World* had been analyzed in detail which had led to the removal of some questions, the rewording of others and additional tailored feedback for common incorrect responses. This detailed analysis had not taken place for *Electromagnetism*, and the difference in question behavior on the two modules points towards the importance of writing high quality questions, monitoring their use, and iterative development (Benson & Brack, 2010; Jordan, 2023).

It is to be expected that students will be more likely to enter a response of some sort rather than leaving a blank response when the answer is easier to guess. Therefore, we would expect more blank responses to constructed-response questions than to selected-response questions, in which students are provided with options from which to guess. Overall, 18.8% of the *Electromagnetism* questions are selected response, while 30.0% of *The Quantum World* questions are selected response.

Despite some concern that students were "gaming the system", for both *The Quantum World* and *Electromagnetism*, the majority of students never repeated questions for exactly the same number of times as there were different versions and then got the answer correct immediately afterwards. Where students did exhibit this behavior, it was usually just on a small number of questions, with only a handful of students doing so on most questions. It seems likely that the behavior was a "helpless reaction" because they did not know how to proceed by other means (Jordan, 2014). One of the factors that has been associated with increased incidence of cheating is a student's perceived inability to complete the assessment task themselves to the standard they feel is required (Sevnarayan & Bridget Maphoto, 2024). It is reassuring that, in general, students who occasionally in early assignments repeated questions unsuccessfully until they obtained a version that they had seen before did not then go on to do so regularly. This gives some optimism that most students appreciated that the assessment's formative focus meant that obtaining the right answer by this method was only cheating themselves (Ashworth et al., 1997). Although the problem is a relatively minor one, it is worth monitoring questions that seem to provoke this behavior and make changes where necessary. It is also advisable for questions to have a variable number of versions.

5.1. Limitations and Future Work to Address These

A known limitation of the study is its reliance on data gathered from a changing student population. We selected student cohorts that were as stable as possible, but we could not entirely remove the risk of impact. In addition, while we hypothesize plausible explanations for some of the different student behaviors on the two modules in the study, it has not been possible to test these hypotheses. The physics curriculum at the Open University is currently being redeveloped, providing an opportunity for further investigation, on a more stable student population, into the impact of assessment strategy on student engagement. However, as the educational community considers future assessment practice beyond the Covid-19 pandemic, we also encourage colleagues in other institutions to conduct similar investigations into the effectiveness of formative thresholded assessment and also to research the impact of a range of apparently minor factors, such as the level of the students and the guidance provided to them on assessment strategies.

5.2. Conclusions

In response to our first research question, we found no detrimental impact following the introduction of modest thresholds for tutor-marked and computer-marked assignments, compared with a previous situation in which some assignments had been summative and some had been purely formative. There was some evidence of improved engagement under the revised assessment strategy, but no change to overall retention on the modules included in the study. Computer-marked questions could be repeated, using a number of versions, as many times as a student wished prior to the deadline, and students were found to repeat questions to improve their score, well beyond what was required in order to reach the required threshold.

In answer to our second research question, students were found to make better use of the model answer provided after three tries than the hints provided after each unsuccessful try. There was evidence that on some occasions some students repeated a question until they obtained a version that they had seen previously, but most students exhibited this behavior not at all or on a very small number of occasions.

More generally, the study provided evidence of the need to address assessment strategy in addition to question type when moving toward online assessment. In addition, it has highlighted the importance of considering detail when designing assessment, as apparently minor factors can have a disproportionate effect on student engagement.

Acknowledgments

The authors gratefully acknowledge the module teams of the modules under consideration and the support of eSTEeM, the Open University's Centre for STEM Pedagogy. Some of the early findings were presented at the 8th EDEN Research Workshop, Oxford, UK, October 2014.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research and publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Authorship Contribution Statement

Sally E. Jordan: Provided wider context within the Open University and externally, some data analysis, drafted paper. **John P.R. Bolton**: Led the investigation into student engagement and attitudes to assessment on the modules in the study, data collection and most of analysis.

Orcid

Sally E. Jordan https://orcid.org/0000-0003-0770-1443
John P.R. Bolton https://orcid.org/0009-0004-7191-6283

REFERENCES

Adedoyin, O.B., & Soykan, E. (2023). Covid-19 pandemic and online learning: The challenges and opportunities. *Interactive Learning Environments*, *31*(2), 863-875. https://doi.org/10.1080/10494820.2020.1813180

Ahsan, K., Akbar, S., & Kam, B. (2022). Contract cheating in higher education: A systematic literature review and future research agenda. *Assessment & Evaluation in Higher Education*, *47*(4), 523-539. https://doi.org/10.1080/02602938.2021.1931660

Ashworth, P., Bannister, P., & Thorne, P. (1997). Guilty in whose eyes? University students' perceptions of cheating and plagiarism in academic work and assessment. *Studies in Higher Education*, *22*(2), 187-203. https://doi.org/10.1080/03075079712331381034

Barthakur, A., Joksimovic, S., Kovanovic, V., Richey, M., & Pardo, A. (2022). Aligning objectives with assessment in online courses: Integrating learning analytics and measurement theory. *Computers & Education*, 190, 104603. https://doi.org/10.1016/j.compedu.2022.104603

Benson, R., & Brack, C. (2010). *Online learning and assessment in higher education: A planning guide*. Chandos Publishing.

Bolton, J. (2010a). CHASE: Prior course grades as indicators of success and failure in SM358. In J. Bolton (Ed.), *e-Tutorials and Learner Support* (pp. 124-140). The Physics Innovations Centre for Excellence in Teaching and Learning. https://www5.open.ac.uk/scholarship-and-innovation/esteem/resources/picetl-book-reports-e-tutorials-and-learner-support-2010

Bolton, J. (2010b). Using interactive computer-marked assignments in level 3 quantum mechanics. In S. Jordan (Ed.), e-Assessment (pp. 82-87). The Physics Innovations Centre for Excellence in Teaching and Learning. https://www5.open.ac.uk/scholarship-and-innovation/esteem/resources/picetl-book-reports-e-assessment-2010

Bond, M., & Bedenlier, S. (2019). Facilitating student engagement through educational technology: Towards a conceptual framework. *Journal of Interactive Media in Education*, 2019(1), 1-14. https://doi.org/10.5334/jime.528

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. *International Journal of Educational Technology in Higher Education*, *17*(1), Article 2. https://doi.org/10.1186/s41239-019-0176-8

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151-167. https://doi.org/10.1080/713695728

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413. https://doi.org/10.1080/02602938.2015.1018133

Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, *25*, 60-117. https://doi.org/10.1007/s40593-014-0026-8

Chen, P.-S.D., Lambert, A.D., & Guidry, K.R. (2010). Engaging online learners: The impact of web-based learning technology on college student engagement. *Computers & Education*, *54*(4), 1222-1232. https://doi.org/10.1016/j.compedu.2009.11.008

del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., & Limone, P. (2023). Automatic evaluation of open-ended questions for online learning. A systematic mapping. *Studies in Educational Evaluation*, *77*, 101258. https://doi.org/10.1016/j.stueduc.2023.101258

Englund, C., Olofsson, A.D., & Price, L. (2017). Teaching with technology in higher education: Understanding conceptual change and development in practice. *Higher Education Research and Development*, *36*(1), 73-87. https://doi.org/10.1080/07294360.2016.1171300

Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, *13*(11), 1109. https://doi.org/10.3390/educsci13111109

Finn, J.D., & Zimmer, K.S. (2012). Student engagement: What is it? Why does it matter?. In S. Christenson, A.L. Reschly, C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 97-131). Springer.

Ghabraie, K. (2020). Computer-marked assessments to enhance learning in engineering education. *International Journal on Innovations in Online Education*, *4*(1). https://doi.org/10.1615/IntJInnovOnlineEdu.2020033192

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, *1*, 3-31. http://eprints.glos.ac.uk/3609/

Guo, Y., Liu, Z., Meng, X., & Yin, H. (2023). Unravelling the relationship between student engagement and learning outcomes in emergency online learning: A synthesis of quantitative and qualitative results. *Assessment & Evaluation in Higher Education*, *48*(8),1325-1338. https://doi.org/10.1080/02602938.2023.2214345

Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment & Evaluation in Higher Education*, *40*(1), 1-14. https://doi.org/10.1080/02602938.2014.881978

Holmes, N. (2018). Engaging with assessment: Increasing student engagement through continuous assessment. *Active Learning in Higher Education*, *19*(1), 23-34. https://doi.org/10.1177/1469787417723230

Jordan, S. (2011). Using interactive computer-based assessment to support beginning distance learners of science. *Open Learning: The Journal of Open, Distance and e-Learning*, *26*(2), 147-164. https://doi.org/10.1080/02680513.2011.567754

Jordan, S. (2014). Using e-assessment to learn about students and learning. *International Journal of eAssessment*, *4*(1), 1-16. https://oro.open.ac.uk/40579

Jordan, S. (2023). Computer-marked assessment and concept inventories. In A.K. Wood (Ed.), *Effective Teaching in Large STEM Classes* (pp. 6-1 to 6-20). IOP Publishing. https://doi.org/10.1088/978-0-7503-5231-4ch6

Kahu, E.R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*(5), 758-773. https://doi.org/10.1080/03075079.2011.598505

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: Effects of incentives on student participation and performance. *Advances in Physiology Education, 31*(3), 253-260. https://doi.org/10.1152/advan.00027.2007

Kirkwood, A. (2009). E-learning: You don't always get what you hope for. *Technology, Pedagogy and Education*, *18*(2), 107-121. https://doi.org/10.1080/14759390902992576

Kuh, G.D. (2009). The National Survey of Student Engagement: Conceptual and empirical foundations. *New Directions for Institutional Research*, *141*, 5-20. https://doi.org/10.1002/ir.283

Kuh, G.D., Cruce, T.M., Shoup, R., Kinzie, J., & Gonyea, R.M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education*, *79*(5), 540-563. https://doi.org/10.1080/00221546.2008.11772116

Kumar, D. & Sajja, R. (2020). Qualifying online assessment during COVID-19 pandemic: Reflecting on our experience under the cognitive lens of Miller's pyramid. *Research and Development in Medical Education*, *9*(1), 1-2. https://rdme.tbzmed.ac.ir/Article/rdme-31855

Lau, A.M.S. (2016). "Formative good, summative bad?": A review of the dichotomy in assessment literature. *Journal of Further and Higher Education, 40*(4), 509-525. https://doi.org/10.1080/0309877X.2014.984600

Lefevre, D., & Cox, B. (2017). Delayed instructional feedback may be more effective, but is this contrary to learners' preferences? *British Journal of Educational Technology*, *48*(6), 1357-1367. https://doi.org/10.1111/bjet.12495

Lipnevich, A.A., & Smith, J.K. (2009). "I really need feedback to learn": Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*, *21*(4), 347-367. https://link.springer.com/content/pdf/10.1007/s11092-009-9082-2.pdf

Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, *24*, 26-34. https://doi.org/10.1016/j.iheduc.2014.09.005

McCallum, S., & Milner, M.M. (2021). The effectiveness of formative assessment: student views and staff reflections. *Assessment & Evaluation in Higher Education*, *46*(1), 1-16. https://doi.org/10.1080/02602938.2020.1754761

Miller, T. (2009). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, *34*(2), 181-192. https://doi.org/10.1080/02602930801956075

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3), e3292. https://doi.org/10.1002/rev3.3292

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior*, *76*, 703-714. https://doi.org/10.1016/j.chb.2017.03.028

Nicol, D.J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Ramesh, D., & Sanampudi, S.K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review*, *55*(3), 2495-2527. https://doi.org/10.1007/s10462-021-10068-2

Rashid, T., & Asghar, H.M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior*, *63*, 604-612. https://doi.org/10.1016/j.chb.2016.05.084

Riegel, K., & Evans, T. (2021). Student achievement emotions: Examining the role of frequent online assessment. *Australasian Journal of Educational Technology*, *37(*6), 75-87. https://doi.org/10.14742/ajet.6516

Roediger, H.L., & Karpicke, J.D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H.L., & Karpicke, J.D. (2018). Reflections on the resurgence of interest in the testing effect. *Perspectives on Psychological Science*, *13*(2), 236-241. https://doi.org/10.1177/1745691617718873

Rolim, C., & Isaias, P. (2019). Examining the use of e-assessment in higher education: Teachers' and students' viewpoints. *British Journal of Educational Technology*, *50*(4), 1785-1800. https://doi.org/10.1111/bjet.12669

Ross, S., Jordan, S., & Butcher, P. (2006). Online instantaneous and targeted feedback for remote learners. In C. Bryan, & K. Clegg (Eds.), *Innovative Assessment in Higher Education* (pp. 123-131). Routledge.

Sangwin, C.J. (2013). *Computer Aided Assessment of Mathematics*. Oxford University Press.

Sangwin, C.J., & Harjula, M. (2017). Online assessment of dimensional numerical answers using STACK in science. *European Journal of Physics*, *38*(3), 035701. https://iopscience.iop.org/article/10.1088/1361-6404/aa5e9d/

Şenel, S., & Şenel, H. (2021). Remote assessment in higher education during COVID-19 pandemic. *International Journal of Assessment Tools in Education*, *8*(2), 181-199. https://doi.org/10.21449/ijate.820140

Sevnarayan, K., & Bridget Maphoto, K. (2024). Exploring the dark side of online distance learning: Cheating behaviours, contributing factors, and strategies to enhance the integrity of online assessment. *Journal of Academic Ethics*, 1-20. https://doi.org/10.1007/s10805-023-09501-8

Sim, G., Holifield, P., & Brown, M. (2004). Implementation of computer assisted assessment: Lessons from the literature. *ALT-J*, *12*(3), 215-229. https://doi.org/10.3402/rlt.v12i3.11255

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2022). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology*, *53*(2), 349-366. https://doi.org/10.1111/bjet.13169

Süzen, N., Gorban, A.N., Levesley, J., & Mirkes, E.M. (2020). Automatic short answer grading and feedback using text mining methods. *Procedia Computer Science*, *169*, 726-743. https://doi.org/10.1016/j.procs.2020.02.171

Winstone, N.E., & Boud, D. (2022). The need to disentangle assessment and feedback in higher education. *Studies in Higher Education*, *47*(3), 656-667. https://doi.org/10.1080/03075079.2020.1779687

Yang, B.W., Razo, J., & Persky, A.M. (2019). Using testing as a learning tool. *American Journal of Pharmaceutical Education*, *83*(9), Article 7324. https://doi.org/10.5688/ajpe7324

Yang, L.P., & Xin, T. (2022). Changing educational ssessments in the post-COVID-19 era: From assessment of learning (AoL) to assessment as learning (AaL). *Educational Measurement: Issues and Practice*, *41*(1), 54-60. https://doi.org/10.1111/emip.12492