Just Five More Minutes: The Relationship Between Timed and Untimed Performance on an Introductory Programming Exam

Dr. Matthew A. Verleger, Embry-Riddle Aeronautical University - Daytona Beach

Matthew Verleger is an Assistant Professor of Engineering Fundamentals at Embry-Riddle Aeronautical University in Daytona Beach, Florida. His research interests are focused on using action research methodologies to develop immediate, measurable improvements in classroom instruction and the use of Model-Eliciting Activities (MEAs) in teaching students about engineering problem solving. Dr. Verleger is an active member of ASEE. He also serves as the developer and site manager for the Model-Eliciting Activities Learning System (MEALearning.com), a site designed for implementing, managing, and researching MEAs in large classes.
Just 5 More Minutes – The Relationship Between Timed and Untimed Performance on an Introductory Programming Exam

Abstract
This evidence-based practice paper explores the relationship between the performance on an untimed exam and performance on other course metrics including later timed exams in an introductory computer programming course.

Background
Introduction to computer programming courses are often viewed as being exceptionally hard for most engineering students not explicitly pursuing careers as professional programmers. The combination of the breadth of material, the complexity of that material, and students’ relative unfamiliarity with the material makes it exceptionally difficult to give a proctored exam during a traditional class period. Students frequently complain that they understood the material but needed additional time to complete exams.

Purpose (Hypothesis)
The purpose of this research is to explore the relationship between time needed to complete the exam and overall course performance. The hypothesis was that additional exam time is of little relevance – students who know the material do better on the exam than students who don’t know the material as well, regardless of the amount of time available.

Design/Method
During the Fall 2015 semester, the instructor of an introductory programming course scheduled the first exam on a Saturday with the provision that students had unlimited time. Completion time for the exam was noted for each student. Due to an administrative request, the second and third exam were held during a 50-minute class period. A comparison of time taken on exam 1, all three exam scores and other course performance indicators (i.e., homework and final project scores, final course grades, course attendance) was conducted.

Results
Analysis revealed two major findings. First, the hypothesis was confirmed; there is no relationship between exam time and course or exam performance. Second, homework, more so than timed or untimed exams, was a better predictor of overall course performance.

Conclusion
The conclusion this research made was that the value of exams in a programming class should be rethought altogether. Isolated exams are a poor predictor of overall performance and the granting of unlimited time does not result in stronger alignment to course performance.

Introduction
Introduction to computer programming courses are often viewed as being exceptionally hard for most engineering students not explicitly pursuing careers as programmers. Intuitively, this makes sense, as learning computer programming generally requires the mastery of three independent components:
1. general problem solving - identifying inputs and outputs, formulating a procedure
2. general programming knowledge - sequential statements, decision structures, and looping structures as general algorithmic concepts
3. language specific syntax - where semi-colons go, where spacing or indentation is required, what function to call to get user input

The combination of the breadth of material, the complexity of that material, and students’ relative unfamiliarity with the material makes it exceptionally difficult to give a proctored exam during a traditional class period. At a small, private, STEM+Business only institution in the southeast, non-electrical and computer engineering students, as well as some non-engineering students, are required to take EGR 115 – Introduction to Computer Programming for Engineers during their first year. The course meets twice per week in a computer lab for 50 minutes each. The course utilizes a flipped/hybrid format with traditional instructional content being posted online in the form of YouTube videos and lab time being used for clarification of concepts and active learning exercises. In prior semesters, students frequently complained that they needed additional time to complete the three in-class exams.

During the Fall 2015 semester, the instructor scheduled the first exam on a Saturday with the benefit being that students had unlimited time. Due to an administrative request, the second and third exam were held during the 50-minute class period. The purpose of this paper is to compare performance on that first unlimited time exam with other course performance metrics to identify the validity of the claim that additional exam time is necessary for students to adequately demonstrate their mastery of the material.

Background

Few studies have explored the relationship between an exam time constraints and performance on complex tasks. Further, most of these studies examine timing in the context of how it effects specific sub-groups of students, not the general student population. What studies have been done have, unsurprisingly, almost universally demonstrated differences in timed versus untimed performance. Onwuegbuzie & Seaman (1995) found that untimed exam performance in a graduate level statistics course was statistically better than timed performance, but that untimed situations are more beneficial for students with high test-anxiety as compared to students with low test-anxiety. Lesaux, Pearson, & Siegel (2006) found that additional exam time helped compensate for learning disabilities on a reading comprehension task, bringing below average readers to average levels. Bergersen, Hannay, Sjøberg, Dybå, & Karahasanovic (2011) combined time and performance into a single metric for assessing programmer capability, basing that metric on the notion that programming speed is one essential component in a high quality programmer. This combination is acceptable for a professional context where speed is of greater importance but fits poorly to an academic context where being able to demonstrate knowledge mastery at all is more important than demonstrating that mastery quickly.

One major area where a better understanding of timing constraints may be useful is in exploring the implications of ADA accommodations. The ADA requires accommodations be made for students with disabilities (US Department of Justice, 2014). Often, accommodations come in the form of additional time on examinations and quizzes. While the ADA does not require
accommodations to be modified to the point of changing what is being measured, there is an open question for all instructors regarding if additional time fundamentally changes the examination process in their context. Zuriff (2000) critically examined five studies whose focus was on accommodation through additional time and raised doubts as to the fairness of the accommodation.

The goal of this study to explore student performance on timed versus untimed exams in this context to better understand how timing relates to other measures of student performance.

**Methodology**

*Participants*
EGR 115 serves both engineers and non-engineers. For engineers, it is a required course typically taken during the first year. Because the course is not required for students in the Electrical/Computer Engineering discipline, the majority of these students do not take the course with the intention of being professional programmers. For non-engineers, the course is either a technical elective or is required but not a strict pre-requisite for their discipline courses, resulting in students from non-engineering disciplines taking it throughout their curriculum. The enrollment during the Fall 2015 semester (n=69) was 78% engineers and 22% non-engineers. It had 20% female enrollment. Both of these metrics are consistent with previous semesters of the course. No students provided documentation requesting ADA accommodations at any point in the semester.

*Exam Structure*
The first exam was scheduled for a Saturday. Students were told they had unlimited time and that the exam would be “open-book, open-note, open-resource, open-internet, closed other people”. The exam consisted of 10 multiple choice theory questions (20% of the exam score) and a MATLAB programming question (included in the appendix – 80% of the exam score). Prior to the exam date, students were required to sign up for an approximate start time. 25 seats were made available starting at 9:00 AM and 5 seats were then released every 15 minutes starting at 10:15 AM until enough seats were released for all students. On the exam day, students arrived and took a number, in the form of a Post-It note on the wall outside the classroom. As students entered the exam room, the current time was noted on the bottom of their exam and again when they completed their exam. This provided a net time spent for each student. As students left, the next number was allowed to enter the room.

Because of an administrative request, the second and third exams were held during the 50-minute class period. The second exam was held during one class period and still contained a multiple choice and programming portion. The third exam was split across two days, with the multiple choice and a small programming portion being done on the first day and a larger programming portion being done on the second day. Both exams were still open-resource. No students finished the second or third exam parts early.

*Course Components*
Course grades are calculated using the percentages listed in Table 1.
Table 1. Course Grade Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1 Total</td>
<td>13.33%</td>
</tr>
<tr>
<td>Exam 2 Total</td>
<td>13.33%</td>
</tr>
<tr>
<td>Exam 3 Total</td>
<td>13.33%</td>
</tr>
<tr>
<td>Attendance</td>
<td>0%</td>
</tr>
<tr>
<td>Exercises</td>
<td>10%</td>
</tr>
<tr>
<td>Homework</td>
<td>20%</td>
</tr>
<tr>
<td>Participation</td>
<td>10%</td>
</tr>
<tr>
<td>Final Project</td>
<td>20%</td>
</tr>
</tbody>
</table>

- Exams are equally weighted at 40% total.
- Attendance at the lab is not required students are told on the first day of class that it is collected and is used as a weighting factor for any curve applied to final course grades. Attendance was taken 16 times between August 28th and November 6th.
- Exercises are conducted using Mathworks Cody Coursework system. They consist of 5-6 problems per assignment and are results-based – does your code calculate the correct answer. Because the system allows for automated grading, students are allowed to submit their assignments late for 50% credit. An average of 27 equally weighted problems were used to calculate the exercise grade.
- Homework assignments consist of 1-2 problems per assignment and are more complex, providing greater context than the exercises and generally requiring more code to be developed. Because these are graded by hand, students are not allowed to submit them late. 13 equally weighted assignments was used to calculate the homework grade.
- Participation is based on students watching the instructional videos and taking the associated video quizzes for each video. Quizzes could be taken up to 3 times with the maximum score counting toward their grade.
- The final project is a large, individual, self-selected program graded against a standardized rubric. Students typically create games or databases designed to demonstrate all of the major course topics. The rubric is split into 33% program design (e.g., use of comments, indentation, variable naming, etc.), 33% demonstration of course material (e.g., presence and usage of FOR loops, programmer-defined functions, etc.), and 33% project-specific functionality focused on how well the achieved the purpose of the problem they selected.

Of these components, only the exams are taken in isolation, meaning that all other grade components have significant external assistance available. The underlying assumption governing this work is that, because that external assistance is largely provided through open channels (e.g., campus tutoring center, faculty office hours), all students are equally capable of benefiting from that assistance, even if not all students partake of that assistance.

Results

Spearman Rho correlation coefficients were calculated between each of the grade components of the course as well as the time taken on Exam 1. These results are shown in Table 2.
Table 2. Correlation of Course Components (n=69)

<table>
<thead>
<tr>
<th></th>
<th>Exam 1 Time (Minutes)</th>
<th>Exam 1 Total</th>
<th>Exam 2 Total</th>
<th>Exam 3 Total</th>
<th>Attendance</th>
<th>Exercises</th>
<th>Homework</th>
<th>Participation</th>
<th>Final Project</th>
<th>Final Course Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1 Total</td>
<td></td>
<td>-0.029</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 2 Total</td>
<td></td>
<td>-0.309</td>
<td>0.480</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 3 Total</td>
<td></td>
<td>-0.071</td>
<td>0.432</td>
<td>0.397</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td></td>
<td>0.096</td>
<td>-0.124</td>
<td>0.082</td>
<td>0.259</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td>0.254</td>
<td>0.135</td>
<td>0.121</td>
<td>0.270</td>
<td>0.418</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homework</td>
<td></td>
<td>0.167</td>
<td>0.235</td>
<td>0.398</td>
<td>0.513</td>
<td>0.443</td>
<td>0.726</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation</td>
<td></td>
<td>0.171</td>
<td>0.176</td>
<td>0.086</td>
<td>0.254</td>
<td>0.356</td>
<td>0.545</td>
<td>0.648</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Project</td>
<td></td>
<td>0.174</td>
<td>0.167</td>
<td>0.290</td>
<td>0.613</td>
<td>0.361</td>
<td>0.457</td>
<td>0.660</td>
<td>0.332</td>
<td></td>
</tr>
<tr>
<td>Final Course Percentage</td>
<td></td>
<td>0.100</td>
<td>0.417</td>
<td>0.549</td>
<td>0.678</td>
<td>0.397</td>
<td>0.647</td>
<td>0.902</td>
<td>0.573</td>
<td>0.801</td>
</tr>
</tbody>
</table>

Correlation is significant at the 0.01 level (2-tailed).
Correlation is significant at the 0.05 level (2-tailed).

Further subdivision was done on the homework scores versus the exam scores by taking only the homework assignments specifically targeted to each exam and calculating an average score. For example, all of the homework assignments that occurred between Exam 1 and Exam 2 were averaged and that average was correlated to the Exam 2 score. The homework set scores were the correlated against the corresponding exam scores using Spearman Rho correlations. The results of this correlation are in Table 3.

Table 3. Exam Scores vs. Corresponding Homework Sets (n=69)

<table>
<thead>
<tr>
<th></th>
<th>Exam 1</th>
<th>Exam 2</th>
<th>Exam 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.278</td>
<td>0.308</td>
<td>0.375</td>
</tr>
<tr>
<td>Sig.</td>
<td>0.021</td>
<td>0.010</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Analysis

Exam 1 versus Time
As Figure 1 shows, there is no meaningful relationship between time spent and exam grade. The Spearman’s Rho correlation was nearly 0 and statistically insignificant (ρ = -0.029, p = 0.811). While students would generally argue that additional time would result in a higher exam score, they typically fail to account for similar increases in their peers scores. Students are also unlikely to dramatically overcome exam unpreparedness with additional time. Students who don’t know the course material won’t magically learn it during the exam if given the opportunity. The exam average was 78.33 and was consistent with the average Exam 1 grade in prior semesters.
**Figure 1. Time Spent vs. Exam 1 Score**

*Exam 1 vs. Exam 2 vs. Exam 3 Scores*

As noted in Table 2, there is a moderate and statistically significant correlation between the three exam scores. This can also be seen in Figure 2, where students who took all three exams (n=60) were ranked by their exam score and divided into quartiles. Their transition from quartile to quartile across the three exams was plotted as a Sankey diagram. The implication of this is that, even with the extra time give during Exam 1, students who do well without a time limit (Exam 1) are likely to do similarly well when a time limit is imposed (Exams 2 & 3), while students who do poorly without a time limit (Exam 1) will likely also do poorly with a time limit (Exams 2 & 3).
Exam Score vs. Exam 1 Time
An interesting negative correlation appears between students’ Exam 2 score and the time taken on Exam 1 ($\rho = -0.309, p = 0.010$). Students who took more time on Exam 1 scored more poorly on Exam 2 than those that finished Exam 1 more quickly. This result is slightly unexpected given the lack of relationship between Exam 1 time and Exam 1 score, and the stronger relationship between Exam 1 and Exam 2 scores. The best explanation the author can make is that perhaps students who took longer on the first exam were less prepared to handle the shorter time constraint of Exam 2 and were therefore likely to perform worse. Only one student completed the first exam in the 50 minutes given for Exam 2, so nearly all students had to adapt to the shorter time, but those who needed more time for Exam 1 adapted more poorly.

Exam 1 & 2 Scores vs. Other Course Components
One of the most intriguing issues that the data highlights is the general lack of relationship between Exam 1 and 2 scores versus the other complete course components. It would generally be expected that students who do well on the other course components (i.e., homework, lab attendance, etc.) should do better on the Exams throughout the course and vice-versa. In contrast to this is the moderate but statistically significant correlations between the exam scores and their corresponding homework sets shown in Table 3. The implication would be that, while the exams do moderately test the corresponding content they are supposed to cover, they are not as cleanly matched to the course content as a whole.

Homework vs. Final Grades
By far the most intriguing finding is the very strong relationship between homework scores and final course grades. While the exams count for a higher percentage of the grade (40% vs. homework’s 20%), homework scores clearly predict overall course performance much better than exams. The existence of a statistically significant relationship is not surprising, but the extent of the relationship is and calls into question the value of the other course components within the course.
Conclusions and Next Steps

There are two primary conclusions from this research. First is that there is no meaningful relationship between giving students additional time and raising exam scores. As long as the testing conditions are fair for all students, there is no reason to believe that additional exam time will result in drastically better performance. While there may be other reasons to give longer exams, such as reduced exam anxiety or better course evaluations because students perceive the additional time as a benefit, the argument that students will generally perform significantly better by getting a longer exam is false.

Second, the instructor should begin questioning the value of the exams as assessment tools in this context. The high-stakes tests in this microcosm do not align well with other course performance indicators. While the results stop short of recommending that exams be removed all together, they do raise the issue of if, as they are currently structured, they are having the desired assessment impact. One alternative structure might involve increasing the value of homework and the final project on the course grade and reducing the perceived value of the exams. This would have the benefit of reducing complaints about exam difficulty (because they would not be as detrimental to overall course grades) without drastically changing final course grades. As a more general recommendation, the author encourages the reader to conduct similar analysis on their own course to explore how well exams align with other course components. Analysis like this may become an instigator for course-level curricular reform.

References


Appendix – Exam 1 Programming Problem

Houses in Florida often use asphalt shingles on the roof to protect against the weather. Shingles are rectangular sheets and are purchased in bundles, typically containing 20 shingles per bundle. A new community is being developed and the contractor has hired you to write a MATLAB program to help her determine the number of bundles to purchase for homes in the community.

Attached is an isometric drawing of a standard house plan in the community. Each house is shaped like a T, with a small front section of the house (CxD in the drawing) centered along the larger main section of the house (AxB in the drawing). Your task is to prompt the user for the five basic dimensions of the house (noted by the letters in the drawing):

• A – Depth of the main body of the house in feet
• B – Width of the main body of the house in feet
• C – Depth of the front part of the house in feet
• D – Width of the front part of the house in feet
• E – Angle of the roof in degrees

You should then ask the user for the number of square feet each bundle covers. You should randomly select a valid price for each bundle between $20.00 and $40.00 (inclusive) and calculate the cost for the shingles needed for the house. You should then print a properly aligned list similar to the one provided summarizing the areas of the roof and the total cost.
Required Input/Output (with user entered values in bold)

A) Depth of Main House Body (ft): 20
B) Width of Main House Body (ft): 40
C) Depth of Front House Body (ft): 10
D) Width of Front House Body (ft): 15
E) Roof Angle (degrees): 30

Square Feet Per Bundle: 32.8

Area of Main Building to be shingled (AxB): 858.81 sqft
Area of Front Building to be shingled (CxD): 238.16 sqft

Cost Per Bundle: $21.95
Bundles Needed: 34 Bundles
Total Cost: $746.30

Detailed Input/Output to help you check your solution

A) Depth of Main House Body (ft): 20
B) Width of Main House Body (ft): 40
C) Depth of Front House Body (ft): 10
D) Width of Front House Body (ft): 15
E) Roof Angle (degrees): 30

Square Feet Per Bundle: 32.8

Area of Main Building to be shingled (AxB): 858.81 sqft
Area of Front Building to be shingled (CxD): 238.16 sqft
Area of Main Building not being shingled: 64.95 sqft
Extra Area of Front Building being shingled: 64.95 sqft

Cost Per Bundle: $21.95
Bundles Needed: 34 Bundles
Total Cost: $746.30