

McCallum, S. and Milner, M. M. (2021) The effectiveness of formative assessment: student views and staff reflections. *Assessment and Evaluation in Higher Education*, 46(1), pp. 1-16.

(doi: 10.1080/02602938.2020.1754761)

This is the Author Accepted Manuscript.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

https://eprints.gla.ac.uk/213472/

Deposited on: 19 May 2020

The Effectiveness of Formative Assessment: Student views and Staff reflections

Suzanne McCallum^a* and Margaret Milner^a

^aAdam Smith Business School, University of Glasgow, Glasgow, UK

*Corresponding author. Email: suzanne.mccallum@glasgow.ac.uk LinkedIn:

https://www.linkedin.com/in/suzanne-mccallum ORCID: https://orcid.org/0000-0002-3587-

2496

Suzanne McCallum is a Senior Lecturer in Accounting & Finance within the Adam Smith Business School at the University of Glasgow. Her research interests include formative assessment, peer review and using technology to facilitate formative assessment.

Margaret Milner was the Director of Undergraduate studies for the Adam Smith Business School at the University of Glasgow, she is now retired. Her research interests include evaluating and embedding graduate skills within Higher Education and the study of Graphicacy within accountancy and accounting education.

Acknowledgements

The contributors would like to sincerely thank Professor Catriona Paisey, Professor Lee Parker and the anonymous reviewers for their very helpful comments on earlier drafts of this paper.

The Effectiveness of Formative Assessment: Student views and Staff

reflections

Abstract

This paper reports on the implementation of formative e-assessments in courses taken

by first year students. The central aim is to measure effectiveness of the formative e-

assessments with reference to the student voice and staff reflections. Students engaged

with the formative assessments and the evidence gathered via questionnaires shows that

students perceived that formative e-assessments helped them to monitor their progress;

encouraged further study and increased their learning and understanding. The

findings also allowed academics to reflect on the benefits of adopting formative e-

assessments to foster student engagement and permit early intervention. By focusing

on the student view (rather than student performance) the findings add to our

understanding of the benefits of integrating regular formative assessment in first year

courses.

Key words: Formative assessment, Feedback effectiveness, e-assessment

Introduction

Educational literature has established that feedback is an active process, rather than simply

the transmission of information. For the feedback process to be effective, students must

decode information received and take action (Nicol 2010; Carless and Boud 2018). While

information can be delivered in a variety of ways, formative assessments are often the vehicle

of choice: as they can focus on not only promoting learning, but also provide diagnostic

information to both students and staff (Harlen and James 1997).

2

The central objective of this paper is to measure the effectiveness of a series of formative e-assessments with reference to the student voice and staff reflections. Evidence suggests that use of formative assessment in the early weeks of the first year of a university degree is associated with student success (Yorke 2003): enhancing students' ability to monitor their own progress is a crucial element of their empowerment in the learning process. However, given the current pressure on formative assessment and the growing view that it is an 'optional extra' (Wu and Jessop 2018) it is important to investigate and reaffirm its value.

Bennett (2011) highlights issues that need to be examined, including the effectiveness of formative assessments. Many studies make it difficult to assess effectiveness of such assessments as they include both a formative and summative element (Dawson et al. 2019). It is therefore difficult to get a clear picture of the impact of purely formative assessments. This study addresses this issue by considering the effectiveness of a series of formative e-assessments introduced across two first year courses. These were carefully designed with reference to best practice (Gibbs and Simpson 2004 and Nicol 2009b) to maximise potential effectiveness. This study reveals that formative assessment effectiveness is enhanced by focusing on student perceptions of assessment impact on their own performance self-reflections and learning engagement. This view of effectiveness contrasts with some other studies (Perera, Nguyen, and Watty 2014) that focus solely on the improvement of student results. The paper also contributes to the literature by reporting on the staff view of the formative assessments, an important aspect that is often overlooked. The remainder of the paper firstly considers the literature on formative assessment and measures of its

effectiveness. It then goes on to describe the methods used in this study, before discussing the results obtained, conclusions drawn and suggestions for future research.

Formative assessment

Formative assessment focuses on promoting student learning (Wu and Jessop 2018), by giving information on performance that students can act on, and by ensuring staff adapt teaching to meet student needs (Black and Wiliam 1998, 2009). Therefore, formative assessment and feedback are inextricably linked.

Pryor and Crossouard (2008, 2010) see formative assessment as taking place when student work is responded to and evidence is given on what constitutes good learning.

Weurlander et al. (2012) showed that formative assessments are an important tool for motivating students to study and for building student awareness of their own learning. They also point to formative assessment being complex and associated with powerful learning.

This view is in line with how the nature of formative assessment can bring about learning with understanding, and how its use can support the learning process (Harlen and James 1997). In this study formative assessment refers to e-assessments intended to generate feedback information that can be used to improve and accelerate learning (Sadler 1998; Nicol and Macfarlane-Dick 2006).

Formative assessment quality can be affected by the nature of the assessment and information delivery. Prior research has reported on the delivery of feedback-rich formative e-assessment. Sim, Holifield, and Brown (2004) review the literature on using computer assisted assessment (CAA) and they identify the need to consider CAA, as the general trend

in higher education is towards increasing class sizes. In his case study research Nicol (2007, 2009b) also highlights the value of using information and communication technology to support formative assessment for large first year classes. Similarly, Carless and Boud (2018) recognise computer-based systems can be a source of information that stimulates the feedback process.

Carless and Boud (2018) also recognise the need to ensure feedback provided is of high quality and aids the development of students' feedback literacy. This involves considering the 'understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies' (p1316). Thus, teaching staff need to consider how students engage with feedback messages and how we can help them to do so (Noble et al. 2020). It is important to consider the students' views of feedback and to ensure feedback messages are clear in order to enhance feedback literacy and feedback recipience (Winstone et al. 2017).

It is therefore essential that feedback is viewed as an ongoing process that offers students a chance to engage with the feedback information provided and improve their work and their feedback literacy: it is more than just an end of module activity, or something that takes place at a single point in time (Carless and Boud, 2018; Molloy, Boud and Henderson, 2019). It is recognised that building feedback literacy is a complex and challenging process (Sutton, 2012), hence it is important to begin in first year (Molloy, Boud and Henderson, 2019), thus offering students an early chance to engage with feedback, actively respond to it and develop deep learning strategies.

Walker, Topping, and Rodrigues (2008) examine student reflections on formative e-assessments and conclude that formative e-assessments can support deep learning, but consideration needs to be given to the role of formative assessment and the quality of information provided.

The higher education literature has examined this issue and identified key principles of assessment and good feedback practice (Gibbs and Simpson 2004; Nicol and Macfarlane-Dick 2006 and Nicol 2009b), covering both formative and summative assessment. Adopting these principles, for example providing 'information to teachers that can be used to help shape their teaching' (Nicol 2009b, 5, principle 12) and providing 'sufficient feedback ... both often enough and in enough detail' (Gibbs and Simpson 2004, 17, principle 4) helps create rich formative e-assessments. Pedagogically designing quality information to stimulate the feedback process within formative assessment, based on the principles above, was a prime focus throughout the project.

The effectiveness of formative assessment

Many studies examine assessment with dual purpose (Dawson et al. 2019) or focus on effectiveness in terms of course results and overall pass rates (Walker, Topping, and Rodrigues 2008; Einig 2013; Perera, Nguyen, and Watty 2014). The accurate measurement of the effectiveness of formative assessment and feedback is recognised as a difficulty (Price et al. 2010), and there is no one measure which can achieve this. While formative assessments may be used to improve student performance, many also have a summative element. Looking to student results on summative assessment as a measure of effectiveness may give some indication of increased student effort through further study (Perera, Nguyen, and Watty 2014). This is important, as 'feedback' is deemed ineffective if students do not act

on it (Gibbs and Simpson 2004). However, summative results do not give any indication of the impact on students' ability to monitor their own progress. Studies have shown that if students learn to monitor their own progress, this makes them more independent learners (Nicol 2007, 2009a, 2009b).

As identified by Harlen and James (1997), purely formative assessment has many purposes, including:

- (1) A focus on promoting learning
- (2) Providing diagnostic information and
- (3) Allowing students to understand strengths and weaknesses and how they might deal with them.

Therefore, any effectiveness measure should be based on these principles. Yorke (2003) agrees, the effectiveness of formative assessment relies on addressing two main questions: firstly, is the 'feedback' the best that could be provided and secondly 'did the formative assessment influence student behaviour?' (Yorke 2003, 484). Therefore, these factors will be used to judge the effectiveness of the formative assessments in this study.

Methodology

Method of implementing regular formative e-assessment:

An integrated series of regular formative e-assessments was introduced across two courses normally taken by first year accounting students in the first semester: namely, Financial Accounting 1 (FA1) and Introduction to Business Statistics (IBS). These are two of the three

compulsory courses taken by first year students at this stage in their studies, therefore the cohort taking both courses is almost identical. In FA1 there were 129 students and in IBS there were 116 students: the difference is due to the presence of visiting students.

To ensure the 'feedback' is the best that could be provided (Yorke, 2003), the formative e-assessments were designed with reference to the principles of good formative assessment and feedback as outlined in Nicol (2009b) and Gibbs and Simpson (2004). All principles were considered in the design of the regular formative e-assessment, but particular emphasis was given to Gibbs and Simpson (2004) conditions 1, 4, 6 and 8 and to Nicol (2009b) principles 2, 3, 7 and 12 (see Table 1).

[Table 1 near here]

The formative e-assessments were delivered via Moodle's quiz function. (Students could complete the e-assessments on or off campus.) Moodle's functionality and flexibility allowed the course co-ordinators to construct quizzes that could be made available at specified intervals: thus, facilitating regularity of student study across the two courses across first semester. Each test was made available for a period of one week only to encourage regular study and to provide regular information that could be used (by the student) to activate feedback processes. The programme of formative e-assessments, as illustrated by the timeline below, began in week 2, starting with low level material then increased in difficulty throughout the semester. Each test covered material from the teaching weeks between the last test and the current test. As FA1 has more credit weighting than IBS it had more formative e-assessments.

Timeline for formative e-assessments

Reading & **Examinations Teaching** Week 2 3 4 7 10 11 12 1 8 Formative FA₁ FA1 **IBS** FA1 **IBS** FA1 assessment Project Summative **Project** Final assessment examinations

The formative e-assessments were not compulsory. The value of regular formative assessment was explained to students and they were strongly encouraged to work regularly across both courses during the semester and to engage with the formative e-assessments. Completion rates for each course were very high, with 94% completing all FA1 assessments and 89% completing all IBS assessments. (The high completion rates may be due to the fact the courses are in year 1, semester 1 and students tend to take a lead from staff in this transition phase.)

Questions on each formative e-assessment (and diagnostic information) were created by the course co-ordinators and covered a variety of different formats including multiple choice, multiple response, narrative and computational questions. Earlier e-assessments tended to focus on lower order skills such as knowledge and comprehension. However, as the semester progressed the complexity of questions increased and moved towards

application and analysis skills. Upon completion of each formative e-assessment, students were given immediate information on their performance in individual questions (tailored according to the answer they provided) and feedback on their overall performance. This information was also available to academic staff – thus allowing identification of problem topics or individual students who were struggling with the course. Having developed the formative e-assessments with respect to the objectives discussed above, it was necessary to evaluate student and staff views on the success of these interventions.

Method of evaluating regular formative e-assessment:

Student opinion was gathered using end of course questionnaires administered via Moodle. Although these were not compulsory, response rates were high with 74% completion for FA1 and 78% completion for IBS. Students were asked questions that were aligned with the following criteria defining effectiveness of the regular formative e-assessments, as developed from Harlen and James (1997), Yorke (2003) and Price et al. (2010):

- Students learning to monitor their own progress.
 This relies on students engaging with the information received to generate feedback and become more reflective learners.
- Encouraging further study
 This is particularly important in this case, as due to the timing of end of semester examinations, the study period at the end of the course is very short.
- Students perceiving an increase in their learning and understanding of the relevant course.

If students see a benefit from the formative assessment process and feel they have improved their learning, this suggests they have engaged with the process.

 Staff gaining useful information that can be utilised in a timely manner to inform and improve their delivery of the course.

As the students complete the formative e-assessment staff are presented with information that allows them to identify students who are at risk and problem topics.

Identical questionnaires were administered in both FA1 and IBS, as the course co-ordinators wanted to be able to compare the effectiveness of the formative e-assessment across each course and identify any differences. A series of Likert scale questions were added to the Quality Assurance required end of course questionnaire and then followed up with narrative questions to elicit more details from students. The Likert scale questions had a 5-point scale ranging from 1=strongly disagree to 5=strongly agree (with 3 indicating a neutral response).

Given the aim of the study is to measure effectiveness of the regular formative e-assessments, based on the student voice, the effectiveness measures (identified above) were incorporated into the questionnaire as follows:

Effectiveness Measure	Corresponding Likert Scale Questions				
Students learning to monitor	27d	Undertaking the Moodle Quizzes helped me to			
their own progress		assess my progress in FA1/IBS			
Encouraging further study	27c	The Moodle Quizzes encouraged me to study			
		regularly			
Students perceiving an	27a and	Having to work regularly (on Quizzes) helped			
increase in learning and		me to learn more			

understanding

27b I found the feedback from Quizzes helped me to understand FA1/IBS

The final measure of effectiveness relating to staff use of information is evaluated via staff reflections. While staff reflections did not follow a formal process, they were effectively aligned with Gibbs' reflective cycle (Gibbs, 1988), following the six stages through description, feelings, evaluation, analysis, conclusion and action plan. After each formative e-assessment staff would consider what happened (description stage); how they felt about the formative assessment (feelings); what went well and not so well (evaluation); the reason behind success and failure (analysis); what could be better / what have we learned about delivering regular formative e-assessment (conclusions) and practical considerations for future improvements (action plan). Staff would meet regularly to discuss their reflections and to try to reach consensus on any alterations required. They also reflected on the overall series of regular formative e-assessments at the end of the semester, once student comments were available.

Descriptive statistics and hypotheses development:

The first step of the analysis for this study was to examine the Likert scale results.

Descriptive statistics were calculated and examined. In order to establish whether or not the formative e-assessments were effective, Sign test of the median were conducted using the following hypotheses:

- (1) H0: Completing the formative assessment has no impact on students' ability to monitor their own progress
 - H1: Completing the formative assessments helped students learn to monitor their own progress
- (2) H0: Completing formative assessments had no impact on further study
 H1: Completing formative assessments encouraged further study
- (3) H0: Completing formative assessments had no impact on a student's perceived level of learning and understanding
 - H1: Completing formative assessments increased a student's perceived level of learning and understanding

In each case the median was assumed to be equal to 3 (the mid-point) and the alternative hypothesis assumed the median was greater than 3. Frequency tables were then constructed (see Table 3) and the analysis extended using non-parametric Mann Whitney U tests (see Table 4) to check if there were any relationships between the student responses and the courses the students were taking. Given the student cohorts were largely the same we wanted to ensure there were no idiosyncrasies in the results. The results of these tests are discussed in the next section.

Further analysis:

After this, the students' comments were subjected to a close scrutiny and thematic analysis was selected as an appropriate method to identify themes within the data. As identified by Clarke and Braun (2013) and Braun and Clarke (2006), thematic analysis normally follows

six stages: although it is a recursive process, rather than a linear model. Firstly, the data was read over several times to gain familiarity with student comments. Second, the data was coded both with reference to the measures of effectiveness already identified and from an inductive view to search for any other prevalent issues. The third stage is to search for themes within the coded data. Clarke and Braun (2013, p. 121) describe themes as "a coherent and meaningful pattern in the data". At the fourth stage themes were reviewed and discussed between the researchers, with coding and themes selection reviewed and updated. Stage five and six are defining and naming themes and writing up. Details of the themes identified, sample student comments and a discussion of these is included in the following sections.

Results

Descriptive statistics for student responses to the 5-point Likert scale questions and results of the Sign test are shown in Table 2.

[Table 2 near here]

The descriptive statistics indicate the median for each question is greater than 3. At a level of significance of 5%, the Sign Test of the Median provides sufficient evidence to infer that the median results are not only statistically different from 3, but suggest they are higher. This shows students agree or strongly agree with the statements and therefore find the formative e-assessments are effective. Having established that the results are significantly different from 3, the analysis then moved to considering the information in Table 3. In both courses, the response to most questions is similar; providing evidence there is a homogeneous cohort.

[Table 3 near here]

The Mann Whitney U Tests (Table 4) was conducted and indicated that in all cases the two populations, based on the sample of students from FA1 and IBS, are identical. This result shows that there is sufficient evidence to suggest that the answers are not dependent on the course taken by the students.

[Table 4 near here]

Additional Chi-Squared tests of independence and goodness of fit were conducted. The test of independence confirmed the level of agreement with each question is independent of the course taken; a similar result to the Mann-Whitney U test. The Goodness of Fit Test showed that there were no differences in the distribution of answers for each course; except for question b. This ties in with the analysis and discussion of the student comments.

[Tables 5 and 6 near here]

The themes identified in the student comments are shown in Tables 5 and 6, along with sample verbatim quotes. Table 5 identifies positive themes/comments while Table 6 shows themes related to negative aspects of the process and changes students suggested. The positive comments mainly relate to the measures of effectiveness, while the negative comments and suggested changes relate to the functionality of the system. These are discussed fully in the next section.

[Table 7 near here]

The final results are staff reflections on the process. Table 7 shows a sample of staff reflection (after the first formative e-assessment took place), matched against the stages of Gibbs' reflective cycle. Staff reflections are also discussed further in the next section.

Discussion:

Each measure of effectiveness is now discussed in turn – with reference to the statistical results and student responses. For the first measure (students learning to monitor their own progress) the Sign Test for question d (Table 2) indicates hypotheses 1 can be rejected and there is evidence that regular formative e-assessment helped students to assess their progress with each course, a necessary pre-requisite for students to be able to take action and improve (Black and Wiliam 1998). Information from the narrative comments are analysed to gain further insight into student views on this measure of effectiveness.

Hypotheses 2 can also be rejected for both courses (per Table 2, question c) providing evidence the formative e-assessments encouraged students to study regularly (the second measure of effectiveness). This agrees with the findings of Weurlander et al. (2012) which shows formative assessment motivates students to study. This was an important objective from the design stage; as there is very little revision time between the end of the course and the final examination – therefore it is imperative students keep up to date with their studies. The Mann Whitney *U* Test showed that both populations are identical i.e. response does not depend on the course taken. The Chi-Squared Goodness of Fit Tests also failed to find a significant difference between the distribution of responses for the two courses. However, Table 3 indicates a higher proportion of students answered negatively (disagree or strongly disagree) for IBS. Therefore, while students found the formative e-assessment encouraged regular study on both courses, this was more prominent for FA1. This will be discussed later, but could be a consequence of IBS having fewer formative e-assessments. These results suggest that the second measure of encouraging the students to work consistently throughout the semester is achieved, but perhaps more frequent formative e-assessment is preferable.

This point will be matched with student comments to obtain more information on the desired frequency of formative e-assessments.

If students perceive a benefit from participating in the regular formative e-assessments, this in itself is likely to lead to greater engagement and therefore more chance of satisfying the first two measures. The results of the Sign Test for questions a and b (detailed in Table 2) show that hypotheses 3 can be rejected. Students believe that completing the formative e-assessment tests helped them to learn more and improve their understanding of each course. Therefore, the third effectiveness measure (students perceive an increase in their learning and understanding) has been successfully achieved. This suggests students can enhance their work by making sense of the feedback information provided, which supports Carless and Boud's (2018) view that students can develop their feedback literacy and enhance their learning.

Although the Mann Whitney *U* Test did not find any significant differences between the two courses for question a or b, the Chi-Squared Goodness of Fit Test did detect a significant difference in the distribution of responses to question b. This suggests students found the feedback from FA1 helped their understanding more than the feedback from IBS – although both sets of feedback were identified as helpful to students. The difference can be seen by looking at the level of disagreement with question b (see Table 3). There was a much larger proportion of students disagreeing or strongly disagreeing with statement b for IBS. This could be a result of fewer opportunities to receive feedback for IBS or a perceived difference in feedback quality. Walker, Topping, and Rodrigues (2008) show the provision of feedback information that is insufficiently detailed can cause frustration and impact on student understanding. This point will be evaluated further when considering student comments.

Student Comments

Students were asked to comment on three areas of the formative e-assessments: what they liked, what they disliked and what they would change. Analysis of the comments revealed recurring themes in student responses (as noted in Tables 5 and 6), these are discussed in turn and related (where applicable) to the measures of effectiveness of the formative assessments.

Positives

Students identified many features of the e-assessment that they liked, but there were four main themes that recurred in the comments of both courses. Each theme will be discussed in turn.

1. The e-assessments aided understanding and encouraged regular study. The formative e-assessments provided students with an opportunity to engage with the content of the courses, test their understanding and then reflect on their progress using the feedback provided. The regular nature of the e-assessments also motivates students to study regularly throughout the semester (in line with the findings of Weurlander et al. 2012), rather than leaving all work until just before the examination. As noted already, response to the Likert scale questions suggest this framework was successful and this is supported further by student comments which explicitly state that their understanding was improved and they studied more regularly (see Table 5).

This relates back to the principles of good formative assessment and feedback as outlined by Gibbs and Simpson (2004) and Nicol (2009b). These criteria were identified at the design stage of the project, which suggests the careful consideration of such criteria is vital. In terms

of the effectiveness of the regular formative assessment the student comments under this theme align with two of the effectiveness measures. Firstly, students perceiving an increase in their learning and understanding of the relevant course and secondly encouraging further study. Both measures are supported by the comments and this adds further weight to the findings from the Likert scale questions.

2. Online functionality of the tests. Although this is not an explicit measure of effectiveness, this was a clear theme identified within the student comments and one which aligns with Gibbs and Simpson (2004) principle 6. Feedback must be timely in order to be of maximum benefit to students. With large staff to student ratios, the use of technology helped to achieve this criterion (Sim, Holifield, and Brown 2004). Students liked the functionality of the formative e-assessments and the freedom given by their online nature, as demonstrated by the comments in Table 5.

They appreciated the fact that tests could be completed at a time to suit them, which fitted with their workloads and other commitments – even if they were not on campus. Technology enabled approaches are generally welcomed by students (Carless and Boud 2018) and they can facilitate timely sharing of comments.

3. Provision of detailed and timely feedback. The quality of the feedback provided is a key element to the success of this project. The design of the formative e-assessments ensured that feedback is timely (as it is provided immediately). However, the course co-ordinators felt strongly if the feedback provided is of poor quality and does not help students to reflect on their performance and improve in future, then the formative e-assessments would be pointless. Students' comments on the feedback given (see Table 5) generally identified the

feedback as one of the positive features – indicating both timeliness and quality are important.

Thus, the effectiveness of the feedback on the formative assessment is twofold. The feedback allows students to monitor their own progress and comments suggest students also perceive an increase in their learning and understanding (in line with Nicol 2009b, principle 3 and 7). The thematic analysis highlighted an interesting point, students noted that there are not many opportunities for them to receive instant feedback, and they find it helpful when this is available. In a practical sense, the most efficient way to achieve this (with larger class sizes) is using technology (Sim, Holifield, and Brown 2004).

The students did comment on some differences between the feedback provided for each course. This aligns with the findings from the Likert scale questions, as the only area where the Chi-Squared Goodness of Fit test found a difference between the two course distributions was in relation to feedback. This will be considered further under student comments on changes they suggested for the future.

4. Ability to assess own progress and evaluate own learning. This is an explicit measure of effectiveness of the formative assessment as Harlen and James (1997) recognise the need for students to use the comments provided in order to make progress while Black and Wiliam (1998, 2009) recognise the importance of providing comments that students can use to identify areas for improvement. Students highlighted these benefits in their comments (see Table 5).

As there were a series of e-assessments which built upon each other, later e-assessments allowed students to test if they had successfully corrected any earlier

misconceptions. This result agrees with the Likert scale question results noted earlier. The students did feel they were better able to monitor their own progress, and this is an important skill to develop for lifelong learning (Nicol 2010).

Negatives and suggested changes

Students were also asked to comment on features of the formative e-assessments that they did not like. There was one very apparent theme to student responses (see Table 6) that related to the functionality of Moodle and its limited ability to recognise the correct input for textual and numerical questions. If the students' response contained spelling mistakes or was input in the wrong format, the system did not recognise their answer. This problem became apparent to staff soon after the first e-assessment commenced. However, it is important to note that the students did not identify any major dislikes in relation to the regular formative e-assessment and feedback information provided – their comments related to the functionality of Moodle, the Virtual Learning Environment (VLE) used to deliver the assessments, rather than measures of effectiveness.

Walker, Topping, and Rodrigues (2008) noted similar problems with fill in the blank style online questions, which raised issues on the fairness and validity of some types of question. This shows how the functionality of the VLE can impact student experience and damage their confidence in the formative e-assessments. However, students' desire to receive credit for marking errors on the system highlights their high level of engagement. As the e-assessments were formative and did not count towards their final grade students could simply have accepted the incorrect grading information. However, based on the feedback they received, they recognised they had the correct answer and did not passively accept the grading information provided but actively contacted staff and sought amendments to the VLE

system. Later formative e-assessments were updated to include instructions on how answers should be entered in each affected question and this solved the problem. There were some other common comments on frequency, style and length of the e-assessments under factors students disliked: these overlap with suggested changes and are discussed below.

Students were asked to list up to three things they would change. There was much repeated reference to the incorrect treatment of answers (discussed above) and some points on the formatting (and differences) between the two courses. For example, each e-assessment was available to students for one week only. In FA1 the co-ordinator took this to be a calendar week (and hence included weekends) whereas the IBS co-ordinator interpreted this as a working week (Monday to Friday only). Students noted this discrepancy and asked that all e-assessments be available over the weekend. Students also requested some minor amendments to the appearance and formatting of the e-assessments – such as including all questions on one page - these are quickly and easily complied with and were actioned by the course co-ordinators.

Otherwise, there was a noticeable difference in the response to this section for each of the courses. For FA1 the majority of students (62%) either provided only a positive comment or stated they would not make any changes to the e-assessment programme. For IBS this proportion was smaller (30%). Analysing the comments on suggested improvements provides some interesting insights into student perceptions of the differences between the two sets of formative e-assessment. The most common suggested improvement on the IBS questionnaires is in relation to feedback information provided (15% of students commented on this aspect). IBS Students indicated that feedback information could be more detailed. In comparison only 2% of students in FA1 noted a comment in relation to the quality of

feedback information provided. Walker, Topping, and Rodrigues (2008) noted similar issues, as students find the quality of feedback information can cause frustration and limit their opportunities to improve their learning. While Dawson et al. (2019) highlight student perceptions show that what makes feedback effective is high quality comments. This aligns with Yorke's (2003) view that 'feedback' should be the best that could be provided in order to be effective. The FA1 feedback was more detailed (utilising all the different levels of feedback the VLE allowed) and students seemed to find this helpful.

This could help explain the difference in the Likert scale distributions for the question 'I found the feedback from Quizzes helped me to understand FA1/IBS' (Table 3). The Likert responses, coupled with student comments, indicate the additional feedback provided for FA1 was beneficial. This is in line with principles of best feedback practice: specifically, Nicol (2009b) principle 3 and Gibbs & Simpson (2004) principle 8. However, it may not be the difference in feedback quality that causes the difference of opinion, but the difference in feedback frequency. Students also noted that they would prefer more frequent (and perhaps shorter) e-assessments, particularly for IBS (see Table 6).

In many ways, these comments are not surprising given that FA1 had twice as many formative e-assessments as IBS. Students generally seem to favour short regular e-assessments over less frequent, but longer tests. This again fits with the Likert scale results. While there was not a significant difference between the responses to question c (on regular study) there was an indication that more regular tests encourage more regular study – hence more frequent formative e-assessments is beneficial. This is in alignment with current literature which suggests early and regular formative assessment can aid first year students'

understanding of what is required (Nicol 2009a) and should focus on helping students understand where they need support (Thomas et al. 2019).

Students also made other suggestions in relation to the e-assessments and noted that it would be helpful if they could access the e-assessments at a later date – particularly when studying for examinations – so they could retry the test and see if they were able to improve their performance. This is a very sensible suggestion and it falls in line with Nicol (2009b) principle 4. Therefore, in future years, formative e-assessments were reopened and made available to students during the revision period prior to the final degree examinations.

Staff reflections

Staff reflections were considered as the final part of the evaluation of the effectiveness of the formative e-assessments. The project team benefited significantly from reflecting on the student comments. In general, not only were staff satisfied with the results of introducing the e-assessments, i.e. they supported students in their learning, but felt that they had gained valuable insight into student progress throughout the semester: a key part of the formative assessment process (Black and Wiliam, 1998, 2009).

After each formative e-assessment was completed staff could review a range of diagnostic information on student performance. This included a graph depicting the overall results for the cohort. Information on an individual student's performance could be broken down either by question or by formative e-assessment. Therefore, staff were able to identify problem topics and they could also identify students who had difficulties and were not progressing well. The diagnostic information was the largest benefit of the e-assessments, for

staff – and also falls in line with Nicol (2009b) principle 12 of best feedback practice and it adds to staffs' knowledge of student achievement (Bennett 2011).

After each formative e-assessment staff could reflect on its successes and failures and identify any necessary actions. Details of actions taken are shown in Table 8 and discussed in full below.

[Table 8 near here]

Students who performed poorly were identified and invited to meet with the course coordinator to discuss any difficulties and identify any action they needed to take, as supportive dialogue is a key part of feedback processes (Price et al. 2010). Most students responded quickly and in the majority of cases students would reply to say they had used the test as a diagnostic mechanism to identify areas of weakness that required further study, and they were confident they could move forward on their own (based on the feedback information received), and therefore did not need to meet with staff to discuss progress. This further highlights the effectiveness of the regular formative assessment as it shows students are learning to monitor their own progress and engaging in further study (as intended) and confirms both staff and students benefit from the diagnostic information provided by formative assessment (Harlen and James, 1997). The remaining students would meet with staff (more than once, if required) to gain assistance on how to improve. Students appreciated this intervention and felt staff cared about their progress. After each test staff would collate results (across their course) to monitor individual student performance over time. Collating information into one Excel spreadsheet could be time consuming, but it aided monitoring and was judged as necessary.

Another benefit from the diagnostic information was the ability to adapt teaching practice in response to problem topics. Rearranging the spreadsheet data to show overall performance by question, allowed identification of these topics. Where a question (or set of questions) indicated a general lack of understanding of a particular topic then immediate action could be taken to rectify this issue (as shown in Table 8). For example, FA1 tests indicated that many students did not understand the difference between bad and doubtful debts. In response, additional lecture time was devoted to this topic and it was revisited during the revision session. The ability to have real time information on student progress that can impact on course delivery highlights the effectiveness of the regular formative e-assessments, as discussed by Black and Wiliam (2009). Therefore, the fourth measure of effectiveness of the formative e-assessments was achieved, staff gaining useful information they were able to utilise in a timely manner to inform and improve course delivery.

While the University's VLE provided the student performance diagnostic information, using the VLE showed other benefits and drawbacks from the point of view of the course coordinators. Joint reflective discussions (at the end of the process) allowed staff to evaluate the overall series of regular formative e-assessments on a holistic basis. They identified benefits and drawbacks of the intervention. Two benefits of using this VLE were: (1) all students on a course would automatically have access to the e-assessments and (2) questions could be individually tailored to the course material. This is particularly important in Business Statistics where different uses of notation could cause confusion. While many publishers offer ready-made question banks, which accompany particular textbooks and would avoid notational differences, access to the test bank relies on students purchasing a

new copy of the textbook. As not all students will buy the textbook, accompanying resources cannot be used as a compulsory element of a course.

While the ability of all students to access course specific and tailored questions is the biggest benefit of using the VLE, it is also the biggest drawback. The course co-ordinators felt that while there is no monetary outlay there is still a significant cost – namely staff time required to create suitable questions and accompanying feedback. This is an initial cost in the first year that will not necessarily be repeated in future, as the question bank can be modified and updated. As publisher question banks develop and different access models are explored, these may become a viable alternative that would eliminate or reduce the initial set-up time.

Other recurring demands on staff time are the administration time spent examining the results of the e-assessments to identify problem areas and students who are experiencing difficulties. Time demands can be significant and can be difficult to accommodate during a busy teaching semester. However, they are worthwhile in terms of the improved information and appropriate action that can be taken on a timely basis; but administrative support could reduce this burden.

After reflecting on the information gathered and the impact of the formative e-assessments, the following action plan was made to identify revisions to the regular formative e-assessments in future years:

 Additional guidance was provided on how to enter numerical and textual answers (to reduce the likelihood of correct answers receiving incorrect grades).

- Unintentional discrepancies between time frames and formatting of e-assessments in IBS and FA1 were eliminated.
- A particularly long e-assessment for FA1 was split into two shorter e-assessments.
- Questions and feedback were updated appropriately.
- Administrative assistance was obtained to aid staff with monitoring of the results.

While the course co-ordinators recognise that the time element is significant, the Likert scale question results and the student comments support the effectiveness of integrating a regular series of formative e-assessments to a first year curriculum. Staff also recognise there are benefits to having a better understanding of formative assessment practice, but feel strongly that administrative support in collating results would free academic time to focus on supporting students.

Conclusions and recommendations

This study added to our understanding of the staff and student experience of regular formative e-assessments and allowed us to judge their effectiveness. It examined the introduction of formative e-assessments, built on the informed views of staff attempting to *improve* or enhance the student experience. The project *involved* both staff reflections and student comments in the continued development of those formative e-assessments. Staff deepened their understanding of the student experience and were able to improve their practice.

Based on the results, there is evidence of the benefits of introducing regular formative e-assessment across year one courses. The results show the students feel the regular formative e-assessment were effective at:

- (1) helping them learn to monitor their own progress;
- (2) encouraging further study and
- (3) increasing a student's perceived level of learning and understanding

These are all important indicators of student engagement and the move towards becoming independent learners.

Staff also benefitted, by gaining insights into student progress. Such insights helped them improve the delivery of their course, by responding to issues in a timely manner.

Reflection on the students' comments also provides supporting evidence which suggests fostering student engagement has the potential to help students build the skills required to become more reflective learners.

The study makes it clear that the best structure for e-assessments needs careful thought. Recommendations based on the evidence gathered are that the e-assessments are: short; frequent; feedback rich – students prefer comprehensive, detailed, multi-level feedback and repeated - giving students an opportunity to revisit the e-assessments so they can determine the level of improvement in their performance.

This paper has added to our understanding by listening to the student voice and reporting on that voice in a much more detailed fashion, and in conjunction with staff reflections. Of course, the project had several limitations. To address these limitations, we suggest future research should consider: (1) Whether the student view changes as they progress through their studies. (2) Looking across all the courses in the same level. This extension could confirm if there are course specific influences. (3) Using a variety of formats

for formative assessments e.g. including peer review exercises and examining if variety improves effectiveness. (4) The nature and quantity of feedback information in relation to student expectations. (5) The balance between question length, question style and test frequency. (6) The time staff have to spend on developing and responding to formative assessments.

Addressing these issues would help realign our efforts in developing effective formative assessments that are fit for purpose and confirm their place within higher education. Given the threat of extinction that formative assessment faces (Wu and Jessop 2018) this is an urgent area for further research.

Reference List

- Bennett, R. E. 2011." Formative assessment: a critical review." *Assessment in Education: Principles, Policy & Practice* 18 (1): 5-25. doi:10.1080/0969594X.2010.513678
- Black, P., and D. Wiliam. 1998. "Inside the Back Box Raising Standards Through Classroom Assessment." *Phi Delta Kappan* 80 (2): 39-148.
- Black, P., and D. Wiliam. 2009. "Developing the theory of formative assessment." *Educational Assessment, Evaluation and Accountability* 21 (1):5-31.
- Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3(2): 77-101. doi: 10.1191/1478088706qp063oa
- Carless, D., and D. Boud. 2018. "The development of student feedback literacy: enabling uptake of feedback." *Assessment & Evaluation in Higher Education* 43 (8): 1315-1325. doi: 10.1080/02602938.2018.1463354.
- Clarke, V., and V. Braun. 2013. "Teaching Thematic Analysis: Overcoming Challenges and Developing Strategies for Effective Learning." *The Psychologist* 26(2): 120-123.
- Dawson, P., M. Henderson, P. Mahoney, M. Phillips, T. Ryan, D. Boud, and E. Molloy. 2019. "What makes for effective feedback: staff and student perspectives."

 **Assessment & Evaluation in Higher Education 44 (1):25-36. doi: 10.1080/02602938.2018.1467877
- Einig, S. 2013. "Supporting student's learning: the use of formative online assessments." *Accounting Education: An International Journal* 22 (5): 425-444. doi: 10.1080/09639284.2013.803868.
- Gibbs, G. 1988. Learning by Doing: A guide to teaching and learning methods. Further Education Unit. Oxford Polytechnic: Oxford
- Gibbs, G., and C. Simpson. 2004. "Conditions under which assessment supports students' learning." *Learning and Teaching in Higher Education* 1:3-31. Accessible at:
- http://eprints.glos.ac.uk/3609/1/LATHE%201.%20Conditions%20Under%20Which%20Asse ssment%20Supports%20Students%27%20Learning%20Gibbs_Simpson.pdf
- Harlen, W. and M. James. 1997. "Assessment and Learning: differences and relationships between formative and summative assessment." *Assessment in Education: Principles, Policy & Practice* 4 (3): 365-379. doi: 10.1080/0969594970040304.

- Molloy, M., D. Boud and M. Henderson. 2019. "Developing a learning-centred framework for feedback literacy." *Assessment & Evaluation in Higher Education* doi: 10.1080/02602938.2019.1667955
- Nicol, D. 2007. "Laying a foundation for lifelong learning: Case studies of e-assessment in large 1st year classes." *British Journal of Education Technology* 38 (4):668-678. doi: 10.1111/j.1467-8535.2006.00657.x.
- Nicol, D. 2009a. "Assessment for learner self-regulation: enhancing achievement in the first year using learning technologies." *Assessment & Evaluation in Higher Education* 34 (3):335-352. doi: 10.1080/02602930802255139.
- Nicol, D. 2009b. "Transforming assessment and feedback: enhancing integration and empowerment in first year", *Quality Enhancement Themes report on first year experience* accessible at:
- https://www.enhancementthemes.ac.uk/docs/ethemes/the-first-year/transforming-assessment-and-feedback.pdf?sfvrsn=c62f981_12
- Nicol, D. 2010. "From monologue to dialogue: improving written feedback processes in mass higher education." *Assessment & Evaluation in Higher Education* 35 (5): 501-519. doi:10.1080/02602931003786559.
- Nicol, D., and D. Macfarlane-Dick. 2006 "Formative assessment and self-regulated learning: a model and seven principles of good feedback practice." *Studies in Higher Education* 31 (2): 199-216. doi: 10.1080/03075070600572090.
- Noble, C., S. Billet, C. Sly, L. Collier, L. Armit, J. Hilder, and E. Molloy. 2020 "It's yours to take": generating learner feedback literacy in the workplace." *Advances in Health Sciences Education* 25: 55-74. doi: 10.1007/s10459-019-09905-5
- Perera, L., H. Nguyen, and K. Watty. 2014. "Formative Feedback through Summative Tutorial-based Assessments: The Relationship to Student Performance." *Accounting Education: an International Journal* 23 (5): 424-442. doi: 10.1080/09639284.2014.947093.
- Price, M., K. Handley, J. Millar, and B. O'Donovan. 2010. "Feedback: all that effort, but what is the effect?." *Assessment & Evaluation in Higher Education* 35 (3): 277-289. doi: 10.1080/02602930903541007.
- Pryor, J., and B. Crossouard. 2008. "A socio-cultural theorisation of formative assessment." *Oxford Review of Education* 34 (1): 1 20. doi: 10.1080/03054980701476386.

- Pryor, J., and B. Crossouard. 2010. "Challenging formative assessment: disciplinary spaces and identities." *Assessment and Evaluation in Higher Education* 35 (3): 265-276. doi: 10.1080/02602930903512891.
- Sadler, D. R. 1998. "Formative Assessment: revisiting the territory." *Assessment in Education: Principles, Policy & Practice* 5 (1): 77-84. doi: 10.1080/0969595980050104.
- Sim, G., P. Holifield, and M. Brown. 2004. "Implementation of computer assisted assessment: lessons from the literature." *Research in Learning Technology* 12 (3):215-229. doi: 10.3402/rlt.v12i3.11255.
- Sutton, P. 2012. "Conceptualizing feedback literacy: knowing, being, and acting." Innovations in Education and Teaching International 49(1), pp. 31-40. doi: 10.1080/14703297.2012.647781
- Thomas, T., Jacobs, D., Hurley, L., Martin, J., Maslyuk, S., Lyall, M. and Ryan, M. (2019) Students' perspectives of early assessment tasks in their first-year at university, *Assessment & Evaluation in Higher Education*, 44(3), pp. 398-414. doi: 10.1080/02602938.2018.1513992.
- Walker, D. J., K. Topping, and S. Rodrigues. 2008. "Student reflections on formative e-assessment: expectations and perceptions." *Learning, Media & Technology* 33 (3):221-234. doi: 10.1080/17439880802324178.
- Weurlander, M., M. Soderberg, M. Scheja, H. Hult, and A. Wernerson. 2012. "Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment." Assessment & Evaluation in Higher Education 37 (6):747-760. doi: 10.1080/02602938.2011.572153.
- Winstone, N.E., R.A. Nash, M. Parker, and J. Rowntree. 2017. "Supporting Learners' Agentic Engagement with Feedback: A Systematic Review and a Taxonomy of Recipience Processes." *Educational Psychologist* 52(1): 17-37. doi: 10.1080/00461520.2016.1207538
- Wu, Q., and T. Jessop. 2018. "Formative assessment: missing in action in both research-intensive and teaching focused universities?." *Assessment & Evaluation in Higher Education* 43 (7): 1019-1031. doi: 10.1080/02602938.2018.1426097.

Yorke, M. 2003. "Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice." *Higher Education* 45 (4): 477-501. doi: 10.1023/A:1023967026413.

Table 1: Principles of good formative assessment and feedback

Gibbs and Simpson (2004)

Condition 1: Sufficient assessed tasks are provided for students to capture sufficient study time.

Condition 4: Sufficient feedback is provided. Both often enough and in enough detail.

Condition 6: The feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance.

Condition 8: Feedback is appropriate, in relation to students' understanding of what they are supposed to be doing.

Nicol (2009b)

Principle 2: Encourage time and effort on challenging learning tasks.

Principle 3: Deliver high quality feedback information that helps learners to self-correct.

Principle 7: Facilitate the development of self-assessment and reflection in learning.

Principle 12: Provide information to teachers that can be used to help shape their teaching.

Table 2: Descriptive statistics and Sign test results

			Median	Mean	Standard Deviation	Sign test p-value
a	Having to work	FA1	4	3.67	1.09	0.001
	regularly (on Quizzes) helped me to learn more	IBS	4	3.77	1.07	0.001
b	I found the feedback	FA1	4	3.71	1.10	0.001
	from Quizzes helped me to understand FA1/IBS	IBS	3.5	3.41	1.15	0.0013
c	The Moodle Quizzes	FA1	4	3.48	1.05	0.001
	encouraged me to study regularly	IBS	4	3.35	1.14	0.0016
d	Undertaking the	FA1	4	3.61	1.01	0.001
	Moodle Quizzes helped me to assess my	IBS	4	3.59	1.02	0.001
	progress in FA1/IBS					

			Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total
a	Having to work regularly (on Quizzes) helped me	FA1	6 14	7	18 19%	43 67%	6 19	93
	to learn more	IBS	2	12	14 16%	36 68%	6 24	88
b	I found the feedback from Quizzes helped me to understand	FA1	4	8	23 25%	31 62%	6 25	91
	FA1/IBS	IBS	5 23	15	24 27%	27	6 17	88
С	The Moodle Quizzes encouraged me to study regularly	FA1	5	9	28 31%	35 54%	6 14	91
		IBS	6 22	15	22 25%	32 51%	6 13	88
d	Undertaking the Moodle Quizzes helped me to assess	FA1	5	6	23 25%	63%	6 14	92
	my progress in FA1/IBS	IBS	3	7%	16 18%	43 64%	6 13	87

Table 3: Student evaluation of formative e-assessments

•

NOTE: Any missing values relate to students who chose not to answer a specific question.

Table 4: Statistical test results – Mann Whitney U Test

		Decision	Z
a	Having to work regularly (on	Accept	0.65
	Quizzes) helped me to learn more		
b	I found the feedback from Quizzes	Accept	-1.77
	helped me to understand FA1/IBS		
c	The Moodle Quizzes encouraged	Accept	-0.68
	me to study regularly		
d	Undertaking the Moodle Quizzes	Accept	-0.07
	helped me to assess my progress in		
	FA1/IBS		

Rejection rule: reject H_0 for z>1.96 or z<-1.96.

Table 5: Themes identified and sample student quotes – positives

1. The e-assessments aided understanding and encouraged regular study

The answers were informative which was good if you got it wrong in helping you to understand. Force you to read and keep up to date with work. (FA1)

it made me keep on top of my work. help me understand topics I didn't understand before through practice of the questions. (FA1)

Helped to motivate me to revise, outlining the difficulties I was having with each topic so I could then address them. (IBS)

2. Online functionality of the tests

Easy to fit around other coursework as they are available 24 hours (IBS) can do at home, can complete at own rate (FA1)

3. Provision of detailed and timely feedback

Immediate feedback. Helpedshow where I was going wrong. (IBS)

I liked the fact that it was one of the only resources we had where there were answers and instant feedback. (IBS)

The immediate feedback I received and how it showed you how to calculate the correct answer were very helpful in th course aswell as helping to find my weaknesses. (FA1)

4. Ability to assess own progress and evaluate own learning

Using the quizzes to assess individually how you are progressing in the course. (IBS)

helped you to identify what areas needed work (FA1)

Table 6: Themes identified and sample student quotes – negatives and suggested changes

Dislikes – functionality of Moodle (problems with how Moodle marks answers for numerical and textual questions)

text questions were very difficult to get right even with the correct answer (FA1)

My answers could have been right but due to the quiz not understanding the way I wrote them, I was marked wrong (IBS)

Suggested changes – more frequent e-assessment (particularly for IBS)

Only two were set over the time, perhaps more would encourage consistent learning ... over semester (IBS)

I think the moodle quizzes should possibly have been more frequent and covered a bit more of the course as there was a more regular pattern in ... Financial Accounting (IBS)

Table 7: Sample staff reflection after the first formative e-assessment

Stage of Gibbs' Reflective Cycle	Reflections
Description	All students completed the formative e-assessment and generally scored well. However, I was bombarded with emails from students stating their grade was incorrect. They stated they had the correct answer but were not receiving credit.
Feelings	Many students were aggrieved, and I was stressed by the volume of emails and worried that I had set up the tests incorrectly. I was also worried students would be discouraged and would disengage.
Evaluation	It was a positive sign that all students completed the formative e-assessment. I had access to a lot of data on student performance. Clearly some students were receiving inappropriate grade information which may damage confidence in the system (and would impact on the information I received). However, feedback was working as they were aware they had the correct answer.
Analysis	The system was working in a way I did not expect and could not control. There was no obvious way to correct the system and the student performance data I received would be affected by this problem.
Conclusion	The formative e-assessments were generally well received, but students were aggrieved when not awarded full credit (despite the fact the grade does not count towards their final mark). This shows engagement, but there is a need to find a solution to the early issues to maintain this engagement.
Action plan	Add additional instructions to numerical questions to help avoid a recurrence of marking errors on the system. Also, identify and contact poorly performing students (to offer additional support) and identify problem topics and take appropriate action.

Table 8: Staff actions

Actions taken after each test:

Download test results to Excel and sort by grade. (After the first test, results needed to be collated across tests. This took longer.)

Pick out students who performed poorly and identify any problematic topics (where many students struggled).

Look up student email addresses and contact poorly performing students and ask them to come and see me. (Non-response would be followed up.)

NOTE: Most students responded quickly to say they used the formative e-assessments as a diagnostic tool to determine how to focus their effort (and therefore didn't need to meet with me).

Meet with students and offer help/advice on areas where they are struggling and how to improve. (Follow up meeting if required.) Students appreciated the intervention and felt supported. Student performance on future tests would be monitored to see if students have progressed.

For any problem topics a range of measures would be considered/taken, including:

- Providing additional notes, examples, instructional videos or self-directed reading;
- Including additional tutorial questions or other worked examples;
- Spend additional lecture time on topics;
- Revisit the topic during revision sessions at the end of the course to help clarify understanding.