Loughborough University

This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Implications of patterns of use of freely-available online formative tests for online summative tasks

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Loughborough University

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Meyer, Jan, Mel Ziman, Sue Fyfe, Georgina Fyfe, Kayty Plastow, Kathy Sanders, and Julie Hill. 2019. “Implications of Patterns of Use of Freely-available Online Formative Tests for Online Summative Tasks”. figshare. https://hdl.handle.net/2134/4597.

Loughborough University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

creative commons

COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor.



Noncommercial. You may not use this work for commercial purposes.



No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

Disclaimer

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

IMPLICATIONS OF PATTERNS OF USE OF FREELY- AVAILABLE ONLINE FORMATIVE TESTS FOR ONLINE SUMMATIVE TASKS

Jan Meyer, Mel Ziman, Sue Fyfe, Georgina Fyfe, Kayty Plastow, Kathy Sanders and Julie Hill

Implications of Patterns of use of Freely-Available Online Formative Tests for Online Summative Tasks

Dr Jan Meyer[1], Dr Mel Ziman[2], Prof Sue Fyfe[3], Ms Georgina Fyfe[3], Ms Kayty Plastow[1], Dr Kathy Sanders[1], Ms Julie Hill[1]

[1] The University of Western Australia
Crawley, Western Australia, 6009
[2] Edith Cowan University
[3] Curtin University of Technology

jmeyer@anhb.uwa.edu.au
m.ziman@ecu.edu.au
S.Fyfe@curtin.edu.au
G.M.Fyfe@curtin.edu.au
kjplastow@anhb.uwa.edu.au
ksanders@anhb.uwa.edu.au
jhill@anhb.uwa.edu.au

Abstract

The use of online assessment tasks in a summative context can create tensions between the institution's need for security to ensure the validity of individual evaluations and the student's need for flexibility of access. This is especially the case in recent years, with the upsurge of students engaged in paid employment while enrolled in full-time study. The lowest rate of engagement of students in paid employment at the three institutions in which our study was based was 65%, the highest 75%. One quarter of all students at this institution spent more than 20 hours a week in paid employment. Ninety seven percent of students in paid work were enrolled on a full-time basis.

This study determined from automatically recorded times of logon, individual question submission and whole test submission the patterns of use of online feedback-enriched MCQ tests by 656 students across the three institutions in Perth, Western Australia. The conditions under which the tests were available to students varied from a strictly secured, summative task available for a limited time on campus within hours governed by the accessibility of automatically locked-down computer rooms and the availability of staff for live or video invigilation to a freely accessible formative learning exercise.

Mismatches between preferred and available times severe enough to exclude some external students from assessment were identified. Evidence was found that for younger (16-18 year old) students especially, meaningful engagement with test-structured tasks lasts no more than 10 minutes, one third of the designed time of our current summative online tests. The one third, approximately, of enrolled students who did not use the online test facility had significantly poorer academic outcomes. The advantage granted by test use increased substantially with repetition.

The question of how to ensure the security and validity of online testing while increasing real flexibility of access remains unresolved for us. We accept the social responsibility of finding a solution.

Introduction

Barbara Stewart (2004) argues persuasively the potential benefits of online learning and assessment for meeting the needs of underserved populations, such as those with physical handicaps, variant personal cognitive and psychological orientations, who are subject to geographic and cultural separation, and operating under gender and occupational constraints. Curtis and Shani (2002) reported an increase in the proportion of students in paid employment from a single department in a British University from 43% to 55% between 2000 and 2001. Levels of participation in paid employment by first year students at Australian universities have increased from 51.3% in 1999 to 54.9% in 2004 (Krause et al., 2005), at the same time as the proportion of students with language backgrounds other than English has risen.

Stewart goes further than pointing out the benefits of online course material, arguing the social obligation to provide flexible access to learner-centred and assessment-centred learning environments. The use of online assessment tasks in a summative context can create tensions between the institution's need for security to ensure the validity of individual evaluations and students' needs for flexibility of access, however. This study examines patterns of use by students of freely-available formative and constrained summative online tests in an attempt to gain some insight into the magnitude of the mismatch between students' needs and preferences and the current manner of presentation of summative online testing in one area of scholarship at three universities in Perth, Western Australia.

Materials and Methods

The data analysed in this study arose from the development and evaluation phases of a trial of automated explanatory feedback comments for single topic online MCQ tests in first year Human Biology units at three West Australian Universities. Approximately 2,000 students in all enrol in these units each year. Demographic information was gathered from 1099 of these students in a survey of perceptions of the adequacy, use of and need for various types of feedback administered at the outset of the courses. The patterns of use were determined from automatically recorded times of logon, individual question

submission and whole test submission for 656 students across the three institutions. Additional demographic information together with self-reports of test scores was gathered from an online survey of 315 students at the end of semester. Final grades and the contribution of different sections of the courses to those grades were taken from unit exam databases.

The project from which this study arose involved the fitting of feedback comments to online test systems already in use at each institution. Thus, while it was not possible to implement a balanced experimental design, the different situations in which the tests were used and the different characteristics of the platforms through which each was presented did present opportunities for gathering information in a form best likened to that arising from a hybrid cross-sectional/ longitudinal study. For example, it was possible to link self-reported test marks with actual test marks at one institution, expected scores with actual scores at another and the time taken over each question with sectional course marks at the third.

At the first institution the online multiple-choice style tests with feedback were only presented summatively, contributing 24% to the final mark for the course (6% per topic test). Students had 40 minutes to complete the 30 item test at a pre-booked time between 9am and 5pm in a secure, invigilated central computing facility. At the second university the test was presented only as a formative learning task, freely available for 24 hours a day for one month. This test comprised 50 unvarying questions. At the third university the test was presented as a freely available formative exercise for 24 hours a day for one week before items from the same test bank were presented in a summative task available from 9am to 5pm under video surveillance in a departmental computing laboratory. The 30 questions in this test were selected randomly from sets of between 5 and 15 alternatives. The hours of availability of the summative tasks were in both cases governed by the accessibility of automatically locked-down computer rooms and the availability of staff for live or video invigilation.

The differences in final course grades of students taking complete, incomplete and no online tests were evaluated using a 1 way ANOVA, with institutions as blocks. At most 6% the final grades used to determine the type of student making use of the online tests was determined by scores obtained in the test (and that for 23% of students). The advantages given by involvement with online testing were assessed by comparing percentage multiple choice question scores in the topic areas dealt with by the feedback-enriched tests in final course examinations in 1-way ANOVAS, with scores in other topic areas as covariates. All analyses were carried out using GenStat ninth edition (2006) and graphs prepared through GenStat and Excel.

Results

The lowest rate of engagement of students in paid employment at the three institutions was 65%, the highest 75% at the institution offering the test as a formative learning exercise only. One quarter of all students at this institution

spent more than 20 hours a week (maximum 70hrs) in paid employment, largely in paramedical areas. Ninety seven percent of students in paid work were enrolled on a full-time basis. Twenty two percent of students spoke one of 42 languages other than English at home, nearly half exclusively so.

Rates of use of the online feedback-enriched tests were significantly lower in the institutions where they were not obligatory (63.4% of enrolment compared with 71.9%, χ^2=8.44, 1df, p=.006) and in the institution where available for a week (60.3%) compared with a month (67.5%, χ^2=4.50, 1df, p=.034). The students who made use of the online tests were the higher achievers, whether or not those taking the test as an obligatory summative task were taken into account (Figure 1)((1-way ANOVA, with Institutions as blocking term F = 60.39, 1& 1167df, p<.001, each level differing significantly from the others at p <.05 by LSD).

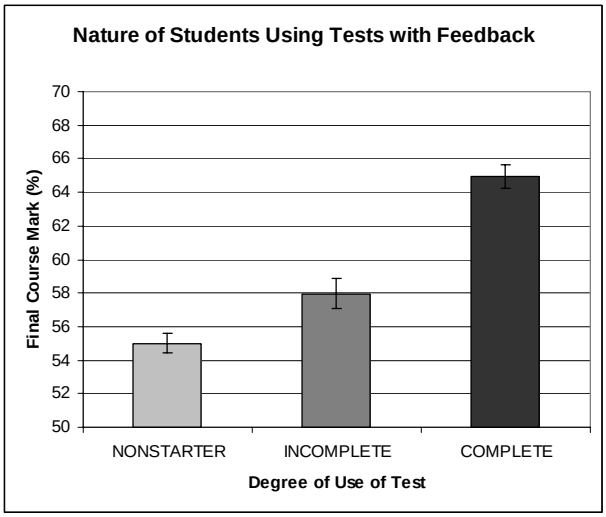


Figure 1. The final course grades achieved by students according to the level of use of the online feedback-enriched tests. Marks from the online tests in question contributed at most 6% to the final course grades of fewer than one quarter of the students. (mean ± SE)

In the institution in which the test was available as a formative review before summative testing use peaked on the days immediately preceding the summative tests. The students who used the test as a review in the week before the summative test gained more advantage from its use than those who had used it earlier, or not at all (Figure 2) (ANOVA F= 4.82, 1& 362 df, p= .003).

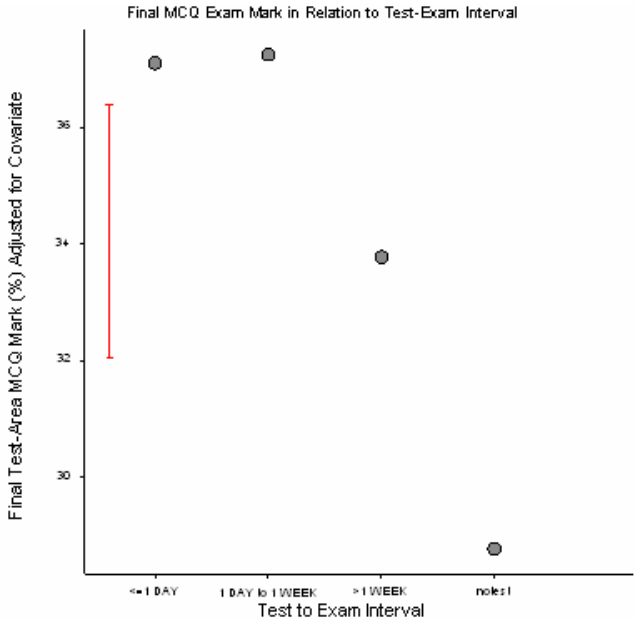


Figure 2. The advantage gained from use of feedback-enriched tests in relation to the interval between test use and exam at institution using tests in both formative and summative tasks.(MCQ means in test topic area adjusted for scores on non-test topics as covariate, bar = SE of difference)

The average time taken by students to read the MCQ stem and the five answer options, decide upon a response, review their grade and read the one feedback comment they received was just under 45 seconds. Only 20% of students spent more than a minute on each question, 56% took between 30 seconds and a minute and 24% less than 30 seconds.

It was only possible to analyse the pattern of interaction of students with the test over time in the institution in which it was offered both formatively and summatively. Only one third of the approximately 60% (281) of the class who logged on to the test at least once, completed it once, straightforwardly from beginning to end. Nearly 40% of those logging on never completed a test, 4% did a little more than one complete test and 9% repeated the whole test (one student 27 times). Students who repeated the test showed significantly greater advantage in the final MCQ exam in the topic covered by the online test in relation to other topics than other students (Figure 3) (ANOVA F= 4.89, 7&459df, p <.001).

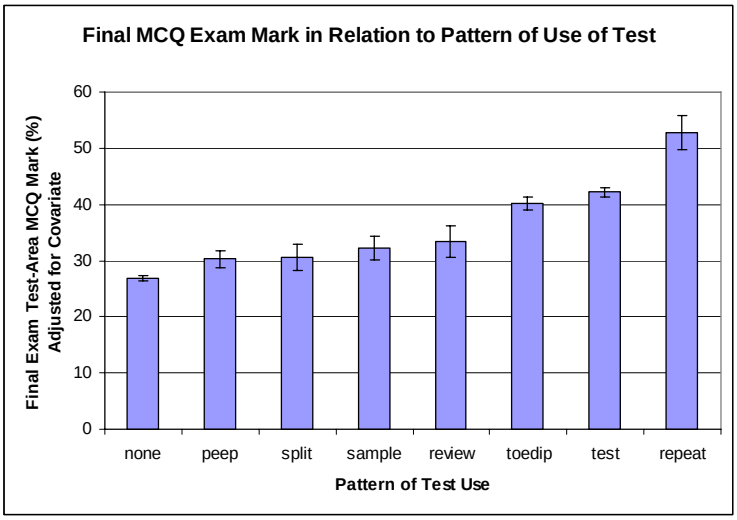


Figure 3. The advantage gained from use of feedback-enriched tests in relation to the degree of engagement with the tests at institution using tests in both formative and summative tasks. None = did not take test; peep & sample = incomplete tests; test, review & toedip = between 1 & 2 tests completed; repeat = 2-27 tests completed. (MCQ means ± SEs in test topic area adjusted for scores on non-test topics as covariate)

The remaining 14% remained logged on for the whole of the interval between the posting of the test and the final examinations, completing a few more questions every few hours or days until they had completed the whole test. There were many comments made in the follow-up evaluation survey to the effect that a 30 question test was too long, for example

> "Also this was a very long test for a computer test after a while at staring at the computer you start to lose concentration."
>
> "Big long sentences get too frustrating to read through and understand so often just picked a letter to save time"

Consideration of the rate of fall-off in discriminatory answering of items on the initial questionnaire survey also indicated that younger (16-18 year old) students in particular experienced difficulty in maintaining concentration on a task for more than 7-10 minutes (Figure 4).

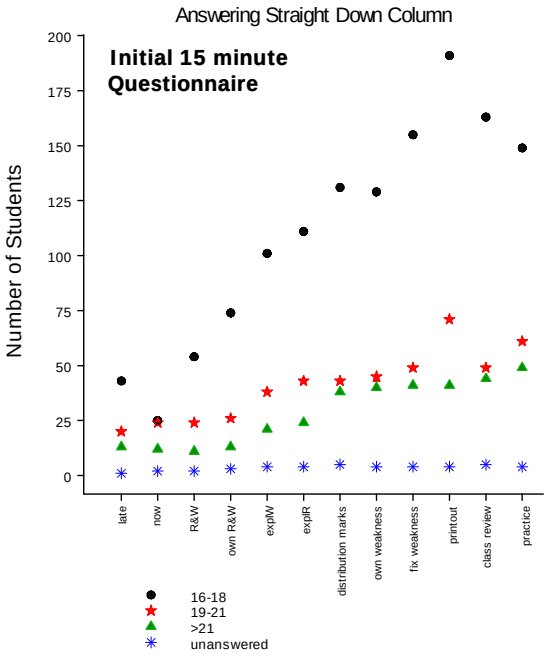


Figure 4. The fall-off meaningful responding across a 15 minute questionnaire, defined by the selection of the same response from the top to the bottom of the column of questions. The first column on the questionnaire is indicated on the far left and the last on the far right. The age groups indicated are 16-18 year olds (circles), 19-21 year olds (stars) and over 21 year olds (triangles).

The only students who did spend more than the intended minute on each test question were those who split the 30 question test across several sessions (Figure 5).

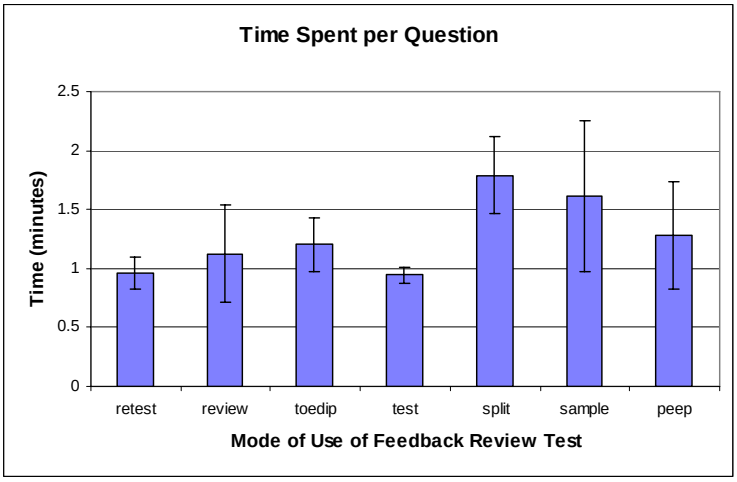


Figure 5. The average time (mean ± SE) spent on each test question in relation to the degree of engagement with the tests at institution using tests in both formative and summative tasks. Peep & sample = incomplete tests; test, review & toedip = between 1 & 2 tests completed; retest = 2-27 tests completed.

Their extra efforts did not yield any obvious performance advantage in the final MCQ exam in the topic covered by the feedback-enriched online test (Figure 3), nor could the extra time they spent on each question be accounted for by language background.

At the two institutions where the test was freely available approximately 40% of students logged on to use it outside of the hours of 9am to 5pm. That the limited times of access to summative testing had an impact on student performance was indicated by the behaviour of a group of externally-enrolled students of the institution offering only the compulsory summative task who, being residents of the metropolitan area, were required to attend on-campus for testing. Despite above-average grades on other aspects of the course, not one of these students presented themselves, and thereby forfeited almost one quarter of their course marks

Discussion

Our study indicates that there is no sign yet of students turning away from paid employment back to full-time engagement in their university studies. Levels of paid employment in our cohort were 10% to 20% above those described by Krause (2005) only two years ago, with similar or higher proportions working outside the university for more than 20 hours each week. The need for the flexible delivery of course materials and assessment which can be provided online cannot be said to have diminished.

It was interesting to find quantitative confirmation of Charlesworth and Vician's (2003) anecdotal observation that, when left to use online tests as they wish,

students like to be able to take breaks. This behaviour, comments provided in an online survey and the failure of younger students to persevere with completion of a paper-based questionnaire provided multiple lines of evidence for the need not only of flexibility of time constraints in online testing, but for the restructuring of tests to maintain engagement, and for exercises to develop the stamina and concentration of our younger students. Since the completion of this study we have restructured the feedback-enriched online tests presented at the institution with 75% of its first year enrolment between 16 and 18 years of age so that only 10 questions are presented in each test. Since these 10 questions are drawn randomly from the same database as served the larger tests, the number of different tests available to a student making multiple attempts has risen considerably, and the rate of repetition of testing risen. The decay over a few weeks of the advantage gained by using the online tests we revealed points to the need for repeated access to online tests for consolidation of learning and the superiority of the effect gained by repetition to its effectiveness. It was interesting to note that a number of students returned to the online tests after their final examinations. While acknowledging the advantages of this approach, we still plan to trial the gradual increase in the number of questions in tests made available across the semester. We have some evidence, in the form of comments such as

> "could have figured out the answers by reading the question again and thinking about how it was worded"

that repeated exposure to explanations of right and wrong answers may be encouraging more than a cursory reading of questions, but have yet to see any evidence that this translates into improved long-term learning.

We found considerable evidence that we are not realizing the potential for flexible course delivery offered by the online learning environment. Like Volery and Lord (2000) we found that when left to their own devices students take full advantage of the flexibility offered by online course activities and log on at all hours of the day and night. We do not offer them the opportunity to take summative online assessments in any location but secure labs on campus, and at any times other than regular 'business hours', however. That the mismatch between preferred and imposed times of access has an impact on students ability to complete our courses successfully was indicated by the failure of a group of high quality externally-enrolled students to access significant components of their summative assessment in the course at all. Their pattern of enrolment is most commonly encountered amongst students in full-time work, but is also employed by women with heavy family commitments. As Stewart (2004) says, it is a social responsibility of education to provide, as far as possible a 'level playing field' with respect to access for students with diverse needs.

The difficulty in realizing the ideal set out by Stewart lies with the issue of security and validity of assessment. Rowe (2004) argues that accurate assessment, including online assessment, is essential to the survival of educational institutions, for it validates student knowledge as certified by degrees and diplomas, that if an institution claims to provide this service, they

must prove to society that they do. Noting that draconian measures to reduce cheating diminish trust, and that people who feel more “distant” cheat, Rowe clearly places us in the position of finding a way of minimizing the impact of teaching while increasing the accommodation of student needs. The chief issue when questions and answers are available before the summative test becomes that of impersonation of the student who is not present and under direct observation. Possible approaches we have considered to the issue include extending the physical access to and supervision of secure computer labs to 24 hours a day (a costly option), decreasing the value of each summative test to the point where it is not worth cheating (but this does disadvantage students with special needs), instigating sectional pass requirements so that cheating on online tests cannot lead to a pass in the unit (but this raises moral dilemmas and invites student appeals against assessment) and finding a computer-based means of verification of identity (does anyone know of one?).

We shared with Morris et al (2005) the experience of having approximately one third of our students enrolled throughout the semester, sharing the same opportunities as the others and yet failing to engage with the online material available. We were able to show the detrimental effect of this lack of engagement upon their achievements in the course, but are no more able than Morris et al to see how to motivate them to sample what we have on offer. Only 2% of students who actually investigated our online tests actually turned away from them without any real attempt at the tests.

Conclusions

The patterns of use of online tests have revealed two ways in which we do not appear to be best serving the needs of our first year students. We do not appear to offer sufficient flexibility of access to online summative assessments, and we press students to complete most of these tasks when they are already fatigued.

The question of how to ensure the security and validity of online testing while increasing real flexibility of access remains unresolved for us. There is no sign yet of students turning back to full-time engagement in their university studies. Levels of participation in paid employment by first year students at Australian universities have increased steadily from 1999 through 2001 to 2004 (Krause et al., 2005) and, now, 2006. We need to address this issue, as well as those of building up the abilities of young students to concentrate on academic work, and engaging that one third of the student body not managing to find their own way to effective learning opportunities.

References

Charlesworth, P. and C. Vician. 2003. *Leveraging technology for chemical sciences education: an early assessment of WebCT usage in first-year chemistry courses*. Chemical Education Research 80(11): 1333 - 1337.

Curtis, S. and N. Shani. 2002. *The effect of taking paid employment during term-time on students' academic studies.* Journal of Further and Higher Education 26(2): 129 - 138.

GenStat version 9.1.0.147, 2006 Lawes Agricultural Trust, Rotherham, Hertfordshire, England

Krause, K., Hartley, R., James, R., & McInnis, C. (2005). *The First Year Experience in Australian Universities: Findings from a decade of national studies*. Canberra: DEST. [Available online: http://www.cshe.unimelb.edu.au]

Morris, L. V., C. Finnegan and S. Wu. 2005. *Tracking student behavior, persistence, and achievement in online courses.* Internet and Higher Education 8: 221 - 231.

Rowe, N. C. 2004. *Cheating in online student assessment: beyond plagiarism*. Online Journal of Distance Learning Administration. v7, n2, p1 – 10

Stewart, B. L. 2004. *Online learning: a strategy for social responsibility in educational access*. Internet and Higher Education 7: 299 - 310.

Volery, T. and D. Lord 2000. *Critical Success Factors in Online Education*. International Journal of Educational Management. v14 n5 p216-23.