AC 2011-337: USING NO-STAKES QUIZZING FOR STUDENT SELF-EVALUATIO OF READINESS FOR EXAMS

Kirsten A. Davis, Boise State University

Kirsten A. Davis is an Assistant Professor in the Construction Management Department within the College of Engineering at Boise State University. Dr. Davis earned a B.Arch. in Architecture and a B.S. in Civil Engineering from the University of Tennessee, an M.S. in Civil Engineering specializing in Construction Engineering and Management from the University of Colorado, Boulder, and a Ph.D. in Civil Engineering specializing in Construction Engineering and Management from Virginia Polytechnic Institute and State University. Her educational research interests are focused on improving construction management education.

Using No-Stakes Quizzing for Student Self-Evaluation of Readiness for Exams

Abstract

Introductory level materials and methods courses in Construction Management (CM) programs generally require students to acquire an enormous new vocabulary. This vocabulary includes terminology pertaining to the materials themselves, as well as terminology defining the construction processes using the materials. Courses such as these form a backbone of knowledge that is necessary in nearly every other CM course, whether they are formal prerequisites or not.

Students are often not prepared for the amount of self-directed studying they must do to be successful in this type of course, particularly as freshmen in college. To assist students in more accurately gauging their readiness for an exam, no-stakes (optional) quizzes were implemented within the Blackboard course management system in a freshman level CM materials and methods course, giving students an opportunity to practice their new language with no grade-related consequences.

This paper presents preliminary results of that effort and illustrates the effects of this no-stakes (optional) quizzing. Specifically, this paper evaluates (1) whether the quizzing helps students to better prepare for medium- and high-stakes exams; and (2) whether the quizzing increases the mean exam scores compared to previous semesters with no quizzing.

Introduction

Introductory courses within an academic field of study often require a student to acquire a new discipline-specific vocabulary in preparation for subsequent courses, as well as for their future career. Students are often not prepared for the amount of self-directed studying they must do to be successful in this type of course, particularly as freshmen in college.

The first year of college is often a year of many new and different experiences. There are both academic and social adjustments to be made by students. While the social adjustments are not trivial and should not be ignored (see [1], for example), this paper focuses on the academic aspects. "The first college year is critical not only for how much students learn but also for laying the foundation on which their subsequent academic success and persistence rest."[2]

Students regularly overestimate their knowledge and abilities[3]. As an anecdotal example, there is the student who receives a D or F on an exam, but claims to have studied for hours (studying what, we don't know…), and another student with a similar grade who didn't study at all because they thought they already knew the information well enough. Students also unreasonably expect that their assignment and test grades should reflect not just their achievements, but also the amount of time and effort they expend[4]. We regularly have students in our classrooms, particularly in first-year courses, who fit these or similar molds.

Accurate self-knowledge, the ability to accurately judge one's level of knowledge, is necessary at the college level. Accurate self-knowledge means that a student must be able to realize what

they already know. More importantly, they must also realize what they don't know so they can take steps to ensure that their end knowledge meets instructor and course expectations[5]. While some students arrive at college with this ability, others do not and need help acquiring this skill.

Tied in with accurate self-knowledge is self-regulation. This concept moves beyond the ability to recognize what one does or does not know, to the ability to adjust one's behavior to ensure that acquisition of the missing knowledge is successful[5]. An internal feedback loop is an essential part of self-regulation[6]. Students with better self-regulation tend to have higher academic achievements[5, 6]. Like self-knowledge, self-regulation is a skill that some students need help acquiring.

This paper discusses a strategy using online self-assessments with feedback to help students acquire skills in self-knowledge and self-regulation. Numerous studies have shown that formative assessments lead to significant learning gains[7]. More recently, online self-assessment exercises have been shown to help students improve their self-knowledge[5], and effective external feedback can help students improve their self-regulation skills[6, 8, 9].

Online self-assessments with feedback, also known as online formative assessments, are a relatively new phenomenon, enabled primarily by online course management systems (for example, Blackboard). Students can now take short quizzes outside of class and have them scored automatically and immediately, as well as retake them to improve their scores. While instructors could have provided students with this type of quizzing option in the past, it was impractical without computerized grading[10]. This form of quizzing is essentially allowing the students to practice taking the quiz until they are ready to officially count their score. Simply providing students with their score could loosely be considered a formative assessment, as the students would have minimal feedback regarding their performance (the score), though research is mixed regarding the effectiveness of this practice[11, 12]. However, to truly improve and support student learning, incorporating specific feedback with information that lets them know why their answer is correct or incorrect is an essential part of formative pedagogy[6, 9, 11].

Objectives

This research project investigates implementing formative no-stakes (optional) quizzing and low-stakes (required) quizzing within Blackboard in a freshman level Construction Management course and evaluates: (1) whether the quizzing helps students to better prepare for medium- and high-stakes exams; (2) whether the quizzing increases the mean exam scores compared to previous semesters with no quizzing; and (3) whether the quizzing can be optional (no-stakes) and be effective, or whether it must be required (low-stakes). This paper presents only the preliminary results from the no-stakes quizzing implementation.

Method

This portion of the project (no-stakes quizzing) was carried out with a group of 38 undergraduate students in a 100-level construction management course entitled "Construction Materials and Methods" at Boise State University in the fall semester of 2010. This course is required for construction management (CM) majors and minors and is a three credit course with 45 hours of

classroom instruction. The course introduces students to construction methods and materials used on building projects, with three main objectives:

- Demonstrate knowledge of the methods and equipment commonly used to construct buildings including the foundation and framing systems.
- Identify and discuss the technical aspects of basic building materials such as steel, concrete, masonry, and wood.
- Utilize correct terminology and nomenclature associated with the materials, methods, equipment and building components found on building construction projects.

These objectives are divided into five topical areas for testing purposes: foundations and construction related math, concrete, masonry, steel, and wood and light gauge steel framing.

Of the 38 students enrolled in the course in fall 2010, 22 had formally declared CM as their major, 2 had declared a CM minor, and 14 were non-majors/minors. Attendance at lectures was mandatory and 90% attendance was achieved.

This group of students is being compared with five previous semesters of students in the course (n=32, 28, 58, 42, and 38 respectively for a total of N=198). The characteristics of the students (age, gender, background, major, etc) in previous semesters are similar to the studied population.

Prior to the changes described here, the assessments comprised five paper-based exams with multiple-choice, true/false, matching, and short answer type questions (16% each, with lowest score dropped, for a total of 64% of course grade), participation and attendance at lectures (6% of course grade), and a comprehensive paper-based final exam with multiple-choice, true/false, and matching type questions (30% of course grade). Comprehensive study guides were made available one week prior to each exam. Feedback was only available through marks made on the exams.

This project was developed as a result of the instructor's desire to help students who were struggling in the course, but were genuinely attempting to do well. Based on statements from several students, they were studying and making efforts, but were not passing the exams. This project was created to help students better assess their readiness to take an upcoming exam and consequently improve their self-knowledge and self-regulation. It was intended that exam scores, particularly for these struggling students, would increase.

In the fall of 2010, the course was modified to include no-stakes (optional) online formative assessments. All other aspects of the course remained the same as the previous five semesters, including the grading scale. In the spring semester of 2011, the online formative assessments will no longer be optional, but will instead be low-stakes quizzes and completion of them (not scores) will count approximately 1% towards the course grade (for all quizzes combined).

The quizzes used consist primarily of multiple-choice and true/false questions, along with a few matching questions. The quizzes were completely optional for students and no grade was assigned to them. Each quiz was made available on the course website (within the Blackboard course management system) one week prior to an upcoming exam and along with the exam study guide. The quizzes consisted of 20 questions randomly chosen from a pool of 20-38 topically relevant questions. Students were allowed to take the quizzes as many times as they wished.

Upon completion of the quiz, students were given their score, the quiz questions were repeated with the answers they chose and each answer was marked correct or incorrect. They were also presented with feedback for each question. Feedback for incorrect answers consisted of hints regarding why the answer was incorrect and prompted students to think about a particular aspect more deeply. The feedback also recommended resources to read that discussed the pertinent information, including specific page numbers in the text or other readings. The correct answer was not given, nor should it be when students can retake a quiz[13]. Feedback for correct answers consisted of congratulations and reference information, including page numbers in the text, if they were interested in learning more about that topic.

The medium- and high-stakes exams the students were preparing for consisted of multiple-choice, true/false, matching, and short answer type questions. There were between 41 and 48 questions with a mean of 45 questions for the five medium-stakes exams and 150 questions on the high-stakes final exam. The terms medium-stakes and high-stakes are subjective, but in this project are defined based on their weight towards the course grade, 16% and 30% respectively. Low-stakes quizzing/exams were not used in this portion of the project, but will be incorporated into the spring semester of 2011, as previously described.

A brief one page questionnaire was attached to the back of each student's exam. The questionnaire asked about the amount of study time spent for that exam along with the student's methods of study and what grade they expected to receive on the exam. The questionnaire also asked if they had taken the study quiz. If they had, they were asked if they thought it helped them do better on the exam, why or why not, and what might have helped them more. If they had not taken the study quiz, they were asked why they didn't, and whether they thought it would have helped if they had taken it.

Results and Discussion

Over the fall semester 2010, seven quizzes were made available to the students. Five 20-question quizzes were directly related to exams during the semester. The sixth 20-question quiz covered material subsequent to the last regular exam, but prior to the final exam, and the seventh quiz was a 40-question quiz that covered material from the entire semester with the number of questions from each topic in proportion to that of the final exam. No new questions were written for this last quiz – they were taken randomly from the existing pools of questions.

A natural concern was that because the quizzes were completely optional, the students would not utilize this resource. This concern appears to be unfounded, at least with this group of students, though other researchers have reported low usage when self-assessments were optional[13]. During the fall semester 2010, the seven quizzes were used 767 times by 38 different students (all students enrolled in the course). The total amount of time spent by all students was more than 96 hours. The author was quite surprised at the amount of intrinsic motivation shown by the students, particularly in a first year course. See Table 1 for a summary of the quiz usage during the semester. This data was collected automatically by the Blackboard course website when students accessed the quizzes. To access the quizzes, students had to be registered for the course and logged in; anonymous logins are not allowed.

Table 1. Summary of Quiz Usage in Fall 2010

Quiz #	# of students who took quiz	Total # of times quiz taken	Total # of minutes used	Range of times taken by single student	Range of minutes used by single student
1	31	55	1364	1 - 5	3 - 235
2	35	145	668	1 - 12	4 - 59
3	29	95	570	1 - 6	4 - 148
4	33	169	762	1 - 18	5 - 50
5	30	103	496	1 - 11	4 - 51
6	30	99	533	1 - 11	3 - 80
7	32	101	1399	1 - 12	6 - 185
Total	38*	767	5792 min = 96.5 hrs		
Mean	31.4	109.6	827.4 min = 13.8 hrs		

* total number of different students who took quiz

Independent sample t-tests were conducted on the medium- and high-stakes tests to determine whether the mean test score from Fall 2010 was higher than the mean test score from the previous five semesters (aggregated). Table 2 shows that the students in Fall 2010 had a higher mean test score than students in past semesters for four out of five medium-stakes exams (tests 1, 3, 4, and 5). In the remaining medium-stakes exam (test 2), they performed slightly worse, though it was not a significant difference. In the comprehensive final exam, the mean scores were nearly identical.

Table 2. Descriptive statistics and 1-tailed Independent t-test table

Test #	Group	n	Mean (SD)	t statistic	1-tailed p
Test 1	5 past semesters	198	82.09 (10.15)	-1.97	0.0248*
	Fall 2010	38	85.50 (7.30)		
Test 2	5 past semesters	193	80.56 (12.68)	0.44	0.6697
	Fall 2010	38	79.61 (9.58)		
Test 3	5 past semesters	192	75.95 (12.54)	-1.65	0.0498*
	Fall 2010	37	79.50 (8.10)		
Test 4	5 past semesters	193	77.94 (11.18)	-2.29	0.0116*
	Fall 2010	38	82.39 (9.96)		
Test 5	5 past semesters	189	77.64 (12.56)	-3.18	0.0008**
	Fall 2010	36	84.51 (7.37)		
Final Exam	5 past semesters	190	79.53 (9.04)	-0.16	0.4373
	Fall 2010	38	79.77 (6.09)		

* $p<0.05$; ** $p<0.01$

Beyond the fact that four out of six tests had a higher average score than in previous semesters, the low scores improved when compared to previous semesters for five out of six tests. The

overall improvement of the students in Fall 2010 is exemplified by fewer D and F grades given (see Table 3). This is also the first time (over the six semesters) that everyone enrolled in the course passed with a C or better grade. The course failure rate (D or F grade) over the previous five semesters has ranged from 3.4% to 26.2% with an average of 14.1%.

Table 3. Percent of D and F Grades Given

Test #	Average of 5 Past Semesters	Range of 5 Past Semesters	Fall 2010
Test 1	13.6%	7.1% - 21.4%	2.6%
Test 2	18.7%	9.7% - 27.8%	18.4%
Test 3	25.7%	10.7% - 45.0%	13.5%
Test 4	23.8%	11.1% - 31.3%	7.9%
Test 5	25.4%	14.3% - 29.8%	2.8%
Final Exam	13.7%	6.5% - 25.0%	5.3%
Course	14.1%	3.4% - 26.2%	0.0%

Fall 2010 students were asked open-ended questions regarding the usefulness of the self-assessment quizzes in a brief questionnaire attached to the back of each of the six tests. The response rate for the questionnaire varied by test but ranged from 81.6% (31 of 38) to 94.7% (36 of 38). Of the students who responded, most who took the quiz at least once believed that it helped them do better on that test (80%-100% felt this way, depending on the test). Some sample comments from the students regarding why they thought the quiz helped are listed below:

> "showed what I needed to learn and that I wasn't ready for the test"
> "it gave me a idea where I needed to started [sic] studying"
> "because it improved my confidence and reinforced what I thought I knew"
> "it made me realize I needed to study more"
> "pointed out areas I needed to work on, and where in the text or lecture slides to find the info"

A majority of the comments of dissatisfaction related to the fact that the questions on the quiz were not the same as those on the test. Additionally, there were a few students who took a quiz only once, got a good score and falsely assumed they were prepared, yet had trouble when they took the exam. Had they taken that quiz more than once, they likely would have realized they needed to study more. All of those who responded that they did not take the quiz believed that taking it would have helped them do better on that exam.

While the intended strategy for students using the quizzes was to study, take the quiz, learn about the material they missed, and then retake the quiz, not all students followed this path. Based on quiz usage and comments on the questionnaire, there were several other strategies used as well. Some students used the quiz prior to studying to judge where they were and what they should study (and, likely, how much time they might need to spend). The students using these two strategies were definitely getting practice developing their self-knowledge and self-regulation skills.

Other students used the quiz itself as the study tool, retaking it over and over again until they were successful, while apparently ignoring any feedback they received for incorrect answers (data showed multiple attempts with 1 minute or less between attempts). This method implies trial, error, and memorization, which would not be the recommended use of the quizzes. This method was discouraged through the technology by having the questions randomized from a larger pool of questions and having the answers to each question randomized as well. This group of students was not really improving self-knowledge or self-regulation. They appear to be more focused on the quiz score, instead of gaining useful knowledge.

Still other students studied first and only took the quiz when they were done studying. It is likely that students using this last strategy are made up of two groups. Some of these students seemed to only use the quiz to get a feel for what grade they might get on the test, and if it met their needs, they were done. They are improving their self-knowledge skills, but are likely ignoring self-regulation. They aren't willing to adjust their behavior any further to improve their knowledge. Others used the quiz to improve their confidence that they were prepared for the upcoming test. These students are likely to be the students who are well-versed in both self-knowledge and self-regulation. They would do well on the exam whether the quizzes were available or not.

Regardless of the strategy used, any improvement in self-assessment skills (i.e., self-knowledge and self-regulation) in students leads to students being more prepared to be lifelong learners[4], something that all teachers can appreciate.

A limitation of this study is the type of questions asked in the quizzes. It was limited predominantly to multiple-choice and true/false questions because of the desire to have the quizzes automatically graded and scores/feedback available immediately upon completion. The author has experimented with other question types (in other courses), such as fill in the blank questions, but they have been largely unsatisfactory. However, the question types chosen do not limit the depth of knowledge required of students. Multiple-choice questions are not limited to only the lower levels of Bloom's taxonomy, such as recall, comprehension, and application. Well thought out questions and answer choices can also test higher levels, such as analysis and evaluation[10]. This course requires students to learn a great deal of terminology related to their major and the quizzes (and corresponding tests) reflect this with many lower level questions. However, the quizzes and tests in this course also include some questions at higher levels because it is important for students to begin analyzing and evaluating choices between materials and construction methods, as this is something they will be expected to do when they begin their career.

Another limitation of the study relates to the tests being compared. This project compared the Fall 2010 tests with tests from five previous semesters and the tests over the six semesters were not the same. While having different tests is not an ideal situation, it would not have been practical to use the same tests for multiple semesters. Using the same tests over and over would likely have introduced more bias when comparing test scores between semesters (due to students sharing information) than having different tests. To limit the issues introduced by having different tests, they were similar in content and format, and they were all written and graded by the same instructor for all six semesters. The descriptive statistics (high, mean, low, s.d.,

median) from one semester to the next have been similar (see Table 4 for an example). To statistically minimize the differences between tests that do exist, the independent t-tests performed compared the aggregate of the previous five semesters with the Fall 2010 semester. Additionally, it is planned that the study will be repeated with additional students in future semesters. With repetition, the differences between the tests (and the students taking them) begin to disappear.

Table 4. Descriptive statistics for Test 1

Semester	# taken	High score	Mean score	Low score	s.d.	Median score
1	32	95	78.59	62	8.48	79.5
2	28	96	81.80	68	6.56	82.75
3	58	103	83.93	60	9.38	85
4	42	100	80.81	56	12.51	82.5
5	38	100	83.84	50	11.28	85.5
Fall 2010	38	97	85.50	69	7.30	86

Conclusions

The study described here was used to supplement an existing course with a number of students who needed help with their self-knowledge and self-regulation skills when studying for medium- and high-stakes exams. It appears to have been successful in addressing this problem. Specifically, nearly all students perceived that the quizzing helped them better prepare for medium- and high-stakes exams, and the quizzing increased the mean exam scores for four out of six exams, compared to previous semesters with no quizzing. Additionally, the quizzing appears to have drastically reduced the failure rate on the exams and the failure rate of the course. Similar quizzes could certainly be used for lower division courses in other disciplines where students have comparable issues adjusting to the requirements of college and the concept is definitely scalable to larger class sizes.

Intrinsic motivation seemed not to be an issue with this group of students based on the high percentage who took the quizzes. However, this study will be repeated in the spring semester 2011 with a low-stakes version of the quizzes to determine whether larger gains can be made with required quizzing.

Acknowledgements

This project was supported by a grant from the Center for Teaching and Learning at Boise State University.

References

1. Wilcox, P., S. Winn, and M. Fyvie-Gauld. (2005). "'It was nothing to do with the university, it was just the people': the role of social support in the first-year experience of higher education." *Studies in Higher Education*, 30(6), 707-722.
2. Reason, R.D., P.T. Terenzini, and R.J. Domingo. (2006). "First Things First: Developing Academic Competence in the First Year of College." *Research in Higher Education*, 47(2), 149-175.
3. Glenberg, A.M., A.C. Wilkinson, and W. Epstein. (1982). "The illusion of knowing: Failure in the self-assessment of comprehension." *Memory & Cognition*, 10(6), 597-602.
4. Taras, M. (2003). "To Feedback or Not to Feedback in Student Self-assessment." *Assessment & Evaluation in Higher Education*, 28(5), 549-565.
5. Ibabe, I. and J. Jauregizar. (2010). "Online self-assessment with feedback and metacognitive knowledge." *Higher Education*, 59(2), 243-258.
6. Nicol, D.J. and D. Macfarlane-Dick. (2006). "Formative assessment and selfregulated learning: a model and seven principles of good feedback practice." *Studies in Higher Education*, 31(2), 199-218.
7. Black, P. and D. Wiliam. (1998). "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
8. Nicol, D.J. (2009). "Assessment for learner self-regulation: enhancing achievement in the first year using learning technologies." *Assessment & Evaluation in Higher Education*, 34(3), 335-352.
9. Miller, T. (2009). "Formative computer-based assessment in higher education: the effectiveness of feedback in supporting student learning." *Assessment & Evaluation in Higher Education*, 34(2), 181-192.
10. Thelwall, M. (2000). "Computer-based assessment: a versatile educational tool." *Computers & Education*, 34(1), 37-49.
11. Gipps, C.V. (2005). "What is the role for ICT-based assessment in universities?" *Studies in Higher Education*, 30(2), 171-180.
12. Davis, K.A. (2009). "Improving Motivation and Knowledge Retention with Repeatable Low-Stakes Quizzing." *2009 ASEE Annual Conference & Exposition*, Austin, TX, USA.
13. Buchanan, T. (2000). "The efficacy of a World-Wide Web mediated formative assessment." *Journal of Computer Assisted Learning*, 16(3), 193-200.