

THE IMPACT OF CHANGING ASSESSMENT: DOES IT MAKE A DIFFERENCE TO STUDENTS' GRADE PERFORMANCE?

Rafiuddin Ahmed, James Cook University

ABSTRACT

Assessment in higher education sector is widely researched for its relevance to outcome measurements in different forms such as grade performance, students' exit skills at graduation, and employability. One of the most widely used forms of assessment, summative assessment, is used to gauge students' performances in these dimensions. The current quasi-experimental study reports the impact of a change in assessment from a single summative paper-based mid-term examination to three short online, open book, continuous mid-term examinations in a third-year management accounting subject. The study finds that students' performance, as measured by marks in the final exam and overall marks in the subject, did not significantly improve as a result of the change in assessment format. Other measures of performance such as changes in student numbers in different grade categories did not reveal changes at any statistically significant level. The findings of this study are contrary to the existing literature that changes in assessment improve grade performance in a subject. The implication of the findings are that changes in assessment need to be weighed against the benefits from it, and that traditional paper-based assessment still works as well, if not better than technology driven assessment such as online open book examinations.

Keywords: *motivation, learning, engagement, continuous assessment, summative assessment, online quiz*

INTRODUCTION

Assessment in higher education is a means to measure students' performances and teaching outcomes. Summative assessment, the most common form of assessment, is mainly used to award students' scores on attempts in examination content, and usually conducted in a supervised time constrained assessment environment (G. A. Brown, Bull, & Pendlebury, 2013; Hernández, 2012; Kuh, 2003). A single end of session assessment or one mid-term and a final exam is commonly used as summative assessments in any subject. Empirical studies report improvements in teaching and learning outcomes when a summative assessment is changed in supervised examination conditions (Greer, 2001; Hernández, 2012; Marriott & Lau, 2008). Very little is known about an assessment in an online, unsupervised examination environment in an Australian higher educational setting.

The current quasi-experimental study explores the impact of a change from a paper-based supervised summative mid-term examination to three short online unsupervised summative quizzes on students' final examination marks and overall marks in a management accounting subject. Using data from two years, before a change and after a change in a mid-term examination format, the study reports a number of important insights contrary to extant literature. The statistical analyses report that after the change in the assessment format, the marks in group essay and the average quiz marks declined slightly in 2014. The findings refute the

conventional wisdom that changes in an assessment improve students' performance (measured by marks or scores in a subject). The study has implications for academics contemplating moves to continuous and unsupervised online examinations.

The rest of the paper is organized as follows: the literature is reviewed in section two followed by the background of the study. The research method and the results are then discussed followed by the conclusions.

LITERATURE REVIEW

Assessment is a way of evaluating students' work, making inferences about the work and estimating the worth of students' actions (G. A. Brown, et al., 2013). Hernandez (2012) adds that assessment is about grading and reporting student achievements and about supporting students in their learning. Brown (2004) recommends the use of a variety of assessment practices so that students can demonstrate their abilities and optimize their potential. Rust (2001) argues that assessment practices should be dynamic enough to have a beneficial effect on students' learning. Empirical studies report a beneficial effect of learning from assessment and improvements in grades from changes in assessment practices (Greer, 2001).

In the higher-education context, two types of assessment items are widely used: summative and formative. Hernandez (2012) labels summative assessment as 'assessment of learning' and formative assessment as 'assessment for learning'. Summative assessment is compulsory in nature, usually completed as a single submission of work and provides little opportunities to students to reflect on how they are progressing (Marriott & Lau, 2008). It is also used as a performance indicator (Knight, 2001) to gauge learner's achievements against predetermined grading criteria. These strengths can be regarded as failings of summative assessment to take a holistic view of learning of subject content and focus on rewarding (related to assessable work) aspects of learning (Marriott & Lau, 2008).

Formative assessment, on the other, hand is an optional assessment which does not contribute to final outcomes (Aisbitt & Sangster, 2005; Marriott & Lau, 2008). It is seen as a lifeblood of learning (Rowntree, 1987) and expected to provide feedback to students to improve, accelerate and enhance learning (Sadler, 1989). The success of formative assessment in terms of increased grades is largely unproven as the majority of the studies shows mixed results (Aisbitt & Sangster, 2005; Sangster, 1996).

Both formative and summative assessments have merits and demerits, so combining the best aspects of these two types of assessment may be appropriate as interventions. Empirical research has revealed successful combination of both types of assessment in a subject (Lewis & Sewell, 2008; Trotter, 2006). Marriott and Lau (2008) argue that summative assessment can be a single piece of assessment or a series of assessments delivered throughout a teaching period which could take the form of essays, tests and presentations (see also Purvis, 1990). When continuous assessment is used, it is aimed to monitor students' performance and provide timely feedback that may be used to improve future performance (Marriott & Lau, 2008). The use of continuous summative assessment throughout the teaching period can be perceived to have formative and summative function in that performance in one test can feed forward to the next, thus possessing the attribute of a formative assessment.

While formative and summative assessment practices have taken different forms, there is an increasing level of use of technology to enhance both types of assessment tasks. Computer-aided assessment (CAA) is one form of assessment technique used in both summative and formative assessment (Bull & McKenna, 2003). It is regarded as an efficient assessment option

(Marriott & Lau, 2008) because this form of assessment does not place excessive burden on staff and students (Light, Calkins, & Cox, 2009). CAA can provide timely feedback to students which can enable students to identify their weaknesses, reflect on their performance and improve their study skills (Aisbitt & Sangster, 2005; Lewis & Sewell, 2008). CAA also offers options for “sustainable assessment” which encompasses knowledge, skills and predispositions required for lifelong learning activities (Boud, 2000, p. 151).

Students are observed to be motivated by extrinsic rewards (e.g. good job, good career) (Ottewill & Macfarlane, 2003) and intrinsic rewards (Biggs, 2011). Assessment is the only way to encourage students to learn (Race, 1995; Rowntree, 1987) though some students may be distracted and enervated by assessment tasks (Rowntree, 1987). Teaching curriculum plays an important role in students’ learning (S. Brown, 2004). So a curriculum that is assessment driven is more useful for students’ learning (Carless, 2007; Joughin, 2009). While learning is desired in any assessment, students must be engaged in learning activities by institutional initiatives (Kuh, 2001a, 2001b) or by coercive practices to ensure learning such as frequent assessment and feedback (Kuh, 2003; Oliver, 1998). Feedback facilitates learning (Gibbs & Simpson, 2004), allows autonomy and responsibility to monitor and manage students’ own learning (S. Brown, 1999; Carless, 2007).

RESEARCH BACKGROUND

This quasi-experiment is based on data collected from a regional university in Australia and involves third-year majors in Accounting and Finance studying a three-credit compulsory Management Accounting subject. The subject is also accepted and taken by students as an elective in other courses offered by the university. An undergraduate program comprises of 72 credits for subjects studied over a three-year period, each year offering eight 3-credit subjects. The subject is delivered over a 13-week period with a break between weeks, once per year. The students attend a two-hour lecture, one-hour tutorial (with a class size of 25 students) and a one-hour workshop. The sessions are optional but students are strongly encouraged to attend as many sessions as they can. Recently the university commenced online studies and is taught online to external students, nationally and internationally.

The assessment items are different in terms of weights and structure between the years. In 2013, the final examination and the mid-term examination were invigilated. The weights of three assessment items were: essay (20%), mid-term examination (20%) and final examination (60%). Achieving an overall pass required only 40 out of 80 marks in invigilated components (50%) and another 10 marks (50%) in the essay. So achieving a pass grade was not too challenging to the students who did not perform well in the mid-term exam still could improve in the final examination.

In 2014, the assessment structure was changed. The paper-based mid-term test in 2013 was changed to three open book online quiz examinations each worth 5%. The two other assessment items remained the same but the weights were changed as essay carrying 20% weight and the final exam carrying 65% weight. The only invigilated component was the final examination and to pass the subject overall, the students were required to get 32.50 marks (50%) of the total (65) and another 17.50 marks from the non-invigilated components (35 marks in essay and three quizzes). In order to make the comparisons meaningful, the assessment items are converted to 2013 weights.

Assessment revision is a common practice in higher education and is motivated by the findings in the literature that a revision in assessment improves learning, engagement and

enthusiasm in students (see, for example, Greer, 2001; Marriott & Lau, 2008). Accordingly, the assessment in the subject was revised in 2014. The paper-based mid-term exam, covering 50% of the total topics, covering lectures one to six, was replaced with three open book, and phased-in online multiple-choice tests worth 5% each, covering only two lectures (two chapters) at a time. The tests were to be completed online (unsupervised), comprised 10 multiple-choice questions to be completed within 30 minutes. Though students were required to attempt only 10 questions, 20-25 questions were kept in two separate pools, a theory pool and a computational question pool, so that each student got different sets of theory and computational questions. In order to prevent plagiarism and collusion, answer choices were also randomized so that each student would see answer choices in different order. As a further precaution, alphabets for answer choices (a, b, c, and d) were removed and replaced with a tick box. The scores of the quizzes were made available through the online testing portal immediately after the submission of the tests by each student. The students were allowed to check their answers against the actual answers after everyone completed their tests.

DATA AND PROCEDURE

Data for the study was collected from the university's central database for the years 2013 and 2014 after receiving ethics approval from the university's Ethics Office. In 2014, 46 students were enrolled, an increase of four (4) students from 42 students enrolled in 2013. All students attempted all assessment tasks in both years, so the data comprises of assessment marks of 100% students on record when the results were finalized at the end of semester two each year (that is, December).

The marks of the students are analyzed in SPSS, Eviews and in Excel. A number of statistical tests are used to analyze the collected data. To learn about the characteristics of the marks in different assessment items, descriptive statistics of all assessment items are analyzed. To ascertain the relations between different assessments items, Pearson bi-variate correlations are used. To determine the effect of intervention on students' marks, an independent sample t-test procedure is used to compare and contrast the marks in different assessment items during 2013 and 2014. Finally, to determine the overall achievements from the intervention, test of proportions of different grades during 2013 and 2014 is used. The analysis and the discussions of the statistical tests are presented next.

RESULTS AND DISCUSSIONS

A number of statistical tests are used to analyze the results of different assessment items during the years 2013-2014. The descriptive statistics below summarizes the marks in different assessment items, after the adjustments, in 2013 and 2014.

	Final exam (60%)		Essay (20%)		Mid-term (20%)	
	2013	2014	2013	2014	2013	2014*
Mean	34.29	35.43	15.51	15.15	11.04	10.83
Median	36.00	35.77	15.63	15.60	11.25	11.00
Maximum	49.00	54.00	17.63	19.55	15.00	14.50
Minimum	0.00	17.08	13.00	0.00	6.00	6.00
Std. deviation	11.07	7.62	0.89	3.12	2.06	2.27
Probability	0.00	0.78	0.00	0.00	0.52	0.49
Observations	N =42	N=46	N =42	N = 46	N = 42	N =46

*adjusted to reflect the total of 2013 mid-term marks

Table 1 above shows the marks distribution of the students enrolled in both years. The table shows that the average marks in the assessment items in 2013 were better than 2014 assessment items. Though the final marks in 2014 were slightly better in absolute terms, the median marks in 2014 (35.77) was inferior to 2013 marks (36.00). The descriptive statistics alone is not sufficiently informative to determine if an intervention in the form an assessment change was effective. In 2014, the change to open-book online examinations is the intervention used to improve the students' engagement with the learning activities and improve the acquisition of skills throughout the semester and be able to keep these skills for the future such as for their jobs and for management accounting subjects taken at professional levels. The correlations table below summarizes the results of causality between different assessment items during 2013 and 2014.

		Essay	Mid-term
Mid term	Correlation	.365*	
	Sig.	0.017	
Final	Correlation	.373*	0.235
	Sig.	0.015	0.134

* Correlation is significant at the 0.05 level (2-tailed).

*5% level of statistical significance

Table 2 above reports the correlations between all assessment items in 2013. A significant correlation between the essay and the final examination marks is observed. The correlation between the essay marks and the mid-term marks is seen as a surprise as the structures of these assessment items are different. The mid-term examination was a closed book multiple-choice exam, and the essay was a take-home group assessment task aimed to improve students' literacy skills. However, the results in Table 2 shows no significant correlation between the mid-term and the final examination marks. Though both assessment pieces required students to solve problems in supervised exam conditions, no apparent connection in the form of a significant correlation is observed between these two assessment pieces. Quite surprisingly a significant correlation is observed between the essay and the final examination marks. From the analysis of the content of the final examination of 2013, it can be inferred that the students' literacy helped the students in the final examination. The final examination in 2013 was quite challenging in that the questions were lengthy, verbose and required significant level of

comprehension skills in a timed test environment. Failure to comprehend the questions and operationalize the variables before solving the problems could have been quite disastrous to students' performance. The concerns over the students' engagement in learning activities in the subject required a reappraisal of the assessment structure in 2014. The mid-term was replaced in 2014 with three online open book take-home examinations, which followed relevant online mock practice exams from each exam topic. The objective was primarily to engage the students to continuous learning of subject materials throughout the semester and also to maintain the rigor of teaching and learning in the subject. An analysis of the correlations between different assessment items of 2014 is reported in the table below.

		Adjusted final	Quiz01	Quiz02	Quiz03	Essay
Quiz01	Correlation	0.243				
	Sig.	0.104				
Quiz02	Correlation	0.234	0.207			
	Sig.	0.118	0.167			
Quiz03	Correlation	0.222	.416**	.489**		
	Sig.	0.138	0.004	0.001		
Essay_2014	Correlation	.368*	0.146	0.217	0.196	
	Sig.	0.012	0.333	0.148	0.192	
Average quiz marks	Correlation	.307*	.724**	.741**	.821**	0.244
	Sig.	0.038	0.000	0.000	0.000	0.102
*Correlation is significant at the 0.05 level (2-tailed).						
**Correlation is significant at the 0.01 level (2-tailed).						

Table 3 above reports the presence of significant correlations between the essay and the final examination marks, as observed in 2013 as well. The online quizzes are not significantly correlated to the final examination marks, but average quiz marks (aggregated) and the final examination marks are significantly correlated. Significant correlations are also observed between quiz one and three, and quiz two and three, which suggests that students who performed well in quiz one and two also performed well in quiz three. In Table 3 above, the other assessment piece (the essay) is kept as a control variable so that the impact of the intervention can be observed in terms of improvements in marks in the final exam. To determine the impact of the interventions, a comparison of marks between different assessment items is reported in Table 4 below.

	2013	2014	t- value	probability
Overall grades				
Final exam	34.29 (11.07)	35.43 (7.62)	0.556	0.573
Class test/average quiz	11.04 (2.06)	10.83 (2.27)	-0.441	0.661
Group Essay	15.51 (0.89)	15.15 (3.12)	-0.767	0.447
Overall marks in the subject	63.80 (10.77)	65.02 (12.29)	0.491	0.624

In the Table 4 above, the results report no apparent significant differences in assessment marks during the years 2013 and 2014. The results above show that the final examination marks in 2014 improved slightly (1.13 in absolute terms or 1.89%) over 2013 marks ($t= 0.553$ and $p= 0.709$). Marks in the other two assessment pieces declined in 2014, but the decline was not statistically significant. The decline in mid-term marks, from 11.04 to 10.83 was not statistically significant ($t= -0.433$, $p=0.3329$). The standard deviation of marks in 2014 is larger while the average score is lower than the marks in 2013 (average = 11.04, standard deviation = 2.06). The variability in marks in 2014 may be due to the level of difficulty invoked to control for cheating in take-home quiz examinations. The essay marks reveal a similar story of decline from 2013 marks, from an average of 15.51 (standard deviation of 0.89) to 15.15 (standard deviation of 3.12) in 2014. The decline in marks, however, is not significant at any statistical level ($t=-0.751$, $p= 0.2297$). The variability of essay marks in 2014 is quite noticeable. One of the possible causes may be the efforts required in the subject, that is, in 2014 there were at least six (6) practice tests and another three (3) graded quizzes which collectively required students to acquire critical reasoning and problem-solving skills.

A departure from these nine piecemeal assessments, mostly problem solving in nature, to a different format of assessment requiring literacy skill, from week 7 to the submission of the essay in week 10, may have added some challenges to the majority of the students doing the subject. This challenge was, apparently, not well handled by the students as evidenced by the decline in the average marks and the increase in standard deviation of the marks in the essay in 2014. Even though there were changes in marks in different assessment items, the ultimate goal was to achieve a better outcome, in terms of intrinsic gain, that is, the acquisition of problem-solving skills and information literacy, and extrinsic rewards, that is, an improvement in grades and overall pass rates in the subject. A comparison of the number of students in different grade categories below sheds some lights on this issue.

Table5
TEST OF PROPORTIONS (Z-TEST) OF STUDENTS' COMPARATIVE PERFORMANCE

Letter grade	2013		2014		Z-value	Probability
Fail	4	9.50%	2	4.30%	0.9622	0.3371
Pass	11	26.20%	19	41.30%	-3.389	0.0007**
Credit	19	45.20%	17	37.00%	0.7892	0.4295
Distinction	8	19.00%	5	10.90%	1.08	0.2801
High Distinction	0	0.00%	3	6.50%	-1.68	0.092*
	42	100.00%	46	100.00%		
** significant at 5% level						
*significant at 10% level						

Table 5 above reports the results of test of proportions of two independent samples (2013 and 2014). The results suggest that there is a difference in proportions of students receiving the Pass and High Distinction (HD) grades between 2013 and 2014, and the differences are statistically significant: the Pass grade recipients are significant at 1% level ($Z = -3.389$, $p = 0.0007$) and the High Distinction grade recipients are significant at 10% level ($Z = -1.68$, $p = 0.092$). The rises in these two categories of student numbers and proportions were counterbalanced by a decline in Credit and Distinction category student numbers. The decline, however, is not different at any statistically significant level. Finally, the overall failures in 2013 (9.50%) and in 2014 (4.30%) were not statistically different ($Z = 0.9622$, $p= 0.3371$). The results

above suggest that overall, the outcomes, in terms of improvements in grade categories, were achieved from the intervention. A slight decline in marks, however, may be due to cohort issue or the number of assessment items in the subject (5 in 2014 compared to 3 in 2013).

CONCLUSIONS

The study seeks to understand the effect of an intervention, in the form of a change in an assessment item, on students' learning habits and learning outcomes of a third-year management accounting subject taught at a regional university in Australia. Two objectives are examined in this paper. The first objective seeks to understand the effect of an intervention (in the form of a change from a paper-based mid-term examination to three online open-book short quizzes) on students' intrinsic learning developments, that is, a change in students' learning habits throughout the semester so that students' attention is moved away from an exam-centered learning to continuous learning and improvement. The finding is that students' study habits changed as a result of the intervention assessment (online open-book quiz) in 2014. Significant correlations between different assessment items, the essay, three quizzes and the final examination, all significant at 5% level, suggest that the students were more involved in learning activities than before. The increased demands to complete more formative and summative assessment items may have driven the learning habits of the students throughout the semester.

The second objective seeks to understand the effect of the intervention on students' extrinsic rewards from the continuous learning activities, that is, a change in grades and overall pass rate in the subject. The findings are that the changes in learning habits affected the average marks in the final examination of 2014 but the marks in other assessment items, that is, average mid-term and the essay, declined in 2014 over 2013 marks. Though the declines are not statistically significant, it suggests that the students were overwhelmed with more assessments in 2014 over the number of assessments in 2013. The results also suggest an improvement of grade distribution over 2013, there were more students in Pass and High Distinction categories in 2014, and the differences were statistically significant within 10% level. Thus, the findings partially support the prior literature on assessment intervention benefits (see for example, Aisbitt & Sangster, 2005; Greer, 2001; Hernández, 2012; Marriott & Lau, 2008). From the instructor's point of view, the reduction in failures from 9.50% to 4.30% was noteworthy though the decline was not statistically significant at any level of confidence.

The study has obvious limitations of any study. Only one subject is examined over two years. The findings can be validated by repeating this study in other subjects with similar subject content, in other assessment formats and in other academic institutions. Other confounding influences such as students' perception about open book examinations, commitments of time and preparation for the online quizzes and time allocation to other subjects based on the belief that open book exams need less time, may have affected the results of this study. The inclusion of these variables in future studies may be worthwhile. Only two years of data is used to report the findings from the intervention. Future research may look into time series data for patterns of effects from an intervention reported in this paper.

ACKNOWLEDGEMENTS

The paper is presented in the Summer Internet Conference, 2015. The research is an ethics approved project (Approval #H6382) of the institution where the research is carried out.

The suggestions of the reviewers for improvement to the earlier version of this paper are gratefully acknowledged.

REFERENCES

- Aisbitt, S., & Sangster, A. (2005). Using internet-based on-line assessment: A case study. *Accounting Education: an international journal*, 14(4), 383-394.
- Biggs, J. B. (2011). *Teaching for quality learning at university: What the student does*: McGraw-Hill Education (UK).
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2), 151-167.
- Brown, G. A., Bull, J., & Pendlebury, M. (2013). *Assessing student learning in higher education*: Routledge.
- Brown, S. (1999). Institutional strategies for assessment. *Assessment matters in higher education: Choosing and using diverse approaches*, 3-13.
- Brown, S. (2004). Assessment for learning. *Learning and teaching in higher education*, 1(1), 81-89.
- Bull, J., & McKenna, C. (2003). *A blueprint for computer-assisted assessment*: Routledge.
- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and teaching in higher education*, 1(1), 3-31.
- Greer, L. (2001). Does changing the method of assessment of a module improve the performance of a student? *Assessment & Evaluation in Higher Education*, 26(2), 127-138.
- Hernández, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education*, 64(4), 489-502.
- Joughin, G. (2009). *Assessment, learning and judgement in higher education*: Springer.
- Knight, P. (2001). *A briefing on key concepts: Formative and summative, criterion and norm-referenced assessment*: Learning and Teaching Support Network.
- Kuh, G. D. (2001a). The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. *Bloomington, IN: Indiana University Center for Postsecondary Research*, 1-26.
- Kuh, G. D. (2001b). Assessing what really matters to student learning inside the national survey of student engagement. *Change: The Magazine of Higher Learning*, 33(3), 10-17.
- Kuh, G. D. (2003). What we're learning about student engagement from NSSE: Benchmarks for effective educational practices. *Change: The Magazine of Higher Learning*, 35(2), 24-32.
- Lewis, D. J. A., & Sewell, R. E. (2008). Instructional design and assessment: Providing Formative Feedback From a Summative Computer-aided Assessment. *American Journal of Pharmaceutical Education*, 71(2), Article ID 33.
- Light, G., Calkins, S., & Cox, R. (2009). *Learning and teaching in higher education: The reflective professional*: Sage.
- Marriott, P., & Lau, A. (2008). The use of on-line summative assessment in an undergraduate financial accounting course. *Journal of Accounting Education*, 26(2), 73-90.
- Oliver, R. (1998). Experiences of assessing programming assignments by computer. *Computer based assessment*, 2.
- Ottewill, R., & Macfarlane, B. (2003). Pedagogic challenges facing business and management educators: Assessing the evidence. *The International Journal of Management Education*, 3(3), 33-41.
- Purvis, R. F. (1990). *Continuous assessment*: Scottish Central Institutions Committee for Educational development.
- Race, P. (1995). The art of assessing 1. *New Academic*(Autumn), 3-6.
- Rowntree, D. (1987). *Assessing students: How shall we know them?* : Taylor & Francis.
- Rust, C. (2001). *A briefing on assessment of large groups*: LTSN York.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2), 119-144.
- Sangster, A. (1996). Objective tests, learning to learn and learning styles. *Accounting Education*, 5(2), 131-146.
- Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education*, 31(5), 505-521.