# *Strategic Engagement: Exploring Student Buy-in across a Formative and Summative Assessment.*

Suzanne Fergus, Ektaa Hirani, Nawal Parkar, Stewart Brian Kirton

**University of Hertfordshire,** s.fergus@herts.ac.uk.

## Abstract.

Assessment greatly influences the approaches students take in their learning. The nature of the assessment task, its educational value, outcomes and perceived importance can influence student engagement. PeerWise, a freely available online tool enables a highly interactive student-led assessment activity that promotes independent learning and peer-feedback. The level and quality of student engagement across a formative and summative PeerWise assignment in two 1st year chemistry modules across three academic years was evaluated. Detailed analysis reports the level of student participation in the task, time-resolved engagement during the assignment window, classification of the student question type (using a revised Bloom's taxonomy), quality of model answers/student comments and occurrence of errors. The level of participation was extremely high in both formative and summative cohorts. There was enhanced engagement from the formative cohorts in answering questions and engaging in PeerWise beyond the assignment deadline. The quality of student work was comparable irrespective of the summative/formative assignment and the majority of questions created tested beyond simple recall. Strategic engagement across both formative and summative cohorts is evident with a greater number of short cuts detected in the formative cohorts. This research study provides insights into some of the common concerns expressed by academics when planning the assessment portfolio within a programme of study.

.

**Keywords:** Formative assessment; PeerWise; Summative assessment.

## 1. Introduction.

Assessment in Higher Education places a focus on challenging students throughout their studies, enabling them to acquire the knowledge, skills and attributes that are both subject specific and generic to equip them for their future employment. Assessment is central to driving motivation for learning and supporting learning. The approach to learning can be described in terms of the distinction between deep learning and surface learning approaches (Entwistle &

Entwistle, 1991; Marton & Säljö, 1976). Deep learning involves an active interaction with a purposeful intention to develop understanding and link ideas together. Surface learning is more passive, focusing on reproducing content as required. A third approach, strategic learning, combines aspects of both surface and deep learning. It is practically impossible for a student to learn every topic in depth and so a choice is made in relation to priorities (Entwistle & Ramsden, 2015). The approach to learning is an important consideration when engineering assessments so that it is difficult for a learner to adopt purely a surface approach and perform satisfactorily. Qualitative evaluations on student approaches to learning (Miller & Parlett, 1974; Snyder, 1971) found that despite carefully considered curricula, students engaged with the assessment components as a priority.

To move beyond simple knowledge or factual acquisition towards higher order intellectual skill development, the assessment must change to encompass a more contextualised format that links to real world scenarios (Darling-Hammond & Snyder, 2000). The work of Carless (2007a) on learning-oriented assessment emphasises the features that enhance student learning whilst also appreciating the inherent tensions or 'double duty' (Boud, 2000) associated with it. Assessment is about grading and about learning, it communicates expected standards and also compares individuals.

## 1.1 Formative and summative assessment.

The function of assessment reflected in the paradigms of a summative approach (assessment *of* learning) and a formative approach (assessment *for* learning) is to promote and enhance learning (Brown, 2005; Carless, 2007b; Harris, Brown, & Harnett, 2015; Nicol & Macfarlane-Dick, 2006). Summative assessment is defined as the process in making a judgement according to a set of standards or criteria (Newton, 2007). These help to determine, at a point in time, knowledge acquisition against specific standards. 'Assessment of learning' describes the use of assessment to evaluate which students have achieved intended learning outcomes. Formative assessment with construct-referenced interpretations aims to improve performance and ensure consistency of meaning across contexts and individuals (Black & Wiliam, 2018; Wiliam & Black, 1996). Formative assessment is part of the instructional process and informs both teachers and learners about knowledge acquisition at a specific point. Learners engage in tasks and data is produced which is used to make inferences about student learning (Black & Wiliam, 2018). The distinction between formative and summative assessment is the distinction in inferences drawn from assessment outcomes. Summative inferences relate to the status of the student or their future potential. Formative inferences relate to helping the student learn and develop. The same assessment instrument can be used in both a summative and formative sense (Trotter, 2006).

For the purpose and discussions in this article, summative refers to an assessment that contributes directly to a grade on the student's performance record whereas formative assessment does not contribute directly to a grade.

Sadler (1989) argues that students need to monitor the quality of work produced and draw on a range of metacognitive strategies to close any gap between their actual performance and the required standard. Feedback, which can be take a variety of formats is an essential component of formative assessment. Nicol and Macfarlane-Dick (2006) offer principles of best practice on effective written or oral feedback that emphasises learners' agentic engagement. Winstone, Nash, Parker, and Rowntree (2017) accentuate learners' proactive recipience of feedback within feedback dialogues. The challenge associated with assessment practice is to appreciate, predict and shape the effects of assessment on learning.

## 1.2 Student engagement and buy-in.

Student engagement is more than participation and requires activity plus the association of cognitive and affective domains. There are a number of different interpretations in relation to student engagement, for example, Trowler and Trowler (2010) categorise three dimensions of engagement: student engagement in individual student learning; student engagement with structure and process; student engagement with identity. This can be articulated through a definition of student engagement (Trowler, 2010)

> '*Student engagement is concerned with the interaction between the time, effort and other relevant resources invested by both students and their institutions intended to optimise the student experience and enhance the learning outcomes and development of students and the performance, and reputation of the institution*.'

Factors that affect approaches to learning and buy-in include motivation (Pintrich, 2004),

self-efficacy (Bandura, 1989), self-regulation (Zimmerman & Schunk, 2001; 2004) including metacognition (e.g. planning and monitoring study habits), perceptions about learning environments and resource management. Self-efficacy refers to learners' perceived capabilities to perform actions while self-regulation refers to the individually generated thoughts and feelings that affect one's learning.

Lizzio and Wilson (2013) examined 1st year students' perceptions of assessment tasks in terms of motivation and manageability. The motivational value predicted the extent of student engagement and the task manageability predicted the task efficacy. Negative attitudes or a lack of value in the assessment will favour surface approaches to learning rather than deeper approaches of conceptual understanding. The role of the instructor has been shown to help build trust in an active learning environment (Cavanagh et al., 2018). Student engagement in formative assessment has been described aligned to five key objectives (Black & Wiliam, 2009):

1. Clarify learning expectations and criteria for success,
2. Engineer discussions and activities to show evidence of student understanding,
3. Provide feedback that moves learners forward,
4. Activate students as instructional resources for one another,
5. Activate students as owners with responsibility for their own learning.

## 1.3 PeerWise.

Student-centred learning reflects the shift from a transmission i.e. didactic mode of learning, to active learning. These *constructivist* and *socio-constructivist* perspectives (Bodner, Klobuchar, & Geelan, 2001; Bodner, 1986) include students and teachers as responsible partners in learning and assessment. The use of peer collaboration is supported by Vygotsky's zone of proximal development theory (Vygotsky, 1978) where the social interactions extend the learner's current competence through guided scaffolding.

A freely available online platform, (PeerWise, 2021) enables students to create and produce their own multiple-choice questions (MCQs) with appropriate distractors. Students are central in the learning process generating a question bank for all students. Other students can answer the questions for self-assessment, have access to the explanations for their own comprehension, engage with the quality of the questions by rating the quality on a scale from 0-5, the difficulty (easy/medium/difficult) and post a comment. This functionality creates an opportunity for anonymous peer review and peer feedback. When a student answers a question, they are provided with immediate feedback including the correct answer, the distribution of answers by all students that attempted the question, the explanation from the author and any comments written by other students. The bank of questions developed by students is then available to be used by the class as a practice resource.

Gamification elements have been designed within the platform. The points system consists of a "reputation" score and an "answer" score. The "reputation" score is awarded for authoring and answering questions and the algorithm considers the ratings of those questions by other students. The “answer” score follows a simpler algorithm and updates on a student’s first attempt at a question. The badge system consists of three categories: basic, standard and elite corresponding to the difficulty required to earn them for example, many standard and elite badges encourage goal-setting behaviours correctly answering 10 or more questions on each of 5 consecutive days. Denny, McDonald, Empson, Kelly, and Petersen (2018) found a causal relationship between gamification and learning outcomes. There were significantly higher levels of voluntary self-testing activity which translated to a significant improvement in exam scores.

This study focuses on student engagement in individual learning, specifically the level of participation and the quality of the student work. Within PeerWise the number of questions authored, questions answered and comments written is consistently reported above the minimum requirement. For example, on a first-year computer programming course, with a requirement to create a minimum of two questions and answer a minimum of ten questions Denny, Hamer, Luxton-Reilly, and Purchase (2008) found on average students submitted 2.6 questions and answered 34 questions. In a first-year chemistry module, Galloway and Burns (2015), found that the number of questions submitted was approximately three times the expected levels whereas the number of answers submitted was approximately 18 times the expected levels.

PeerWise is typically reported as a low-stake coursework assessment where credit (between 1% and 6%) is given towards the student’s total mark (Denny, 2015; Galloway & Burns, 2015; Kay, Hardy, & Galloway, 2018; Kay, Hardy, & Galloway, 2019; Tatachar & Kominski, 2017) or in some cases as a formative assessment (Howe, McKague, Lodge, Blunden, & Saw, 2018; Luxton-Reilly & Denny, 2010; Walsh, Harris, Denny, & Smith, 2018). Where marks are awarded, this is based on the student’s participation against a minimum criterion and the internal PeerWise scores attributed to each participant.

In relation to the question quality produced, evaluations by Galloway and Burns (2015) in a 1st year chemistry course and Bates, Galloway, Riise, and Homer (2014) in a 1st year physics course were determined using the cognitive domain levels of the revised Bloom’s taxonomy (Anderson et al., 2001). Both studies found that many questions classified were of high quality, requiring more than just factual recall and used plausible distractors with valid explanations. These studies involved a PeerWise assignment with an associated credit (5% and 3% respectively). A study by Bottomley and Denny (2011) with 2nd year biomedical science students utilising PeerWise with an associated 10% credit allocation found that the majority of questions produced by students were classified on the lower order cognitive levels.

There are no studies that examine the quality of student contributions in a PeerWise task when used as a formative assessment or compare the quality of student work if a higher credit is adopted. There is evidence that grading systems used within assessment influence students' approaches to learning (Dahlgren, Fejes, Abrandt-Dahlgren, & Trowald, 2009). We have not found any studies that detail the extent of student engagement and any differences in specific aspects of student contributions when an assessment is presented in a formative or a summative assessment format within a programme of study. The variety and number of requirements within PeerWise presents a useful platform to analyse in detail any differences in student contributions between a formative and summative task.

In this study, we aim to understand some consequences both intended and unintended of the assessment format on student engagement between a formative and summative PeerWise assignment. The research question is:

To what extent does a formative versus a summative (contributing 20% to module mark) PeerWise chemistry assignment impact on the level and quality of student engagement in a 1st year programme within a UK institution?

# 2. Materials and Methods.

'In this study, PeerWise was implemented in two 1st year chemistry modules at a UK based institution. One module, Foundations of Pharmaceutical Chemistry (F) is a 30-credit module co-taught to Pharmacy and Pharmaceutical Science students. PeerWise is implemented as a formative assignment during the first term. The second module is Molecular Structure and Reactivity (S) which is a 15-credit module delivered to Biochemistry students. PeerWise is implemented as a summative assignment during the first term with an allocation of 20% weighting towards the total module mark. Both modules will be referred to hereafter as F (for formative module) and S (for summative module).

There is a cohort size difference between both modules; the cohort sizes were: F-2015 (n=143); F-2016 (n=130); F-2017 (n=139); S-2015 (n=24); S-2016 (n=13); S-2017 (n=16). All students registered for both modules have prior chemistry learning although the pre-requisite for the formative module is a grade level higher (B grade) than the summative module (C grade). A cross-sectional study across three academic years (2015/16, 2016/17 and 2017/18) has been undertaken so that research findings can be compared across multiple cohorts. The data was not aggregated to understand if any patterns were consistently exhibited across the three academic years. The cohort size is the major point of difference and has been considered.

The scaffolding approach used by Casey et al. (2014) with an introductory workshop prior to the PeerWise assignment was adopted for both modules. This introduced students to the structure of a MCQ, the importance of plausible distractors and the challenge to create a good quality MCQ that goes beyond simple recall.

The minimum expected engagement for both PeerWise assignments was identical: create 2 questions; answer 5 questions from their peers; rate and comment on 3 questions. Within the formative module, students were allocated two topics from their module content whereas in the summative module, students had a free choice on topics. The rationale for this difference was to guard against a student in the summative module underperforming due to the assigned topic. All students had 7 weeks during the first term to complete the PeerWise assignment. The PeerWise platform was available for students to use after the assignment deadline for exam preparation. The extent of academic staff involvement during the PeerWise assignment is minimal thus promoting a student-led collaboration. There is staff presence to deal with any alerts flagged by students within the system and moderate but similar to Galloway and Burns (2015) the students take ownership and responsibility within the online environment.

The systematic item analysis approach used by Bates et al. (2014) and Galloway and Burns (2015) to classify the student-authored questions informs our approach. The revised Bloom's taxonomy (Anderson et al., 2001) was used as a framework (Table 1) to categorise the cognitive level attributed to each question created.

Table 1: Categorisation levels and explanations for the cognitive domain of Bloom's taxonomy.

| Level | Identifier | Description |
|---|---|---|
| 1 | Remember | Factual recall, definition. |
| 2 | Understand | Basic understanding, no calculation necessary. |
| 3 | Apply | Implement, calculate or determine. Single topic calculation or exercise involving application of knowledge |
| 4 | Analyse | Multi-step problem, requires identification of problem-solving strategy before executing. |
| 5 | Evaluate | Compare and assess various option possibilities, often qualitative and conceptual questions. |
| 6 | Create | Synthesis of ideas and topics from multiple course topics to create significantly challenging problem. |

The solution explanation provided by the author was visible during the classification procedure along with any comments from other students thus providing a more contextual overview to the ratings awarded. The quality of the explanation in the solution written by the author was categorised as per Table 2.

Table 2. Categorisation levels for explanation of solution to questions.

| Level | Identifier | Description |
|---|---|---|
| 0 | Missing | No explanation provided or explanation incoherent. |
| 1 | Inadequate | Wrong reasoning and/or answer. Solution may be trivial, flippant or unhelpful. |
| 2 | Minimal | Correct answer but with insufficient explanation or justification. Some aspects may be unclear or incorrect |
| 3 | Good | Clear and sufficiently detailed exposition of both correct method and answer. |
| 4 | Excellent | Thorough description of relevant chemistry and solution strategy. Contains remarks on plausibility of answer and/or other distractors. Beyond normal expectations for a correct solution. |

The comments left by peers were evaluated and categorised according to Table 3 which contains an additional level of distinction to that used by Galloway and Burns (2015).

Table 3. Categorisation levels for quality of comments left by peers.

| Level | Description |
|---|---|
| 0 | No comment provided. |
| 1 | Comments left contain simple phrases such as ‘Good Question’. |
| 2a | Comments left contain phrases with a simple qualifier/justification such as ‘good question because it tested my knowledge’. |
| 2b | Comments left contain phrases with a detailed qualifier/justification such as ‘good question because it required me to think carefully as the other answer options were good distractors and I had to double check my answer’. |
| 3 | Comments left suggest improvements and/or new ideas leading to discussion and/or improvements to the question as a whole. |

Table 4 shows two examples of comments that contain a phrase with a qualification. The level of detail in the justifications provided were not considered equal and thus a differentiation in level 2 was included in the classification of comments (Table 3).

Table 4. Examples of comments with qualifications.

| | |
|---|---|
| Comment | This is a good question as it tests the students' knowledge on bonding, although it is at a basic level |
| Comment | This is a simple question, as it is just a definition, the person answering must only recall the definition to get the marks. The stem of the question is perfect and unambiguous, the distractors are all good as they are all definitions, this makes it harder to eliminate answers. This would be a hard question for someone who has not revised the definition because the distractors are so similar, therefore I think it is a good question. I do not think the question could be improved as it is already a straightforward question, and does not require much thought, just memory. |

The data analysis process was carried out by two researchers (E.H. and N.P.) who were assigned formative and summative cohorts. One researcher (N.P.) carried out the classification of student content for F-2015 (292 questions), S-2015 (50 questions) and the other researcher (E.H.) carried out the classification of student content for F-2016 (258 questions), F-2017 (290 questions), S-2016 (26 questions), S-2017 (33 questions). Prior to the classification process, both researchers discussed the categorization levels with an experienced educationalist (S.F.) to ensure there was a shared understanding of the intended outcomes. Using a sample of 10 questions selected at random (using the question ID generated in the PeerWise system) from each of the cohorts the rating results were compared pairwise between researchers (N.P. and S.F; E.H. and S.F.) to identify similarities and differences and any clarifications on the levels were agreed. This iterative cycle was repeated until consistent results were obtained and then the researchers (E.H. and N.P.) completed the classification process.

The research project was approved by the ethics board at the University of Hertfordshire. As per the Data Protection Act 2018, any identifiable and traceable information has been removed.

# 3. Results

The formative and summative modules are used to compare the levels and quality of student engagement within a 1st year PeerWise assignment.

## 3.1 Level of Student Participation.

Table 5 shows the cohort sizes, the expected minimum engagement and the actual engagement against the assignment criteria.

Table 5. The expected minimum engagement and actual engagement for all cohorts.

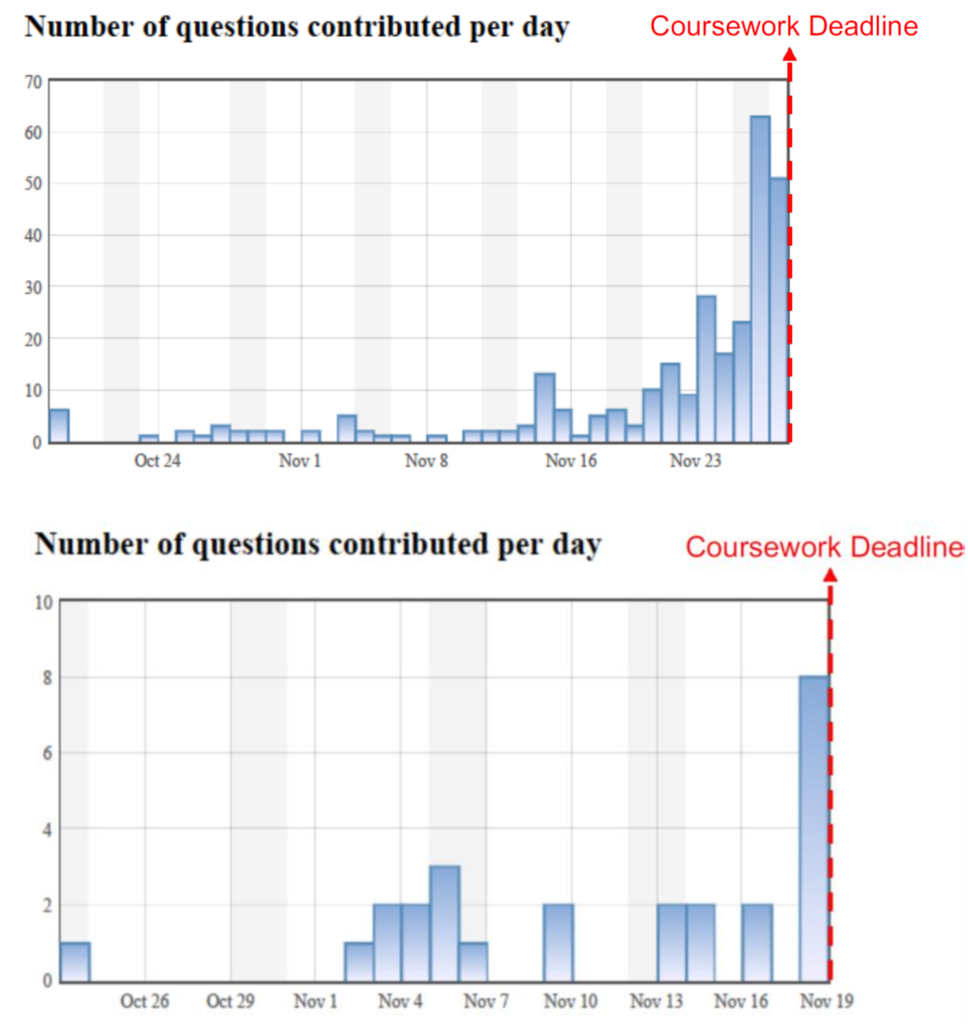
| Module | Year | Cohort size | Questions Authored | | Questions Answered | | Comments written | |
|---|---|---|---|---|---|---|---|---|
| | | | Expected | Submitted | Expected | Submitted | Expected | Submitted |
| F | 2015 | 143 | 286 | 292 | 715 | 2022 | 429 | 605 |
| | 2016 | 130 | 260 | 260 | 650 | 1556 | 390 | 469 |
| | 2017 | 139 | 278 | 290 | 695 | 2785 | 417 | 537 |
| S | 2015 | 24 | 48 | 50 | 120 | 157 | 72 | 91 |
| | 2016 | 13 | 26 | 26 | 65 | 72 | 39 | 44 |
| | 2017 | 16 | 32 | 33 | 80 | 83 | 48 | 53 |

There is a much higher level of engagement in the formative modules in relation to the number of answers submitted with up to 4 times the level of questions answered for the F-2017 cohort. A closer inspection of student participation shows that for the summative cohorts, only 2 students did not meet the threshold criteria on answering questions (one student answered 4 not 5 questions) and commenting on questions (one student commented on 2 not 3 questions). Across the formative cohorts, the total number of students that did not meet the criteria is: for creating questions (9 students, 2.2%), answering questions (15 students, 3.6%) and commenting on questions (57 students, 13.8%).

Within the summative modules, the maximum engagement levels observed by different individual students were 3 authored questions, 15 answers and 7 comments and within the formative modules, 6 authored questions, 288 answers and 35 comments.

The student activity timelines illustrate the level of engagement over the assignment period. The number of questions contributed per day for F-2015, F-2016 and F-2017 show a similar pattern where the highest levels of activity are observed days prior to the assignment deadline. For example, with F-2016, 116 questions (45%) were contributed in the last 5 days before the deadline. For F-2017, 182 questions (63%) were contributed 5 days before the deadline out of which 114 (39%) were contributed in the final 2 days. A comparison with the summative modules S-2015, S-2016 and S-2017, shows that questions were contributed throughout the assignment period with a similar increased level of activity prior to the assignment deadline. For example, S-2016 the highest number of questions contributed in one day, was the day of the assignment deadline. For S-2017, 21 out of 32 questions were contributed 4 days before the assignment deadline. Figure 1 shows a comparison of the timeline activity in relation to questions contributed per day for one of the formative modules F-2017 and one of the summative modules S-2016.

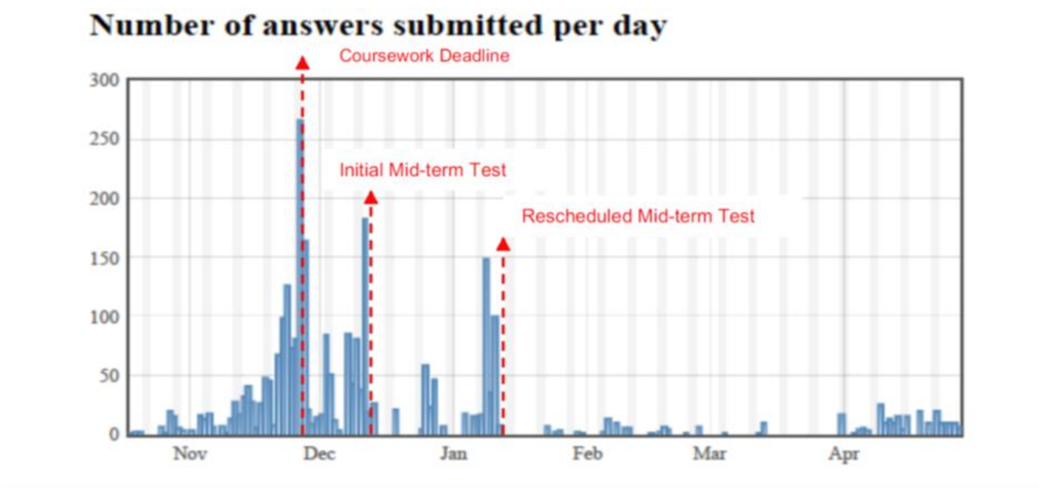
Figure 1: Plots for student engagement over time in relation to questions contributed per day for the formative module F-2017 (top) and the summative module S-2016 (bottom).

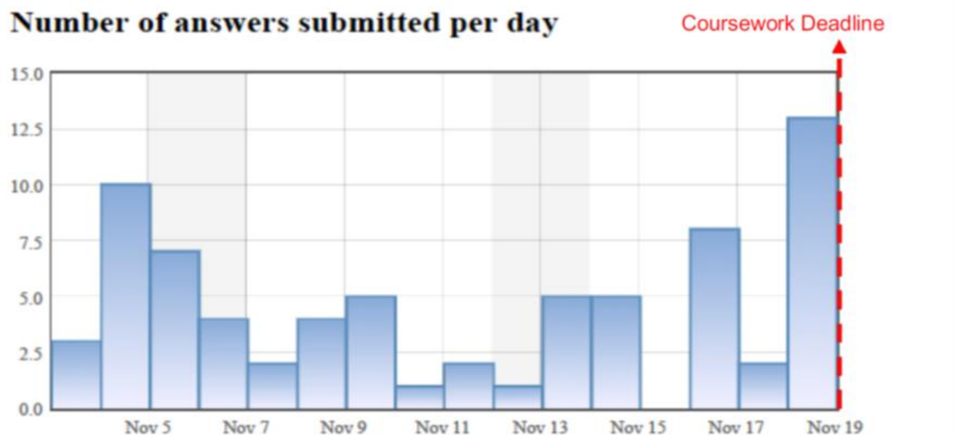



The timeline activity for the formative modules showed a level of activity after the PeerWise assignment deadline and ahead of a summative mid-term test. Figure 2 shows the timeline plot for F-2017, where a mid-term test scheduled on 11th December 2017 was rescheduled to 11th January 2018 due to adverse weather conditions. The end of module examination was

3rd May 2018, where small but frequent levels of activity were observed. In contrast, the timeline activity patterns for the summative modules showed a more consistent pattern throughout the assignment period, however, no answers were submitted after the assignment deadline or prior to an end of Semester examination.

Figure 2: Plots for student engagement over time in relation to answers submitted per day for F-2017 (top) and S-2016 (bottom).



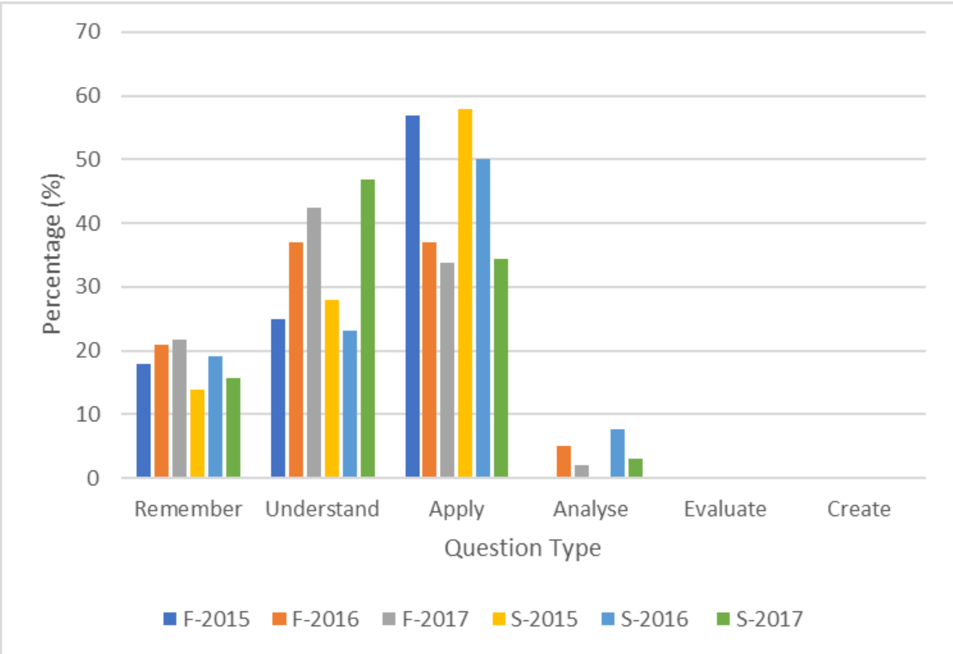




All cohorts were made aware of the purpose associated with implementing the PeerWise assignment and the workshop format emphasised this aspect. The workshop discussed with examples the features of a good quality MCQ, a constructive comment and introduced Bloom's taxonomy. Students had an opportunity to write their own MCQs in small groups and critically evaluate these questions. After the assignment deadline, there was no direct communication to encourage PeerWise activity.

## 3.2 Quality of student engagement.

Most questions across all cohorts go beyond the simple factual ‘remember’ types (~20% of questions at ‘remember’ level). Most are ‘understand’ and ‘apply’ type questions with a low number of ‘analyse’ questions and no ‘evaluate’ and ‘create’ type questions (see Figure 3).

Figure 3: Distribution of question type classification using the cognitive levels of the revised Bloom’s taxonomy for all cohorts.

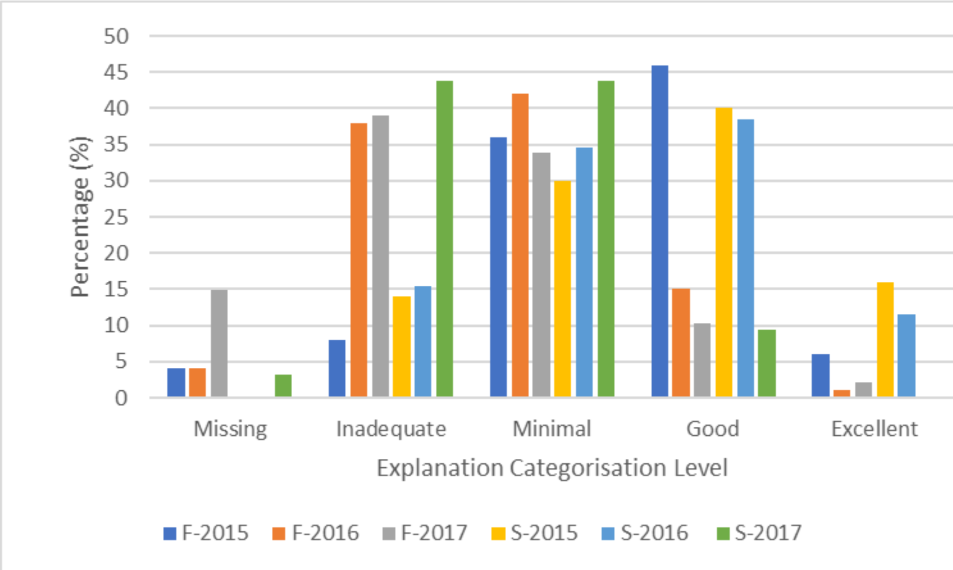



The percentage of ‘understand’ type questions is observed as: F-2015 (25%); F-2016 (37%); F-2017 (42%); S-2015 (28%); S-2016 (23%) and S-2017 (47%). There is no distinct difference between the formative and summative cohorts on question types; a higher percentage of ‘apply’ questions was found across S-2015 (58%) and S-2016 (50%) but a similarly high level was observed in F-2015 (57%).

Upon a closer inspection of the topics covered in questions within the formative cohorts, calculation questions accounted for the majority of the ‘apply’ type questions, for example in F-2015, 73% of the ‘apply’ questions were calculation based compared to 12% in the S-2015 cohort. The formative module covers more calculation-based topics, such as percentage yield, acid-base and buffer calculations within the PeerWise assignment period with more conceptual topics considered in the second semester. The summative module has a focus on more conceptual topics throughout the PeerWise assignment period.

The distribution plot of explanation classification for solutions to students’ questions is shown in Figure 4.

Figure 4: Distribution of explanation classification for solutions to questions completed by students for all cohorts.




From the results, minimal explanations with some unclear aspects and details were most regularly observed throughout all cohorts. The percentage of missing explanations is higher for the formative cohorts with one question in S-2017 having a missing explanation. This study also found a higher percentage of inadequate explanations containing incorrect or unhelpful information (F-2015 (8%), F-2016 (38%), F-2017 (39%), S-2015 (14%), S-2016 (15%), S-2017 (44%)) than that reported by (Galloway & Burns, 2015), who reported a small percentage (~5%). A positive learning outcome is that students clarified aspects of 'inadequate' explanations in their comments and engaged in constructive dialogue (Table 6).

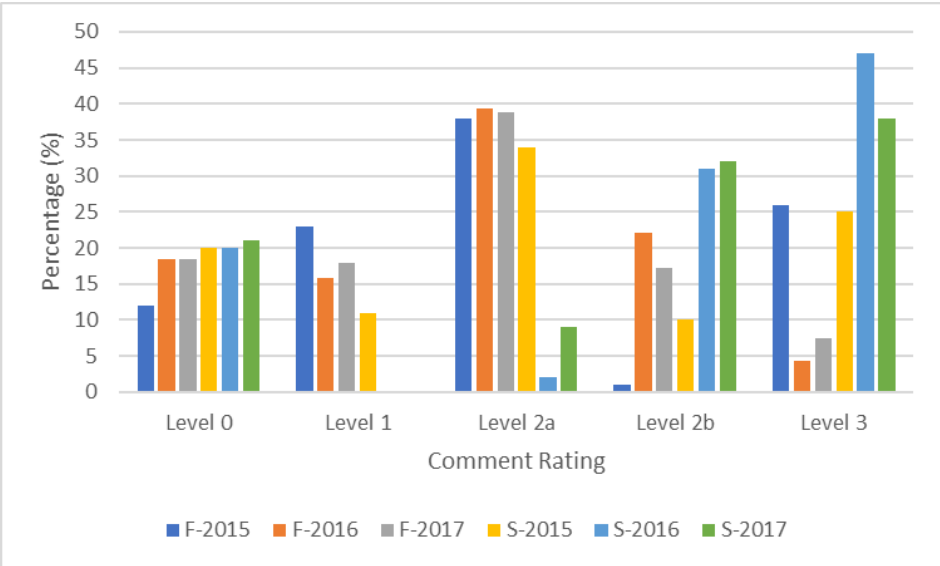
Table 6. Errors highlighted by other students in their comments.

| | |
|---|---|
| F-2017 | The answer is wrong because $CO_2$ is sp hybridised for the central carbon atom and $C_2H_2$ is $sp^2$ hybridised due to the double bond containing one pi bond and 3 sigma bonds.<br><br>Reply:<br><br>Yes $CO_2$ is indeed sp hybridised. However, $C_2H_2$ is also sp hybridised due to the fact that it contains 2 sp hybridised orbitals and two p orbitals. The sp hybridised orbitals is bonded to two groups thus it is sp hybridised. |

| | |
|---|---|
| S-2015 | The stem itself is very easy to understand and the majority of the distractors seem plausible as an answer however there are some problems in the explanation as well as the answer (in my opinion). First of all there is no tertiary amine present in the molecule, there is only a secondary amine group present however there is an ammonium ion present in the molecule. There is also a methyl group present however I do agree with your statements about the molecule having an amide group, a thioester group, an alkene group, a primary and secondary alcohol group and a phendiol group. The main problem with this question is the explanation and the answer however the stem and the alternatives are completely fine. |

For the summative cohorts, marks were assigned for factually correct answers, thus emphasising an expectation on this aspect. Some students in the formative cohorts simply stated in their explanation 'Refer to X's lecture' thus demonstrating less effort and simply directing other students to the source of the information. The percentage of excellent explanations is higher for the summative cohorts compared to the formative cohorts although the summative S-2017 cohort showed no 'excellent' level explanations. Figure 5 shows the categorisation levels for the quality of comments.

Figure 5: Distribution of quality classification of comments left by peers.




There is a similar percentage (~20%) of questions across all cohorts with no comments. Typically, this was found to be with the later submitted questions in the summative cohorts, as questions which had been submitted earlier generally have more comments. In the formative cohorts, some questions have more than 10 comments. Where comments have been provided the distinction between the formative modules and summative modules is more pronounced. Most comments (S-2015 (25%), S-2016 (47%), S-2017 (38%)) for the summative cohorts are at the highest ‘Level 3’ category and suggest improvements and/or new ideas to support the development of the question. The majority of comments for the formative cohorts (F-2015 (38%), F-2016 (39%), F-2017 (39%)) are at the ‘Level 2a’ category and contain a simple phrase such as ‘good question’. The category ‘Level 2b’ containing a simple phrase and detailed qualifier illustrate that a higher percentage of comments in the summative cohorts contain this extended detail compared to comments in the formative cohorts. The marking criteria for the summative assignment emphasise the requirement for reflection and constructive feedback in comments. The students have produced work to meet this expectation. Only S-2015 is an exception where 34% of comments were categorised at ‘Level 2a’. Across both cohorts, most students wrote accurate, coherent questions. The accuracy in the correct answer options was shown to be higher for the summative cohorts, with only one S-2015 cohort having 2 questions with errors (4%). In contrast there were some errors in all the formative cohorts: F-2015, 9 questions with errors (3%); F-2016, 7 questions with errors (3%); F-2017, 19 questions with errors (7%). There were errors across different topics and included both conceptual and procedural topics. What is pleasing to note is that many errors were detected by other students and their constructive comments highlighted this (Table 6).

There was a higher occurrence of questions with only 2 or 3 answer options in the formative cohorts: F-2015, 17 questions (6%); F-2016, 16 questions (6%); F-2017, 21 questions (7%); S-2015, 1 question (2%); S-2016, all questions had 5 answer options; S-2017, 2 questions (6%). Looking at the most answered questions across both the formative and summative cohorts, there was no distinct pattern. The highest rated questions did not have more answers, for example, in the formative F-2016 cohort, two questions with a maximum average rating of 5 had 4 answers, the highest answered question with 25 answers had an average rating of 3. The questions with most answers were all created in the first six weeks of the assignment window.

Both formative and summative cohorts contained questions of the highest rating (see Table 7). Regarding the quality of questions (with the highest average ratings from peers) created the day before or on the day of the assignment deadline: in the formative cohorts, half (8 questions) were created and in the summative cohorts, 13% (2 questions) were created.

Table 7: The highest and lowest average ratings awarded by peers.

| | Average Rating | | | |
|---|---|---|---|---|
| | Highest | Lowest | % Less than 3.5 | % Between 3.5-5.0 |
| F-2015 | 5 | 0.25 | 73.2 | 26.8 |
| F-2016 | 5 | 0.67 | 84.5 | 15.5 |
| F-2017 | 4.67 | 1 | 89.5 | 10.5 |
| S 2015 | 4.33 | 2 | 63.0 | 37.0 |
| S 2016 | 5 | 1.5 | 43.8 | 56.3 |
| S 2017 | 5 | 2 | 52.6 | 47.4 |

# 4. Discussion.

The level and quality of student engagement in the context of this study provides some interesting insights and distinctions into how students manage and prioritise assignment tasks.

The overall participation for the formative cohorts is extremely high [F-2015 (94%), F-2016 (91%) and F-2017 (84%)]. For the formative cohorts, the students answered more questions overall but commented least. This demonstrates a level of strategic participation where students in the formative cohorts are choosing to spend their time on select aspects of the assignment. Previous research by Fergus (2019) indicated that 1st year students perceived the creating of questions in PeerWise as the most challenging. The most beneficial aspect was answering questions. In this study, both summative and formative cohorts have engaged well with creating questions despite the inherent challenge. A key difference between the summative and formative cohorts is that the summative cohorts did not engage with answering questions beyond the assignment criteria. The summative cohorts viewed the assignment deadline as the end of the PeerWise activity and did not revisit the question repository. It could be argued that the smaller summative cohorts reduce the size of the question banks available and hence restricts the level of engagement although there is no wider evidence to support this. Previous research (Fergus, 2019) has shown that some 1st year students recognised the benefit of the PeerWise repository but time barriers did not permit further engagement whereas other students did not trust the quality of other students' work and chose not to engage further. This would be described as a dualistic thinking where there are right answers and the lecturer has the right answers in Perry's scheme of epistemological development (Perry, 1999). The factors of time and trust in question quality could also explain the differences observed in this study.

The decisions learners make to self-regulate tends to be driven by deadlines (Yan, Thai, & Bjork, 2014) and the influence of the assignment deadline is a strong feature irrespective of the formative or summative nature of the assignment. From previous reports using PeerWise, activity is evident around module examination dates (Duret, Christley, Denny & Senior, 2018) or when used as a coursework task (Casey et al. , 2014). One example where PeerWise contributed 20% to the final module grade (Grainger, Dai, Obsorne & Kenwright, 2018) does not report any timeline data so we cannot make comparisons to our findings.

An important aspect is the role of the academic in establishing expectations. The students were made aware of PeerWise in the introductory workshop and they had the opportunity to create a question with feedback in small groups. One of the most effective ways to help create a strong sense of efficacy is through mastery experiences. The workshop experience provides an opportunity to successfully perform the task required in their first assignment experience at university and addresses the first three of the five objectives for student engagement in formative assessment (Black & Wiliam, 2009). Students also view questions created by their peers within the PeerWise platform and this would help strengthen self-beliefs of efficacy. The formative cohorts received an email from the tutor with their assigned two topics. This level of personal interaction between student and academics helps to create a connection and it could be argued an expectation of engagement.

The gamification aspects within PeerWise is a valid consideration in relation to engagement. The smaller summative cohort sizes and hence smaller question bank repository would make it more difficult to attain particular badges for example, correctly answering at least 10 questions correctly in a row and impossible for answering at least 50 questions. Although game elements in PeerWise have been shown to motivate students to create questions, the evidence in a large cohort (~500 students) does not suggest that students persist with answering questions (Denny et al., 2018).

A potential limitation with using Bloom's Taxonomy to classify questions is that calculation questions are categorised at the 'apply' level; the cognitive load (Paas, Renkl, & Sweller, 2003) associated with creating a calculation-based question that is more formulaic could be reduced particularly if a student models the question stem from a sample question and simply changes the values. Galloway and Burns (2015) similarly found fewer 'Analyse' type questions but a larger proportion of 'Evaluate' type questions. The first-year chemistry modules in this study are components of applied degree programmes. Examination questions within both modules address cognitive development at the knowledge, comprehension and application levels and examples in the PeerWise workshop are purposefully designed at these levels. It is expected that students create questions at these levels too. Due to the variation in different research contexts, drawing conclusions and making direct comparisons to other research findings is advised with caution.

In relation to student buy-in, there is evidence of short cuts that some students have taken in the formative cohorts when compared to the summative cohorts. There were a higher number of missing explanations. Perhaps the students did not provide an associated explanation as there no was perceived benefit of the task or marks gain, by using the extra 'effort' to add an explanation. Some students simply stated 'Refer to X's lecture' rather than generate an explanation. The principle of *generation,* where learners generate components of a learned stimulus promotes better learning outcomes but requires more effort from the learner hence the term "desirable difficulty" (Bjork & Bjork, 2011; Richland, Bjork, Finley, & Linn, 2005). Another example is the higher occurrence of questions with only 2 or 3 answer options and not 5 answer options. The extent of errors in questions is higher for the formative cohorts. Similarly, Denny, Luxton-Reilly, Hamer, and Purchase (2009) reported that the 11% of questions with incorrect answers were picked up on and corrected by students. This is the strength of this student-centred online community; it enables constructive dialogue between students to enhance and transform their learning.

Students are choosing to engage with the assignment close to the deadline, but this does not mean that their work is of lower quality. The percentage of questions with an average rating between 3.5 and 5 is higher in the summative cohorts again illustrating the range of quality work produced (Table 7).

The purpose of a strategic approach is to seek the optimal outcome where the ratio of achievement to effort is high. The level of engagement and the student's approach to learning whether strategic, deep and surface are linked (Yorke, 2006). The strategic engagement observed in this study involves the use of both deep and surface learning decided by the student where appropriate. The mark scheme for the PeerWise summative assignment (20%) includes scores for the correct explanation and detailed constructive comments. Not surprisingly, the summative cohorts have completed these aspects fully.

## 4.1 Conclusion.

The level of student participation was high in both the summative (20%) and formative PeerWise assignment with an extended level of engagement from the formative cohorts beyond the assessment deadline. The quality of engagement across a formative and summative online assignment is consistent in relation to the questions created however, the formative cohorts adopt short cuts in some respects such as explanations, number of answer options and providing constructive comments. Summative assignment criteria will directly influence strategic engagement as evidenced by the quality of student comments and explanations. Factors such as self-efficacy and self-regulation influence the student role in assessment and their buy-in. Modelling and using appropriate examples initially prior to the assignment can support student buy-in. Communication with students can support connection and establishes expectations for engagement.

**Acknowledgements.**

The authors wish to thank Paul Denny, University of Auckland (creator and developer of the PeerWise system) for his continuous support and help, particularly in relation to user data analytics.

# 5. References.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Wittrock, M. C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom’s taxonomy of educational objectives, abridged edition. White Plains, NY: Longman.

Bandura, A. (1989). Human agency in social cognitive theory. American Psychologist, 44(9), 1175.

Bates, S. P., Galloway, R. K., Riise, J., & Homer, D. (2014). Assessing the quality of a student-generated question repository. *Physical Review Special Topics - Physics Education Research,* 10(2), 020105.1-020105.11.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society, 2*(59-68).

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*(1), 5.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551-575.

Bodner, G., Klobuchar, M., & Geelan, D. (2001). The Many Forms of Constructivism. *Journal of Chemical Education, 78*(8), 1107. doi:10.1021/ed078p1107.4

Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education, 63*(10), 873.

Bottomley, S., & Denny, P. (2011). A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochemistry and Molecular Biology Education, 39*(5), 352-361. doi:10.1002/bmb.20526

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151-167.

Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education* (1), 81-89.

Carless, D. (2007a). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57-66.

Carless, D. (2007b). Conceptualizing pre-emptive formative assessment. *Assessment in Education, 14*(2), 171-184.

Casey, M. M., Bates, S. P., Galloway, K. W., Galloway, R. K., Hardy, J. A., Kay, A. E., McQueen, H. A. (2014). Scaffolding student engagement via online peer learning. *European Journal of Physics, 35*(4). doi:10.1088/0143-0807/35/4/045002

Cavanagh, A. J., Chen, X., Bathgate, M., Frederick, J., Hanauer, D. I., & Graham, M. J. (2018). Trust, growth mindset, and student commitment to active learning in a college science course. *CBE—Life Sciences Education, 17*(1), ar10.

Dahlgren, L. O., Fejes, A., Abrandt-Dahlgren, M., & Trowald, N. (2009). Grading systems, features of assessment and students' approaches to learning. *Teaching in Higher Education, 14*(2), 185-194.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16*(5-6), 523-545.

Denny, P. (2015). Generating practice questions as a preparation strategy for introductory programming exams. Paper presented at the *SIGCSE 2015 - Proceedings of the 46th ACM Technical Symposium on Computer Science Education, p. 278-283.*

Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. (2008). PeerWise. Paper presented at the *Koli Calling 2008 - 8th International Conference on Computing Education Research .*

Denny, P., Luxton-Reilly, A., Hamer, J., & Purchase, H. (2009). Coverage of course topics in a student generated MCQ repository. Paper presented at the *Conference on Integrating Technology into Computer Science Education, ITiCSE, p.* 1-13.

Denny, P., McDonald, F., Empson, R., Kelly, P., & Petersen, A. (2018). Empirical support for a causal relationship between gamification and learning outcomes. Paper presented at the *Conference on Human Factors in Computing Systems - Proceedings,*

Duret, D., Christley, R., Denny, P & Senior, A. (2018). Collaborative learning with PeerWise. *Research in Learning Technology,* 26, 1-13.

Entwistle, N., & Ramsden, P. (2015). *Understanding Student Learning (Routledge Revivals)*. New York: Routledge.

Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: the student experience and its implications. *Higher Education, 22*(3), 205-227.

Fergus, S. (2019). Using PeerWise to support the transition to higher education. *Teaching Chemistry in Higher Education*, 1, 181-194.

Galloway, K. W., & Burns, S. (2015). Doing it for themselves: Students creating a high quality peer-learning environment. *Chemistry Education Research and Practice, 16*(1), 82-92.

Grainger, R., Dai, W., Osborne, E. & Kenwright, D. (2018). Medical students create multiple-choice questions for learning in pathology education: A pilot study. *BMC Medical Education*, 18(201). Available: Medical students create multiple-choice questions for learning in pathology education: a pilot study (biomedcentral.com)

Harris, L. R., Brown, G. T. L., & Harnett, J. A. (2015). Analysis of New Zealand primary and secondary student peer- and self-assessment comments: applying Hattie and Timperley's feedback model. *Assessment in Education: Principles, Policy and Practice, 22*(2), 265-281.

Howe, P. D. L., McKague, M., Lodge, J. M., Blunden, A. G., & Saw, G. (2018). PeerWise: Evaluating the Effectiveness of a Web-Based Learning Aid in a Second-Year Psychology Subject. *Psychology Learning and Teaching, 17*(2), 166-176.

Kay, A. E., Hardy, J., & Galloway, R. K. (2018). Learning from peer feedback on student-generated multiple choice questions: Views of introductory physics students. *Physical Review Physics Education Research, 14*(1).

Kay, A. E., Hardy, J., & Galloway, R. K. (2019). Student use of PeerWise: A multi-institutional, multidisciplinary evaluation. *British Journal of Educational Technology*, 51(1), 23-34.

Lizzio, A., & Wilson, K. (2013). First-year students' appraisal of assessment tasks: implications for efficacy, engagement and performance. *Assessment & Evaluation in Higher Education, 38*(4), 389-406.

Luxton-Reilly, A., & Denny, P. (2010). Constructive evaluation: a pedagogy of student-contributed assessment. *Computer Science Education, 20*(2), 145-167.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology, 46*(1), 4-11.

Miller, C. M., & Parlett, M. (1974). Up to the mark: A study of the examination game.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education, 14*(2), 149-170.

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.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

PeerWise. (2021). PeerWise website. Retrieved from https://PeerWise.cs.auckland.ac.nz.

Perry Jr, W. G. (1999). *Forms of Intellectual and Ethical Development in the College Years: A Scheme.* Jossey-Bass Higher and Adult Education Series: ERIC.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). *Linking cognitive science to education: Generation and interleaving effects.* Paper presented at the Twenty-seventh Annual Conference of the Cognitive Science Society.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Snyder, B. R. (1971). *The Hidden Curriculum*. Cambridge, MA: MIT Press.

Tatachar, A., & Kominski, C. (2017). Assessing a traditional case-based application exercise and a student question creation exercise on student performance and perceptions. *Curr Pharm Teach Learn, 9*(4), 689-697. doi:10.1016/j.cptl.2017.03.007

Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education, 31*(5), 505-521.

Trowler, P., & Trowler, V. (2010). Student Engagement Executive Summary: Higher Education Academy Student Engagement Project. HEA. Available Student_Engagement_Project_Executive_Summary._Nov_2010.pdf (lancs.ac.uk)

Trowler, V. (2010). *Student Engagement Literature Review*. The Higher Education Academy. Available: Student engagement literature review | Advance HE (advance-he.ac.uk)

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children, 23*(3), 34-41.

Walsh, J. L., Harris, B. H. L., Denny, P., & Smith, P. (2018). Formative student-authored question bank: Perceptions, question quality and association with summative performance. *Postgraduate Medical Journal, 94*(1108), 97-103.

Wiliam, D., & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537-548.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17-37.

Yan, V. X., Thai, K. P., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they varywith mindset. *Journal of Applied Research in Memory and Cognition, 3*(3), 140-152.

Yorke, M. (2006). *Student engagement: Deep, surface or strategic.* Paper presented at the Keynote address to the 9th Pacific Rim Conference on the First Year in Higher Education, Griffith University, Australia.

Zimmerman, B. J., & Schunk, D. H. (eds) (2001). *Self-regulated Learning and Academic Achievement: Theoretical Perspectives*. (2nd edition). Mahwah, N.J: Erlbaum.

Zimmerman, B. J., & Schunk, D. H. (2004). Self-regulating intellectual processes and outcomes: A social cognitive perspective. In Y. Dai & R.J. Sternberg (eds). *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development. Mahwah, N.J.: Erlbaum*, pp. 323-349.