

EXAMINATION TIMING AND STUDENT PERFORMANCE: THE LONGER YOU TAKE THE WORSE YOU WILL DO?

Hirotohi Takeda
Georgia State University
University of Paris Dauphine
htakeda@cis.gsu.edu

Abstract

This study is a research in progress taking an exploratory look at the student exam scores dependent on the amount of time that a student takes to finish the exam. Four different multiple choice exams were used for data analysis. Students were undergraduate students in an Introductory Computer Information Systems Course from a university in the Southeastern US. Linear regression analysis was used to see how much exam scores were explained by the time that the student took to finish the exam. The data indicates that a trend where the more time one takes to finish the exam, the worse the score gets.

Keywords: Examination, Information Systems, Multiple Choice, Timing.

Introduction

The time required by students to take a multiple choice exam is different, dependent on the student. Students take different amounts of time when given a multiple choice exam, seemingly in a random fashion. As instructors we are obligated to give good instruction and good test taking advice as well. We often give the advice to “go with your initial gut instinct” (Dutch, 2006, <http://www.couns.uiuc.edu/brochures/testanx.htm>)

There are different ways of reasoning how a student may use their time. One train of thought seems to be that good students know the material so they are prone to take less time. Another way of reasoning is that good students are diligent in checking their answers during test taking so they would take more time. There seems to be no logical conclusion as to how good students will behave. Conversely there seems to be no logic in the behavior of poorly performing students. One thought is that poor performing students take less time because they do not know the answers. Another way of reasoning is that poor performing students take more time because they are struggling through the exam.

One aspect that might be a factor in the time it takes a student to finish a multiple choice exam is the time it takes the student to review the question. The old logic is that the first choice is probably the correct choice. If you go back and are unsure, the best choice was your first instinct. But if this is true, and students follow this strategy, then high performing students should take less time to finish an exam.

This research is an exploratory research in progress, where testing times were measured for a course for “introduction to computer information systems”. The students were given several multiple choice exams and the scores and time that the students took to finish the exam was recorded. The research will try to make some sense in how students take multiple choice exams.

Related Research

A literature review for a study that looks at the time that students take to finish a multiple choice exam was done with no results. Past studies have been done on the ordering of multiple choice exams, test taking strategies, and the amount of time that students put into their studies.

The ordering of multiple choice exams has been studied. This is the order of the questions, whether sequentially as introduced in the readings or lectures, or a scrambled ordering. When the ordering of questions is scrambled, the

student scores are lower (Balch, 1989) but research exists that denies these findings (Neely, Springston, and McCann, 1994).

There are papers that study test taking strategies. The old adage that, when in doubt go with your first impression on a multiple choice exam, has been confirmed and debated. Some say that this is a good strategy and that when you re-read and consider other choices, you start to read things that are not in the question. Some say that be careful but if you have to change but the myth of always sticking with the first impression may not be true (Dutch, 2006, Edwards and Marshall, 1977, <http://www.couns.uiuc.edu/brochures/testanx.htm>). Other studies look at strategies for students with disabilities (Barry and Moore, 2004).

The amount of time that a student puts into studying inside and outside of class has also been addressed. The time for self study may be of little significance to the exam scores (Dolton, Marcenaro and Navarro, 2003). Other studies looked at how students allocate their study time with respect to the material being presented (Son and Metcalfe, 2000).

Data Collection Methodology

The observation methodology was used in this study. The pilot study was conducted on an “introduction to computer information systems” course required by all business undergraduate students. The course was offered at a large urban south east US university. There was no prerequisite for the course. The course offers a challenge to students as there is a wide variety of topics covered in the course (breadth) but each topic is not covered in much detail (depth). The majority of the material of the course is remembering terminology and facts within the information technology (IT) field. This requires that the course be heavily oriented towards multiple choice testing as opposed to essay questions.

One course under one instructor was used in this study. Students enrolled in this course were required to finish four multiple choice exams for this study. Three of the exams were of the same format, where there was 55 minutes to finish the exam, and there were 50 questions for a total of 50 points. The final exam was longer, with 120 minutes to finish with 100 questions for 100 points. The data was taken from a real course, therefore the timing of the tests were set by a course coordinator. Timing was not controllable by the researcher.

All exams had three versions. This was to control for academic dishonesty. The order of questions in all the exams was the same across versions. The order of the multiple choice answers were the only items that were scrambled. The difference in the three tests (referred to as Test1, Test2, and Test3 in this study) and the final exam was controlled for by measuring exam scores by percentages, and dividing the time the students took for the final exam by a factor of 2 to make the minutes used similar between the exams.

All questions for the four exams were multiple choice. Exams scores were taken from the student records. Student names were coded so that anonymity would be retained. The timing was taken during the exam. The instructor had a tabulation sheet to indicate the number of students that turned in the exam at a certain minute past the hour. The tabulation was only to indicate the number of students that turned in during each minute and not the actual student name. The exams were kept in a pile to preserve the order that the exams were submitted. After each exam, the researcher took the tabulation sheet and the exam pile to correlate each exam with the order in which the exams were submitted. The time it took for each student to finish the exam was thus recorded.

Make up exams were given to students that had valid excuses. For several students on some of the exams time data was not achieved due to this fact. When student were given a make up exam, the instructor was not present on some occasions, so only score information is achieved.

Some students missed exams altogether. Some students dropped the class after taking one or two exams. These data points were kept for students that didn't continue in the class. The data is analyzed by a test by test basis and not on data from test to test (see Table 1).

Table 1. Test Statistics

	Number of Problems	Time Allocated	Average Score	Average Time	N
Test 1	50	55 minutes	76.4 %	29.6 minutes	41
Test 2	50	55 minutes	76.9 %	45.5 minutes	38
Test 3	50	55 minutes	78.9 %	33.5 minutes	34
Final Exam	100	120 minutes	74.8 %	38.0 minutes	36

The first three tests were given in chronological order. Each test was devised to cover the same amount of information. Test 2 was considered to be harder than Test 1 and Test 3 because Test 2 covered material on the Database portion of the course. The Database portion of the course consisted of learning to use Microsoft Access and recognizing Database code and SQL code. This portion of the test was the most technical, and thus, foreign to most students. Yet the data indicates the students performed close to average compared to the other two tests. The data does indicate the students did take the most amount of time during Test 2. The Final exam was cumulative with an emphasis on the material covered between Test 3 and the last day of class. The data seems to indicate that students struggled on the Final exam compared to the other three tests given the low average score (see Table 1).

Findings

The analysis began with looking at the edge points of the data. The four edge points would be the data point where the least amount of time was taken, the data point where the most amount of time was taken, the best scoring exam, and the worst scoring exam. When looking at the time spent, test 2 shows there were four students who took the maximum 55 minutes to complete the exam. This may be an indication (along with the average time being 45.5 minutes) that the time allocated for the second exam may be too short. When there was a tie such as the case where four students took the maximum 55 minutes for test 2, all the corresponding data points are reported in Table 2.

Looking at the students that took the longest to finish the test, their average grade was 69.4, which compared to the average overall scoring over all four exams (76.7) was considerably lower (by 7.3 percent). On the opposite side of the spectrum, the students that finished the test first had an average grade of 81.0, which was 4.3% higher than the average. Five of the seven students that took the longest were getting grades of D or less, while three of the four students that finished first were getting a B, and the fourth student was right at the average for average of all four tests. The student taking the longest scored the worst for test 1 and test 2, but the fastest student was never the highest scoring student. This tends to indicate that you can have the worst student struggling and trying to take as much time to try to find the answer during the test. While the best student will not be the first, typically a student that finished first is going to get a good grade on the test.

Looking at the top finishers in scoring, their average time was 37.6 minutes, which compared to the average time for all four exams (36.5 minutes), was around the average. The bottom finishers in scoring averaged 44.3 minutes which is much longer than the average time. This seems to echo the result found when looking at the first and last finishing student with respect to time, where the students that did well finished an average of about 7 minutes faster than the student that the students that did poorly. Students that are struggling were taking more time to finish the exam.

Table 2. Edge Data Point Analysis

	Longest Time (in minutes)	Longest Time Score	Shortest Time (in minutes)	Shortest Time Score	Best Score	Best Score Time (in minutes)	Worst Score	Worst Score Time (in minutes)	Average Score	Average Time (in minutes)
Test 1	42	54	16	84	92	23	54	36, 42	76.4 %	29.6
Test 2	55	62, 66, 80, 88	32	82	92	54	62	42, 55	76.9 %	45.5
Test 3	54	68	20	76	94	27, 35, 41	52	48	78.9 %	33.5
Final Exam	54.5	68	26	82	90	45.5	58	43	74.8 %	38.0

Linear regression analysis was run on the four tests with the dependent variable being the score and the independent variable being the time taken. A fifth linear regression was run to see how the final grade will be predicted by the four times taken. For this equation the dependent variable is the final grade, and the independent variables are the four exam times. The result of the regression analysis is in Table 3. There is little correlation from the time to the exam score but the coefficients are all negative, indicating a negative slope or a degradation of grade as you take a longer time to finish your exam. Test 1 seemed to indicate the highest correlation between the longer time taken and the loss of score. This correlation is consistent with the edge point analysis. When the time for all four tests was used to explain the final course grade, a greater understanding was found.

Table 3. Regression Analysis Output: The Coefficients for the Final Grade Are in the Order Test 1, Test 2, Test 3, and Final Exam

	R	R ²	Beta	Coefficient(s)	Significance
Test 1	0.256	0.066	87.239	-0.365	0.106
Test 2	0.065	0.004	80.735	-0.074	0.700
Test 3	0.156	0.024	85.271	-0.193	0.379
Final Exam	0.182	0.033	83.322	-0.215	0.287
Final Grade	0.358	0.128	79.058	0.072, 0.407, -0.314, -0.178	0.429

The following are the scatter plots for the four tests and a combined scatter plot. The fit line has been added to all the plots and they are all negatively sloping, indicating the relation that the longer you take the worse your grade will be.

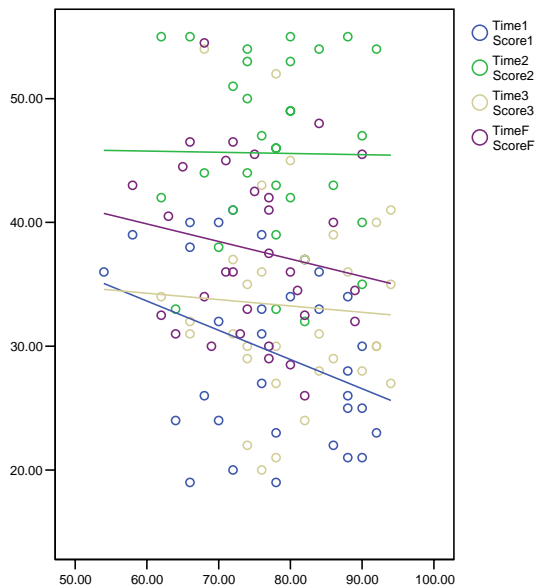
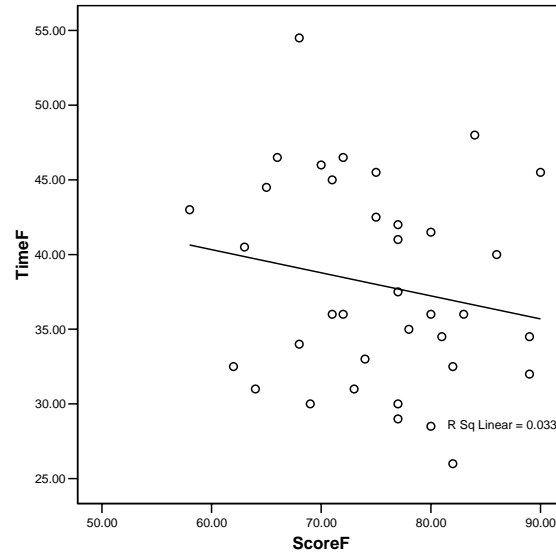
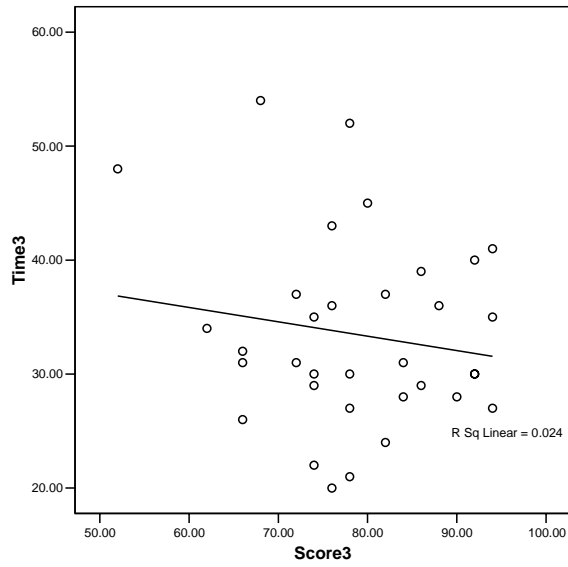
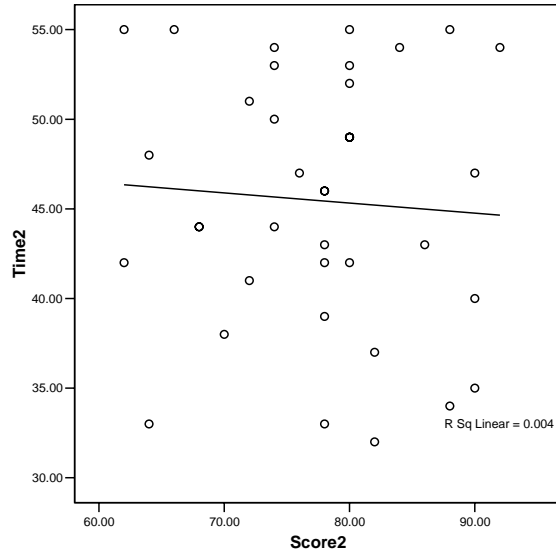
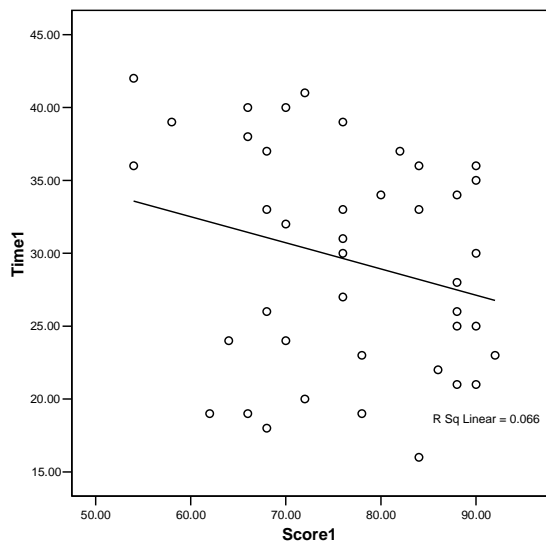


Figure 1. Scatter plots of four exams and combined plot

Limitations

The study did not find high significance. Lack of high significance is expected as students do not behave in the same exact way. There are students who do well whether they finish early or finish late. There are students that do poorly, but who finish early (even though the data did not find this too often) as well as some who do poorly when finishing late. So the study was taking the overall picture and cannot say that just by taking more time, you will do poorly.

The study was taken on only one class during one semester. More data and different courses need to be included. The continuation of this project will allow for more data taking and analysis.

Some data was missing due to lack of control by the instructor in special situations. These situations were those such as where make up exams were given outside of the control of the instructor.

Demographic information was not taken, thus not controlled. The only demographic information available to the instructor is the gender, class year, and major of the student. In the future demographic information such as work experience, course experience, reading ability, testing ability, and English as a foreign language status, may be taken and controlled for.

Conclusion

This study is a research in progress that took an exploratory look at the student performance on multiple choice exams with respect to how much time they used for the exam. The findings tend to indicate that students who use more time tend to do poorly than students that finish earlier. The data is a trend and not a conclusive correlation. This study can allow educators in multiple choice exam situations to give advice to students as to how to try to take the exam. Students that are changing answers naturally are taking longer, thus the old adage of going with your first instinct may hold true. Other findings indicate that the first finisher tended to do better than the average, and that the last one was usually a poor performer.

References

- Balch, W. R. (1989). Item Order Affects Performance on Multiple-Choice Exams. *Teaching of Psychology*, 16(2), 75-77.
- Barry, L. M., & Moore, W. E. I. (2004). Students with Specific Learning Disabilities Can Pass State Competency Exams: Systematic Strategy Instruction Makes a Difference. *Preventing School Failure*, 48(3), 10-15.
- Dolton, P., Marcenaro, O. D., & Navarro, L. (2003). The effective use of student time: a stochastic frontier production function case study. *Economics of Education*, 22, 547-560.
- Dutch, S. (2006). Taking Multiple Choice Tests. Retrieved 1/5/07, from <http://www.uwgb.edu/DutchS/multchc.htm>
- Edwards, K. A., & Marshall, C. (1977). First Impression on Tests: Some New Findings. *Teaching of Psychology*, 4(4), 193-.
- Illinois, U. o. (2006). Counseling center. Retrieved 1/5/07, from <http://www.couns.uiuc.edu/brochures/testanx.htm>
- Neely, D. L., Springston, F. J., & McCann, S. J. H. (1994). Does Item Order Affect Performance on Multiple Choice Exams? *Teaching of Psychology*, 21(1), 44-45.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and Control Strategies in Study-Time Allocation. *Journal of Experimental Psychology*, 26(1), 204-221.