

What Online Quizzing Can Tell Us About Our Students

Jonathan P. Mathews*, Sarma V. Pisupati*, and Andrew Wiesner[±]

***Energy and Geo-Environmental Engineering Department and
John A. Dutton e-Education Institute
College of Earth & Mineral Sciences**

**[±]Schreyer Institute for Teaching Excellence
The Pennsylvania State University
University Park, PA 16802**

jmathews@psu.edu | spisupati@psu.edu | ajw13@psu.edu

Abstract

Computer based quiz and exam results from a large enrollment general education class were analyzed to determine what an in-depth analysis of the quizzing data could tell us about our students. Analysis of variance methods were implemented to study the effects of gender, grade point average, and class standing on overall test performance as well as the multiple choice and short answer sections of the tests. The effects these factors held on student quiz behavior in relation to their exam scores was also evaluated. Important considerations were given to behavioral differences based on gender and GPA. The “B” or better students were less likely to miss the weekly quiz (3%) than the lower GPA student (20%). Students were likely to at least match their incoming GPA with course grade in this general education course. Multiple comparisons for essay time determined that significant differences were present between the A, B and the C, D, F students, with the better performing students taking more time to complete this portion of the exam. Within these timed exams, there was no significant difference in quantity typed between the sexes, 752 words for males in comparison to 742 words (35 minute exam with instructions to answer 4 of the 6 questions.) Evaluating short answer essay questions against “multiple choice” scores within three timed exams indicated that 71.5%, 90.6%, and 77.2% of the students did better in the essay component in comparison to the multiple choice component for the 3 exams. For the “multiple choice” portion of the exam, surprisingly with 38 questions only 14% male and 16% of the female students used >20 minutes in answering the questions. The average time required being 15 of the 25 available minutes. An analysis of the essay portion of the exam indicates that both genders submitted the exam after an average of 32 minutes, rather than utilize the full 35 minutes. Approximately 20% of the class, and each gender submitted the exam in less than 30 minutes. Most of the students were tenacious in obtaining the 100% score for the weekly quiz, despite the value being 1% of the grade. Hence, this approach was considered useful for a directed review of the material.

Introduction

Computer based testing can: be convenient, reduce academic dishonesty, provide quick/immediate feedback, has the potential to offer better testing, use media, and enable more consistent grading when applied to large enrollment classes¹. However, within a large institution such as Penn State the use of computers for exams is rare. A major barrier is the availability of a facility to accommodate students. The administration does not permit general student computing facilities to be scheduled (for exams), due to the high usage and negative impact of reduced computational availability. There are, however, several classrooms with a computer for each student but they are in use throughout the normal class schedule. Currently the administration is considering a significant investment in a centrally located, high stakes, large scale proctored computer based testing facility. With increasing number of online classes, use of multimedia and technology there is interest in enabling large scale computer-based testing to better serve the students and enable better testing methodologies that the traditional paper and pencil “fill in the circle” approach.

For one WEB class “Energy & the Environment” (class is conducted entirely online) computer-based testing has been utilized for weekly quizzing, and for examinations in proctored rooms. The use of computer-based testing enables the collection of student usage data that would be difficult to obtain for paper tests. An analysis of the timestamps, submission frequency (quizzes), duration, and performance was performed for the low-stakes weekly quizzes that can be taken multiple times, and high-stake proctored exams. The computer-based software used currently is “Test Pilot” from “ClearLearning Corporation”². This particular software was chosen for the response-specific feedback ability and is delivered via web browsers to students in a location of the student’s choice for the low-stake weekly quizzing. The questions are delivered from a question bank, thus, each quiz can be somewhat unique to the student. The use of random number generation (within specified ranges) ensures calculations are also unique. These approaches were taken to discourage academic dishonesty. However, the philosophy of the lead author is that this type of weekly quizzing is an additional directed learning opportunity rather than an evaluation of the student performance. Thus, the use of response specific feedback for questions was adopted with textual or visual feedback. These weekly quizzes are computer graded and include a variety of question types including: True/False, Select all that apply, Fill in the Blank, and Mathematic calculations. Questions can be textual, based on tables, graphs, or other media such a browser plugin enabled molecular structures. The collection of these question type will be referred to generically as “multiple choice” questions throughout this paper. About 300-400 students take 11 weekly quizzes every fall and spring semester. These weekly evaluations can be taken multiple times, the highest grade being recorded (via an excel macro) immediately after the due date. These assessments contribute 13% of the total grade.

For each of the 3 examinations, two proctored rooms containing 80 computers each are utilized over a four-hour period in the evening to accommodate the 400 students. Evening exams are required because of scheduling limitations on the heavily utilized computer

classrooms and due to student class conflicts. Students sign up for a 1-hour slot for the exam. The exam contains two components: an essay portion with 4 out of 6 questions to be answered during a 35 minute computer timed quiz, and a multiple choice style portion (incorporating True/False, multiple choice, select all that apply question types) that runs for 25 minutes. Access to the examination is restricted to computers in the room. Access is controlled by the Internet Protocol (IP) address. Exam security is enhanced by preventing printing via a JavaScript code. Secure windows prevent saving the browser page. The right-click mouse button is also disabled to prevent copying and pasting content. When the time limit expires (based on when the student accessed the exam) the exam is automatically submitted. Teaching assistants and the instructor grade the written portion of the exam with one person grading all the submissions for a specific question, to ensure consistency.

Results

Low-Stakes (Quizzes)

In a class of 333 students (spring 2004), 854 submissions were made to one of the weekly Test Pilot computer graded quiz. The average number of submissions prior to the deadline was 2.8 per student. Only student submissions are recorded, with questions being delivered one at a time with immediate feedback between questions. Some students, upon failing a question, choose to close the quiz window and restart the quiz. Most of the students were tenacious in obtaining the 100% score, despite the value being 1% of the grade. Hence, this approach was considered useful for a directed review of the material. Students value points, particularly 100% and can be lured into time on task reviewing the material. The time on task for a weekly quiz, and quiz submission frequency is shown in Table 1. However, this can also have negative consequences. One student utilized 13 attempts, during 55 minutes of effort for one particular quiz. Effort that would have been better spent comprehending the material rather than the “try until you succeed” approach lacking the comprehension. It is permitted for the student to access the notes or any other electronic source in search of the answer(s). Nearly all the class obtained 100% for these quizzes (enhanced (directed) learning opportunities). This particular quiz contained the 20 questions from a question bank of about 40 questions. Often this activity was immediately prior to the deadline. Another companion paper, in these proceedings, discusses the procrastination issue and consequences.

Table 1. Time on Task for quiz submission(s) and submission frequency

Submissions	1	2	3	4	5	>6
Frequency (%)	38	15	16	13	9	8
Total Time (mins)	8	16	27	30	36	

The number of students who failed to attempt the weekly quizzes are shown in Table 2 for an earlier semester. The numbers shown are for the students who finished the course. Of the 315 students, 73% (231/315) had a score in each of the 11 quizzes, typically 100 or high 90's. Forty-five students missed one lesson quiz. However, for 6 of these students

(13%) it reduced their letter grade one notch. The majority of these grade reductions being A to an A-. As expected, the percentage of students who were influenced by multiple missing quizzes increased with increasing missed quizzes. Missing 3 quizzes (10 students) resulted in 5 grade reductions. For the 6 students who missed 4 of the quizzes there was no impact on the grade. All these students were at the “bottom of the grade barrel”. Simply put, missing assignments even low-stake 1% assignments will influence the letter grade. As the quizzes are designed to be a directed additional learning experience, the cumulative GPA was compared with likelihood of missing assignments using ANOVA and multiple comparison methods. The differences were hypothesized: students with higher GPA are more likely to submit assignments than the lower GPA students. The p-value for the overall ANOVA was less than 0.001. It was found that the A, B and C students were more likely to submit the web quizzes and reflective assignments than students with a GPA below a C average. As expected the better students were less likely to miss assignments. Similar results linking frequent testing and exam performance have been previously reported³. In another study, females were shown to be less likely to miss a quiz⁴.

Table 2. Number of quizzes missed and number of grades it influenced

orig	no missing	1 tp	2 tp	3 tp	% change
0	231				
1	45	6			13
2	18		3		17
3	10			5	50
4	6				
Total	315				

The student performance in the class was compared to student incoming grade point average (GPA). The results are shown in Table 3. Students were likely to at least match their incoming GPA. Table 3 also indicates that a student’s potential to fail the course increased as student overall GPA declined, particularly the C and D students. As general education classes are stand-alone classes it is no surprise that the B and C overall GPA students tend to achieve a better grade in this class. Alternatively, this could also be interpreted as an example of grade inflation

Table 3. GPA versus final course grade (values shown are percentages) spring 04.

Incoming GPA	Course Letter Grade				
	A	B	C	D	F
A	87	10	0	0	3
B	41	53	4	1	1
C	15	42	27	4	12
D	0	25	12	0	63

High stakes (Exams)

A plot of the time utilized for the “multiple choice” style portion of an exam is shown in Figure 1. It shows considerable scatter and diversity of student behavior during the exam. The shortest submission time was slightly over 5 minutes with a respectable grade. For this particular exam a grade of 72% was considered a “B” grade. Thirty-two students failed (score of <60%) this portion of the exam. Overall, there is a large spread for time utilized and no clear relationship between time utilized and performance. All the “multiple choice” questions were presented on a single webpage (drawn randomly for each student). A countdown timer was displayed at the top of the webpage. A 5-minute warning alerted students of the impending automatic submission. Only a few required the full 25-minute allocation to answer all 38 questions. Each student had the same number of question types, for example this exam had 4 True/False questions, and there were no calculations. The mean submission time was 15 minutes, with only 15.5% of the class requiring > 20 minutes. An analysis based on gender indicated very similar behavior.

A similar analysis was performed for the essay portion of the exam where the student had answered 4 of the 6 questions. If 5 were attempted, the lowest score was dropped. Within these timed exams genders wrote the same volume: 752 words (male) in comparison to 742 words (female). Both genders submitted the exam after a mean of 32 minutes, rather than utilize the full 35 minutes. Approximately 20% of the class, and each gender submitted the exam in less than 30 minutes. Applying ANOVA techniques to analyze exam grade versus time on task for the essay and “multiple choice” portions of the exam resulted in non-significant differences for multiple choice ($p = 0.624$), but a significant difference resulted for the factor of essay time ($p < 0.001$). Multiple comparisons for essay time determined that significant differences were present between the A, B and the C, D, F students, with the better performing students taking more time to complete this portion of the exam.

When moving from paper to electronic exams, several years ago, the authors expected some student apprehension over typing issues. However, it is rare to have students cite typing as an issue for poor exam performance. Poor typing skills can be mediated with the appropriate exam performance strategies. The exam scores in this class are dependent on content, knowledge, and synthesis not word count limited. Often students add correct, yet not applicable information, to add bulk to their answer. In online exams at the pre-college level the digital divide many place the disadvantaged students at a further disadvantage, particularly with typing answers. Surprisingly, the countdown timer has often been cited (by students) as adding to the stress of the event. Women students may be more exam anxious in such settings⁵. Students officially identified as having learning difficulties take the exam in quiet rooms with extra time allocated.

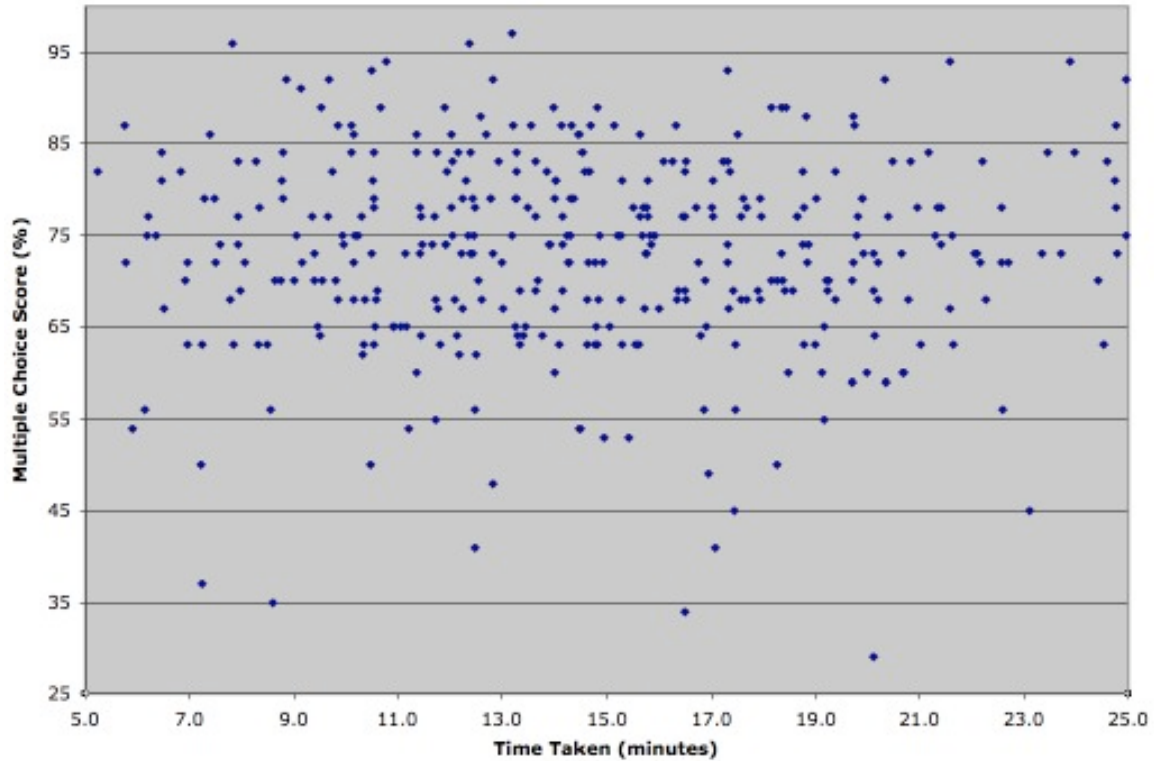


Figure 1. Plot of time utilized prior to submission for the “multiple choice” portion of exam one (fall 2004 semester)

A plot showing the percentage difference between the essay portion and the “multiple choice” style portion of the exam is shown in Figure 2. It is not common in large enrollment classes to have essay questions given the grading burden and limited grading assistance. Thus, it was interesting to compare the difference in scores between the essay and “multiple choice” style portions of the exam. Students had a wide range of variances with high and low differences of +35% and -37%. Seventy one percent of the students obtained a higher grade in the essay portion of the exam. Sixty three percent of the students are within +10 to -10% band. Dropping the essay portion of the exam in favor of a multiple choice style exam would have implications for exam scores. The students taking this general education class are generally math and science phobic. It is expected that these students will perform better in the essay portion where quantification is less prevalent. Figure 3 shows the plot of the difference between exam components and the overall exam score. As expected the top performing students did better in both portions of the exam.

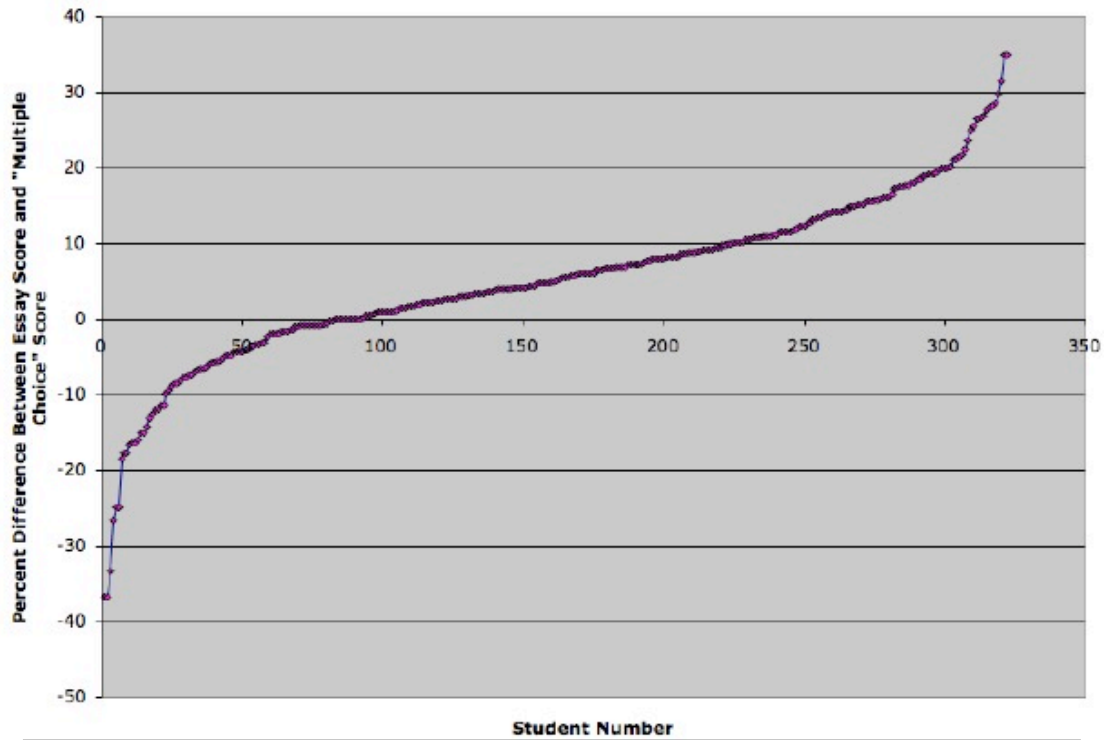


Figure 2. Plot of exam 1 score difference (essay – multiple choice) vs. student number (spring 2004 semester)

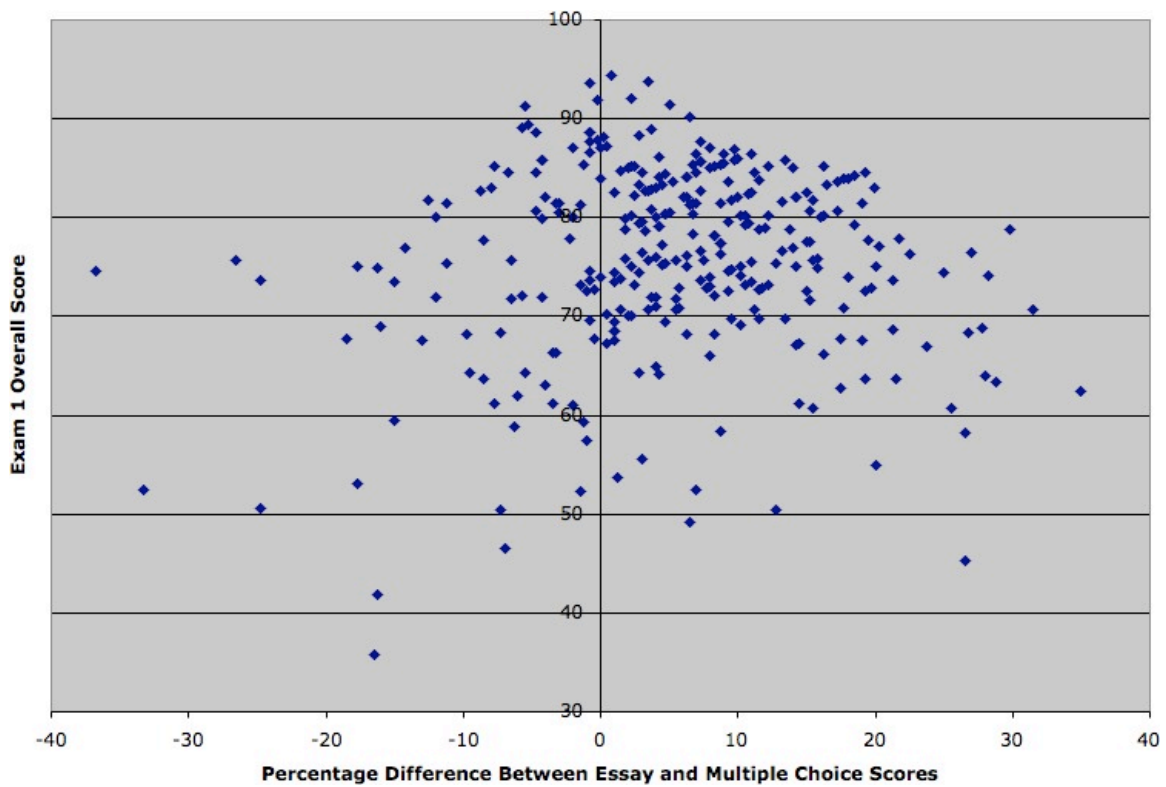


Figure 3. Plot of exam 1 score difference (essay – multiple choice) vs. the overall exam 1 score

Conclusions

An analysis of exams and quizzes within a large enrollment general education classes was performed. “Multiple choice” style testing was performed rapidly by the students with a mean submission time of 15 minutes, with only 15.5% of the class requiring > 20 minutes for the 38 questions. A gender analysis indicated very similar behavior. There was no relationship between score and time utilized for the multiple choice portion. For the essay portion a multiple comparison determined that significant differences were present between the higher (A, B) and the lower GPA students. The higher GPA students took more time to complete this portion of the exam. Missing one weekly quizzes valued at approximately 1% of the course grade resulted in a grade reduction (from A to A-) for 13% of the students. The likelihood of a grade reduction increased with frequency of missed quizzes. Based on the incoming GPA the A, B and C students were more likely to submit the weekly quizzes and reflective assignments than students with GPA below a C average. Students were likely to at least match their incoming GPA with course grade in this general education course. Most of the students were tenacious in obtaining the 100% score for the weekly quiz, despite the value being 1% of the grade. Hence, this approach was considered useful for a directed review of the material.

Bibliographic Information

1. McGraw-Hill, C., 2005, <http://www.ctb.com/articles>
2. ClearLearning, 2005, <http://www.clearlearning.com/>
3. Tuckman, B. T., 2003, <http://all.successcenter.ohio-state.edu/belgium-paper/BWT-belgium-paper.htm>
4. Gunn, C., McSporran, M., Macleod, H. and French, S., Dominant or different? Gender issues in computer supported learning, *Journal of Asynchronous Learning Networks*, 2003, **7**, (1).
5. Steveler, R. A., Hoeglund, T. and Stein, C., The study strategies of academically successful students at the Colorado School of Mines, *33rd ASEE/IEEE Frontiers in Education Conference*, 2003, Boulder, CO, **T2A**, 1.

Bibliography

JONATHAN P. MATHEWS is an Assistant Professor in the department of Energy & Geo-Environmental Engineering, and a Fellow of the John A. Dutton e-Education Institute. His educational interests are enhancing undergraduate online learning for success in lifelong learning. He teaches the largest enrollment WEB course at Penn State.

SARMA V. PISUPATI is an Associate Professor of Energy & Geo-Environmental Engineering Department and a Faculty Fellow of the John A Dutton e-Education Institute of the College of Earth and mineral Sciences. He is Chair of the General Education Program of the EGEE Department and has been teaching undergraduate and graduate courses at Penn State University since 1992.

ANDREW WIESNER is an educational testing specialist in the Schreyer Institute for Teaching Excellence.