Correlation between Exam Completion Sequence and Resulting Scores in a Construction Management Program

Bolivar A. Senior, PhD, Scott Glick, PhD, LEED AP, and Carla Lopez del Puerto, PhD
Colorado State University
Fort Collins, Colorado

This study explores the correlation between the order in which multiple-choice exams are completed and the average scores obtained in these exams. Midterm exams of four undergraduate courses in construction management offered in the same semester were analyzed and compared with other sections of the same course and with the results from the other courses. Average scores were computed for sets of five consecutive exams in sections ranging in size from 33 to 80 students. Results show a non-linear pattern between exam averages and the order in which the exams were turned in. Correlation was high for sections of courses taught by the same instructor. A Discussion and Conclusions section addresses these findings and discusses their significance.

Key Words: Exam scores, exam administration, undergraduate education.

Introduction

This paper reports the results of a study that explored whether the completion order of multiple-choice exams were correlated to their respective average scores. This study addressed an important aspect of college education. Exams are crucial tools for the assessment of a student's academic performance, and even small variations in their administration can have significant consequences for a student. From a student's perspective, exams are stressful events that not only determine a course grade, but also can make the difference for getting accepted to a college program or having a good resume for a job recruiter. From an instructor's viewpoint, exams are almost always the main instrument for gauging a student's mastery of a course contents, and therefore, for assigning an equitable grade to the student's performance. Exams are imperfect instruments to measure many of the skills relevant to a student's professional life (Stenberg, 2004, Foster, 2008). Despite their limitations, exams are likely to remain serving as the main means to measure a student's mastery of a course material, and therefore, every effort to understand their structure and administration is important.

The effect of external factors on exam scores (i.e., factors other than subject mastery) has been the focus of many studies. For example, students with positive perfectionist personality traits have been found to get better than average scores (Stoeber and Kersting, 2007). Cognitive abilities, i.e., innate skills such as working memory and processing speed, have been researched by Rohde and Thompson (2007), who found that mental processing speed can be a predictor of better academic grades. Stress coping strategies before and after an exam period have been analyzed by Folkman and Lazarus, (1985), who found that positive minded, problem solving-oriented coping mechanisms resulted in better GPAs. Another study (Connelly et al., 2005) found that students with slow handwriting are at a disadvantage during exams compared to faster writing classmates. Blake and Lesser (2006) explored the influence of middle school students' perception of self-sufficiency in their exam scores. Their study found that students with low self-appreciation of their ability to get good grades indeed got lower grades than their classmates with better sense of self-sufficiency. Haile and Anh Ngoc (2008) researched the influence that racial and family background factors may have on academic attainment, among other factors, and concluded that these two factors have less influence than commonly assumed. Llabre and Froman (1987) looked into the cultural differences between Hispanic and non-Hispanic students in their strategies to complete exams. They found that Hispanic students tend to take more time than non-Hispanic white students to complete the same exam, and therefore, imposing a time constraint may penalize the Hispanic examinees. This insight is reinforced by the work of Wild and Dorso (1979), who reported that providing more time to finish an exam ameliorates the differences among students of different cultural origins.

Few studies have been published about the connection between completion order (or completion time) and grades. Obligacion (2004) explored the relationship between time and grades for a single exam of an introductory course in sociology, finding that the best scores were found in exams taking more time than average to be completed. A personal website (League, 2008) plots the order of test completion against the grades obtained in an exam. It analyzes the final exam scores a 101-level course in computer science for non-majors and majors with limited prior experience. The graph shows a positive correlation between completion order and grades, with a coefficient of determination of 0.56.

Recently, some standardized, high stakes examinations have attempted to reduce the difference in grades due to the time required to respond a question. In the NCLEX-RN exam for nurse licensing, the time length of the examination is determined by each examinee's responses to test items (NCLEX, 2009). In effect, the duration of this computer-based exam is tailored to each student's strategy and style to take it.

Objectives

The purpose of this paper, as discussed in the Introduction, is to report the results of a study comparing average exam scores with the order in which the exams were completed. The potential correlation between these parameters, i.e., exam grades and completion order, is significant to students and instructors. From an instructor's perspective, it is important to know, for example, whether the last students to finish an exam tend to get high scores (and are just reviewing their answers) or not (consisting of students waiting for personal enlightenment, requiring more time than allotted for the test, or postponing the inevitable). There is anecdotal evidence that instructors do not have a uniform perception of which of the two scenarios is more common. Repeatedly announcing the remaining test time is less pressuring for students just reviewing their good exams than for those behind in the number of question answered. This research does not include normative aspects such as schemes to avoid or alleviate test administration practices hindering the performance of the latter type of students. However, it should be evident that without the objective data provided by this study, any new technique to improve the end-of-test period administration would be arbitrary, since the beneficiaries of the technique would be unknown. From a student's viewpoint, this study provides a straightforward metric for identification of poor test preparation and taking practices. If students with better grades tend to finish early, then a student that consistently finishes in the middle of the total number of exams raises a red flag that counselors can use to recognize that there is a problem (of course, other indicators also would have to be considered).

Methodology

The authors were instructors of the four college courses in the Construction Management program at Colorado State University involved in this study. Each course comprised two to three sections, and varied in size between 33 and 80 students. Course levels went from first year to seniors. Course details are shown in Table 1.

The tests were regularly scheduled midterm exams, and were not changed in their format, content or length. They consisted of multiple-choice questions, with one exception which included short-answer and multiple-choice segments. The multiple-choice format removes any difference in grading criteria among the instructors. Contrary to a common misconception, the score obtained in a multiple-choice exam is an excellent predictor of the performance of the student in a similar constructed response test (Lukhele and Thissen, 1994). A shortcoming of multiple-choice exams is that students lose the individualized comments that some instructors include as marks in graded exams, if returned to students. All authors of this paper conduct a discussion of key test questions in the session following the exam.

Each completed exam response card was marked with a sequential number as students turned it in. These numbers were later matched to the score obtained by the student for each course and section. To avoid the possibility of a breach in confidentiality, only student IDs were used to match each score to its completion order.

Table 1: Details of courses included in the study

Course Number	*Title*	*Section*	*Number of exams*	*Instructor (author initials)*
CON 151	Construction Materials and Methods	1	52	MCN
		2	53	MCN
		3	50	SAG
CON 265	Construction Estimating I	1	63	CLP
		2	80	CLP
CON 317	Safety Management	1	54	CLP
		2	61	CLP
CON 462	Financial Management for Construction	1	33	SAG
		2	47	BAS
		3	57	BAS

The scores obtained in the exams were generally consistent with those of previous semesters. The maximum absolute possible exam score differed among courses, and to compare results, scores were normalized by dividing each score by the respective exam average. If the average score for a given exam was, for example, 80 points, then a particular exam with a score of 80 was normalized to 80/80 = 1.00. An exam with a score of 90 would be normalized as 90/80 = 1.125. Moreover, scores were averaged for each five consecutive exams. All results refer to the behavior of these averages. The sample bin size of five exams provides a meaningful level of detail to the analysis while facilitating the detection of trends in the data.

Bivariate statistics were used to probe the correlation between the two variables of this study. A non-linear regression analysis was performed to each individual section sample and to the aggregate of all sections for each course, consisting of a third-order polynomial. Moreover, Pearson's correlation coefficient was computed for each section pair within each course. Finally, scatterplot diagrams were plotted for visually reinforcing the analytical analysis.

Results

The results of the analyses detailed above are shown in Tables 2, 3 and 4 and Figures 1 and 2. Table 2 shows the score range for each course, as well as the average range for the entire sample. Table 3 shows the coefficients of determination (R^2) for each section and the combined sections for each course. Table 4 consists of a correlation matrix summarizing the possible combinations for the sections within each course. Figure 1 contains scatterplot diagrams showing the sample data for each course with all combined sections, and Figure 2 shows a detail of results for CON462. The corresponding coefficient of determination is included in each figure.

Table 2: Score range

	Max	*Min*	*Range*
CON 151	1.09	0.94	0.15
CON 265	1.17	0.93	0.24
CON 317	1.08	0.94	0.14

CON 462	1.15	0.95	0.20
Averages	1.12	0.94	0.18

Table 3: Coefficients of Determination (R^2)

	Section 1	*Section 2*	*Section 3*	*Combined*
CON 151	0.7297	0.3562	0.6250	0.5475
CON 265	0.3882	0.5994		0.5252
CON 317	0.5017	0.3146		0.0601
CON 462	0.5633	0.7042	0.6165	0.8311

Table 4: Correlation Matrices Showing
Pearson's Correlation Coefficients

	Section 1	*Section 2*	*Section 3*
CON 151			
Section 1	1.00		
Section 2	-0.04	1.00	
Section 3	0.56	0.76	1.00
CON 265			
Section 1	1.00		
Section 2	-0.30	1.00	
CON 317			
Section 1	1.00		
Section 2	-0.47	1.00	
CON 462			
Section 1	1.00		
Section 2	0.04	1.00	
Section 3	0.04	0.43	1.00

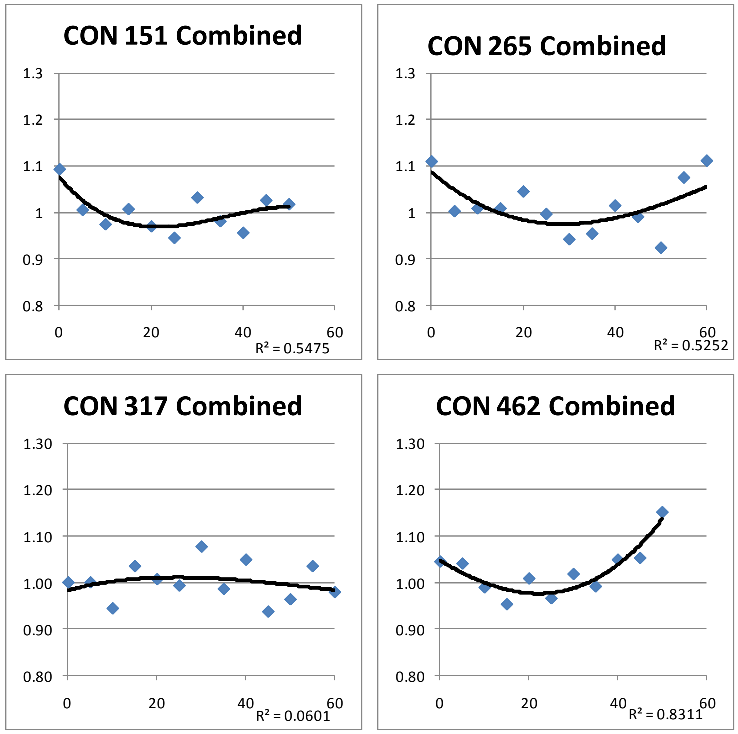

Figure 1: Scatterplot and trendline for combined sections of each course

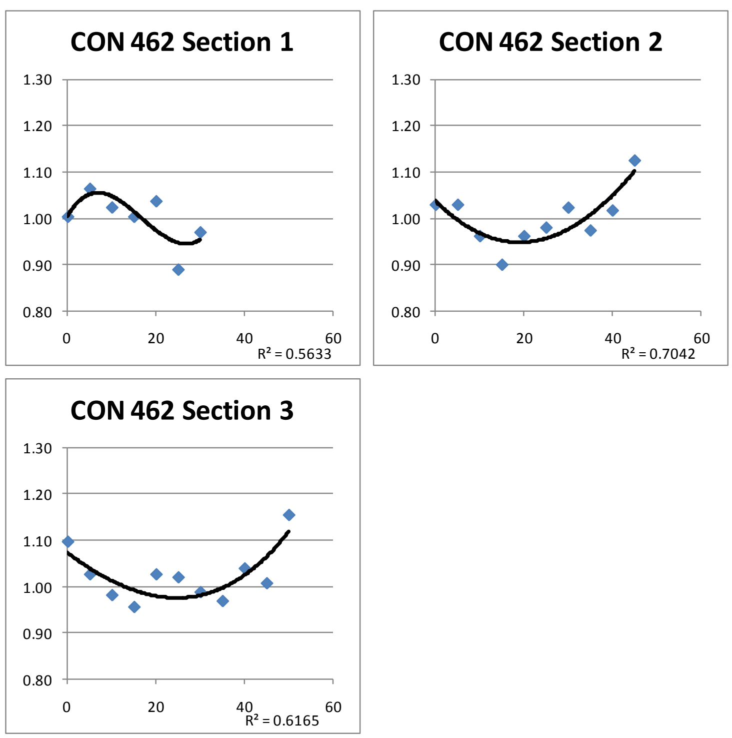

Figure 2: Scatterplot and trendline for each section of CON 462

The average difference between the maximum and minimum scores for each of the four courses was 0.18, or 18% of the normalized scores. This substantial range points to the significance of the analyzed variable (i.e., completion order) in the exam grades.

The strong coefficient of determination R^2 found in almost all the cases analyzed here suggests that there is an underlying student strategy to the order in which exams are completed, as discussed in the next section. However, an inspection of the correlation matrices in Table 4 and the scatterplots in shows that the correlation among sections can be much weaker. A closer examination of the correlation among sections reveals that the best correlations are found between sections taught by the same instructor. This behavior seems logical. The teaching style, details in class contents, and subtle differences in an exam's administration depend in each instructor's preferences, and therefore, are more similar for sections sharing their main instructor.

The best fitting was nonlinear. Specifically, a third-degree polynomial line was fitted in all cases. For the majority of the 10 sections, the curve was concave. The first exams turned in tended to have high scores, decreasing to a minimum after around 20 exams had been completed, and then showing a gradual uptick in average grades. The larger sections tended to have a plateau in grades towards the end. As can be seen in Figures 1 and 2, this general pattern differed somewhat among courses, with the exception, discussed next, of CON 317.

CON 317 was the one exception to the general completion-score patterns, not only by not having large differences in average scores, but also by displaying a convex curvature instead of a concave one as in the other courses. This course was the only one whose exam, for both sections, contained a short-answer segment additionally to a multiple-choice part. The peculiarities of this course may be related to this difference in exam format. Students may feel that they need to reserve time for the short-answer section, effectively overriding any strategy or behavior revealed by the analysis of the other courses. As discussed in the Conclusions, the unique pattern of this mixed-format exam merits further research.

Discussion and Conclusions

This study found a significant nonlinear correlation between the order in which completed exams were turned in and their scores. This correlation was strongest among sections of courses taught by the same instructor, but considerably smaller when the instructors were different. In particular, this study found that the best scores are those at the beginning and end of the exam, and the lowest ones tended to avoid these extremes.

This study had limitations that must be clearly indicated. The scope of this study made appropriate the use of bivariate statistics, and therefore, its results cannot be interpreted for the inference of cause and effect between the examined variables. The rationale for the use of multiple-choice exams was previously discussed. Despite the research (Lukhele and Thissen, 1994) strongly suggesting that the obtained results should hold for other type of tests (e.g., essay questions), this generalization is an inference not directly supported by the present data. Moreover, the results of this research could be affected by the subject being tested. Although courses in three disparate areas provided the data for this analysis, courses in other topics such as structural analysis could yield different results.

A comprehensive study of the causes and effects of external factors affecting exam performance would have to include the issues studied by previous research and discussed in the Introduction, such as exam's duration (Obligacion, 2009, Llabre and Froman, 1987, Wild and Dorso, 1979), student self-confidence (Blake and Lesser, 2006) and stress coping mechanisms (Folkman and Lazarus, 1985), among others. These factors were not part of this research's scope, but could influence its results.

Despite the above limitations, there are possibilities that merit speculative discussion, especially when they align with the personal experience of many instructors. The differences in grades vs. completion order could be due to deliberate student strategies for exam taking. Students first completing the exam may happen to have prepared well for the exam, and are confident about their knowledge of the covered material. Their completion speed could reflect that they do not need to spend much time reviewing their responses. Moreover, an important factor for the higher score obtained by the last completed exams could be that conscientious students tend to be the ones willing to spend more time making sure that all responses are the most logical option, and that there are no clerical mistakes in their answer card. In either of the scenarios above, higher scores would be the consequence of good strategies, such as thorough advance preparation and meticulous scrutiny of the questions.

It was mentioned in the Objectives that students with lower grades could benefit from recognizing that their approach for exam-taking may not be effective. There are many ways to squander an exam's time, or to use it inappropriately. In the authors' personal experience, some students obsessively attempt to respond a hard question before finishing the others, or mark the first choice that seems to be reasonable without attempting to discard the other choices. The authors (as many other educators) attempt to underscore the need to avoid these poor practices in each exam review session and in the discussion of its results. But, many of the students that could benefit the most of changing their exam taking strategies seem to not realize that they suffer from these poor habits. Including the completion order in each returned exam could help students with chronic lower grades in the realization that they have a problem. A consistent tendency to finish in the middle portion of the exam would be a simple and palpable warning sign that the instructor may point out, and which could be more effective to elicit a change in behavior from these students than a generic admonition.

The results of the present study show the need to improve current exam administration practices. An exam score frequently tells as much about a student's knowledge of the material covered as to the student's strategy to take exams efficiently. There are many factors that can hinder the intended purpose of any exam, which should be the true gauging of each student's mastery of the course material. This study can help in the improvement of classroom dynamics during an exam and the strategies followed by students to complete it.

References

Blake, S. and Lesser, L. (2006). Exploring the relationship between academic self-efficacy and middle school students' performance on a high-stakes mathematics test. *Proceedings of the 28th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Mérida, México

Connelly, V., Dockrell, J. and Barnett, J. (2005).The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology*, 25(1) 99-107.

Folkman, S. and Lazarus, R. (1985). If it changes it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology*, 48(1), 150-170.

Foster, C. (2008, May 23). If the answer is instant, the question is pointless. *The Times Educational Suppleme*nt, p. 33.

Haile, G. and Anh Ngoc, N. (2008). Determinants of academic attainment in the United States: A quantile regression analysis of test scores. *Education Economics*, 16(1), 29-57.

League, C. (2008, December 17) Completion time correlation. Retrieved *October 18, 2009 from Contrapunctus.net* website: http://contrapunctus.net/index.php/2008/completion-time-correlation

Lukhele, R., Thissen, D. and Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, 31(3) 234-250.

Llabre, M. and Froman, T. (1987). Allocation of time to test items: a study of ethnic differences. *The Journal of Experimental Education*, 55(3), 137-140.

NCLEX Test Breakdown (n.d.). Retrieved October 11, 2009 from TestPrepReview.com website, http://www.testprepreview.com/nclex_test_breakdown.htm.

Obligacion, F. R. (2004). Test-Taking Time as a Correlate of Student in Introductory Sociology. *Paper presented at the annual meeting of the American Sociological Association, San Francisco, CA*. Online <.PDF>. 2009-05-26 from http://www.allacademic.com/meta/p109661_index.html

Rohde, T. and Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35 83-92.

Stoeber, J. and Kersting, M. (2007). Perfectionism and aptitude test performance: Testees who strive for perfection achieve better test results. *Personality and Individual Differences*, 42 1093-1103.

Wild, C. and R. Dorso, R. (1979). *Effect of increased test-taking time on test scores by ethnic groups*. New Jersey: Educational Testing Service