ON THE EFFECT OF THE ORDERING OF QUESTIONS IN EXAMS — A VISUAL ANALYSIS

Jürgen Symanzik*, Utah State University,
Robert Erbacher, Utah State University,
Palyne Gaenir, CyberGnostics, Inc.,
Natascha Vukasinovic, Utah State University,
Alex Cheng Ki Wun, Hong Kong.
*Department of Mathematics and Statistics, Logan, UT 84322–3900,
e–mail: symanzik@math.usu.edu

Key Words: CyberStats, Electronic Textbook, Long–Distance Course, Question Order Effect, Student Behavior, Teaching.

Abstract

The effect of the ordering of questions in sample surveys has been well investigated. However, there has been no study so far that examined the effect of the ordering of exam questions on the exam behavior of students. In this paper, we will discuss the effect of the ordering of questions on the student behavior for an introductory long–distance statistics course. Because we were working with the electronic textbook CyberStats that records exact submission times of student answers to exam questions, the following research question could be answered: Do students usually answer exam questions in sequential order (independent from the exam questions) or do they tend to jump from one exam question to another while leaving some types of (apparentely more difficult) exam questions until the very end? As it turned out, many students answer exam questions in almost perfect sequential order. This may easily lead to the result that those students will miss simple exam questions towards the end of an exam when there is a tight time limit for the students.

1. Introduction

In Stat 1040, a course regularly offered at Utah State University, a course coordinator prepares a common final exam for all sections. In previous years, the questions in the exam were arranged by topic (easier material discussed early in the semester being assessed first, and more difficult material discussed late in the semester being assessed towards the end of the exam). To prevent students from copying from each other, two versions of the exam had been handed out, with slightly different numerical values but otherwise identical. In a recent offering of the exam in Fall 2003, a new course coordinator modified the exam structure. Instead of using slightly different numerical values but the same question order, two versions of the exam were given, with exactly identical questions but two different random question orders. One of the exams appeared far more difficult than the other exam because several of the more difficult questions were asked early in this version of the exam. However, a statistical test (based on the outcome in one of the sections of this course) did not reveal any significant difference in the performance of students that were given the apparently simpler version of the exam compared to those that were given the apparently more difficult version of the exam (due to the ordering).

In the Spring 2004 semester, the International Program at Utah State University offered for the first time an extension course on Stat 2000 for its international students in Hong Kong. The course was based on the electronic textbook CyberStats (http://cyberk.com). CyberStats has been developed since the late 1990ies, with Editor in Chief Jessica Utts from the University of California at Davis. It is distributed by CyberGnostics, Inc. CyberStats has been used at more than 50 universities; more than 600 introductory Statistics courses have been offered based on CyberStats, mostly throughout the US and Canada. Reviews of CyberStats and comparisons with other electronic textbooks can be found in Dear (2001) and Symanzik & Vukasinovic (2003). Teaching experiences with CyberStats for campus–based courses have been reported in Symanzik & Vukasinovic (2002) and Utts, Sommer, Acredolo, Maher & Matthews (2003).

CyberStats has build–in testing facilities. An instructor can select n (say 100) questions overall and specify that each student has to answer m (say 30) questions in a particular exam. CyberStats then randomly selects m questions for each student and presents them in a random order to the students.

It can easily happen that two students do not get any question in common or some get easy questions at the beginning while others get difficult questions at the beginning of their exams. For our study, we modified this random question selection mechanism by using a fixed set of questions for all students and by placing constraints on possible question orders.

CyberStats is set up to record all student input — from answers to self–assessment questions to answers to homework exercises. All student answers are recorded — even if an answer is changed, the previous answers are still visibly for the instructor and authorized TAs. Some other variables of interest that are recorded are the times of submission and the Internet Protocol (IP) address of the computer from where an answer has been submitted.

In the remainder of this paper, we discuss our research questions and possible outcomes in Section 2. We describe the exam structure for this Stat 2000 course in Section 3. A visual analysis of our data follows in Section 4. We finish with our conclusions and an outlook on our future work in Section 5.

2. Research Questions and Possible Outcomes

From the literature on sample surveys, e.g., Perreault Jr. (1975), Sigelman (1981), McFarland (1981), Crespie & Morris (1984), McClendon & O'Brien (1988), and Drew (1991), it is well known that there exists a question order effect. This means, in experiments where the order of survey questions was exchanged, researchers found significant differences how the same question was answered — based on whether this question was asked toward the beginning or towards the end of a sample survey. Based on this knowledge and the outcome of the Stat 1040 exam with different question orders, one might be wondering about the following research questions:

(i) In which order do students answer the questions, i.e., do students prefer some type of exam questions and answer this type first and initially skip other types (that seem to appear more difficult)?

(ii) Do students speed up towards the end of the exam period?

(iii) Do students have a preference when to answer multiple–choice (MC) and text–based answers (TA) questions?

(iv) How often do students change their answers? Do they more often end up with a correct answer at the end or more often change a previously correct answer to an incorrect answer?

(v) Does the presented order of exam questions make any difference on the performance of students?

One might speculate that some of the following outcomes may be observed:

(i) Students prefer one type of exam question (e.g., MC) over another type (e.g., TA) and answer those preferred questions first.

(ii) If exam questions are presented in a random order (e.g., MC and TA questions randomly mixed), students may work sequentially through the questions; however, if there are two blocks of questions of the same type, students may answer questions of one type (e.g., MC) first.

(iii) There may be some association between the overall performance and the order in which exam questions are presented (although there was no such effect when comparing the two random question orders of the Stat 1040 course).

In this paper, we will concentrate on research questions (i), (ii), and (iii) only. The remaining two questions will be addressed elsewhere. Therefore, we do not want to reveal yet whether there is any association between overall performance and the question order.

3. Exam Structure

Instead of assigning random question sequences (as is the default in CyberStats) to each student, we worked with the CyberStats developers to be able to assign specific random question order sequences that follow particular constraints for each student. We dealt with a fixed set (e.g., 25 questions for Midterm 1) of MC questions and of TA questions for each exam. In each exam, each student was assigned to one of these three types of sequences of questions:

- `[1]` first MC, then TA questions
- `[2]` first TA, then MC questions
- `[3]` order of MC and TA questions fully randomized

Within each sequence type, the order of the questions of the same type was further individually randomized for each student.

We were running three exams (2 midterms, 1 final) throughout the semester. Each student was assigned a different sequence type in each exam, e.g., type `[3]` in Midterm 1, type `[1]` in Midterm 2, and type `[2]`

in the Final. There are 3! = 6 possible permutations of these types of sequences of questions over the three exams. With 57 to 59 students participating in the three exams, about 9 to 10 students were randomly assigned to each of these six permutations. Thus, we are considering a cross–over design with all orders of treatments (i.e., type of sequences of questions) present in the study design. If there is indeed an advantage of having questions in a particular order (e.g., a block of MC questions), each student enjoyed this advantage in one of the three exams.

Midterm 1 consisted of 25 questions: 15 MC and 10 TA questions. 59 students participated and had 60 minutes to answer all questions. Midterm 2 also consisted of 25 questions: 15 MC and 10 TA questions. 57 students participated and had 60 minutes to answer all questions. The Final consisted of 50 questions: 25 MC and 25 TA questions. 58 students participated and had 180 minutes to answer all questions.

Students had access to all 25 questions in the two midterms and to all 50 questions in the final exam via a single html page. Students were allowed (and encouraged) to answer questions in any order. The introduction for Midterm 1 stated: "You have **exactly 60 minutes** to complete the exam. Any items submitted after 60 minutes will not be graded. Try to correctly answer as many questions as possible during this time period. **You are allowed to answer questions in any order.** Start with a question that seems the easiest for you. If you cannot answer a question within a short time, move to another question, and come back to the previously unanswered questions toward the end of the exam." All exams (and the corresponding solutions) can be accessed at `http://www.math.usu.edu/~symanzik/teaching/2004_stat2000/stat2000_index.html`.

To guarantee that an answer got recorded at the CyberStats site, students had to click on a *Submit* button for each answer. Answers not confirmed via *Submit* were lost in case of a computer crash or the (unintentional) closing of the Web browser window that contains the questions and answer fields. Students were frequently reminded before and during the exams to click on *Submit* after each answer.

The data that eventually got recorded electronically at the CyberStats site consisted of about 16 variables, most importantly the student ID, all given answers (even if a student changed an answer, all previous answers remained in the data base), and the time each answer got submitted (down to seconds). Additional data recorded for Midterm 2 and the Final was the time the students first accessed the exam (unfortunately this time was not recorded for Midterm 1) and the time students confirmed that they had completed the exam. In Midterm 1, we had to make the assumption that the time of the first answer being recorded equals the time a student first accessed the exam. We will further discuss on the validity of this assumption.

There was some initial concern whether students participating in this course should be asked to give their consent or not to have their exam answers being recorded for purposes beyond determining the students' grades. But, by default, when enrolling in a course based on CyberStats, students are aware that homework submissions and exam submissions are stored electronically at the CyberStats site. While in a medical study where some of the patients receiving a placebo just react because they think they are being treated with a real treatment, the situation might be different here. Assuming that students often react negatively to exams, just the idea that there might be a study going on related to exams might have an adverse effect on such students. Therefore, it was decided to conduct this study double–blindly, i.e., neither the students nor the local instructor in Hong Kong who interacted with the students were told that there was a study going on related to the order of exam questions. Other than working on exams with a question order type randomly assigned as part of the study, there was no further intervention with the students.

4. Visual Analysis

Figure 1 shows the answer times for each question for the first twelve students recorded in the CyberStats data base for Midterm 1. Questions are displayed in relative order (REL Question) as they were visible to these students. Students 2, 3, 5, 7, 8, 9, 11, and 12 apparently had sequence type `[1]`, i.e., the 15 MC questions first, followed by the 10 TA questions. Student 6 had sequence type `[2]`, i.e., the 10 TA questions first, followed by the 15 MC questions. Students 1, 4, and 10 had sequence type `[3]`, i.e., MC and TA questions in fully randomized order. A few patterns are obvious: Eight of these twelve students (1, 2, 6, 7, 8, 10, 11, and 12) used a mostly sequential question answering method with a few questions possibly being skipped early in the exam. Almost all students corrected some of their answers at some point. Students 6, 10, 11, and 12 wanted to be certain that their answers got recorded and resubmitted their answers towards the end of the 60 minutes exam period. Student 3 ignored the exam instructions and did not click on *Submit* after each question. Instead, this student solved all ques-

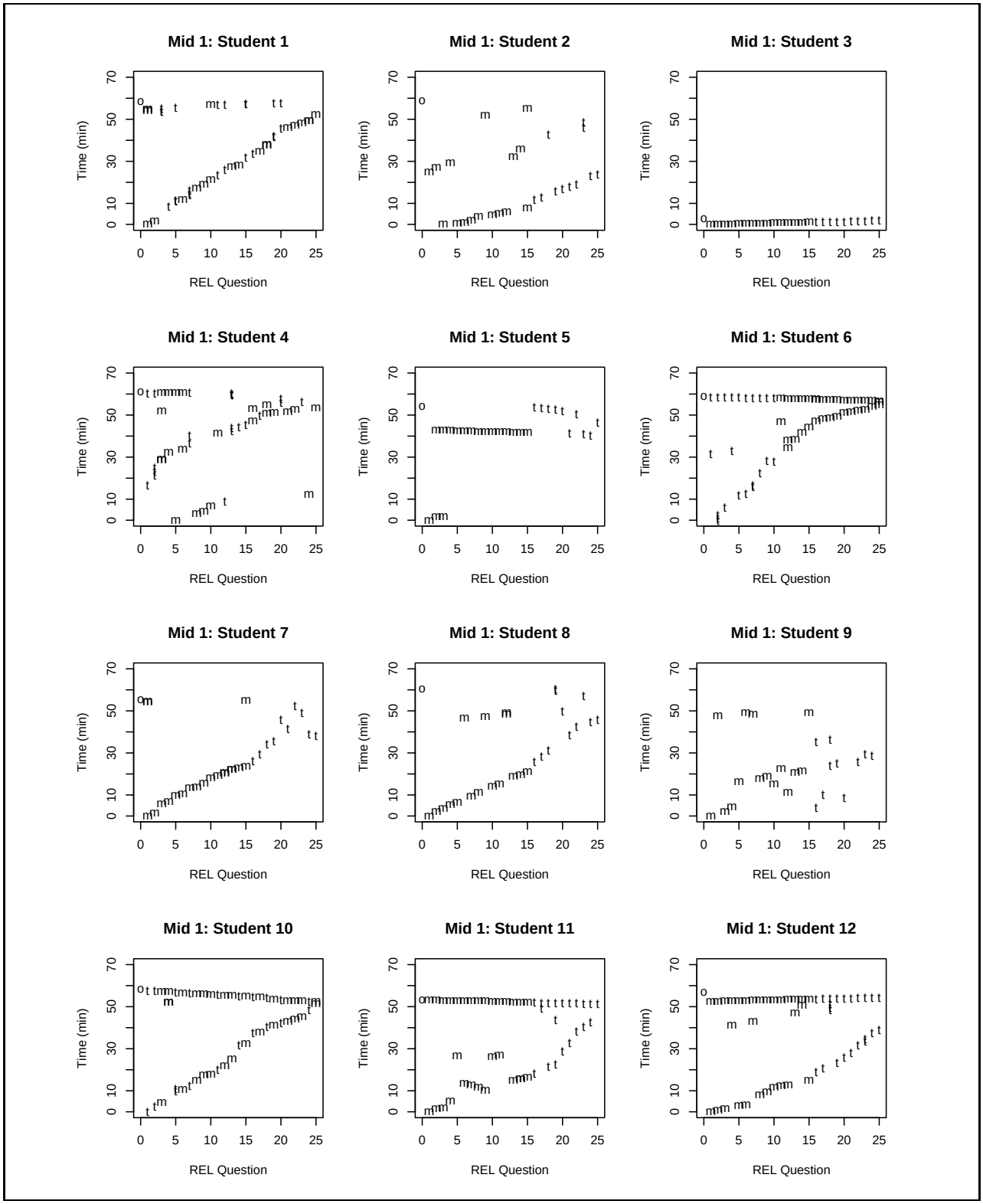

Figure 1: Answer times for 12 students for Midterm 1. m represents a MC question, t a TA question, and o any other entry (such as begin/end exam). REL Question represents the order of questions as seen by the students.

tions (in any order) and submitted only the final answers within a five minutes time interval. This, and similar data, should be discarded from the data set for a future numerical analysis.

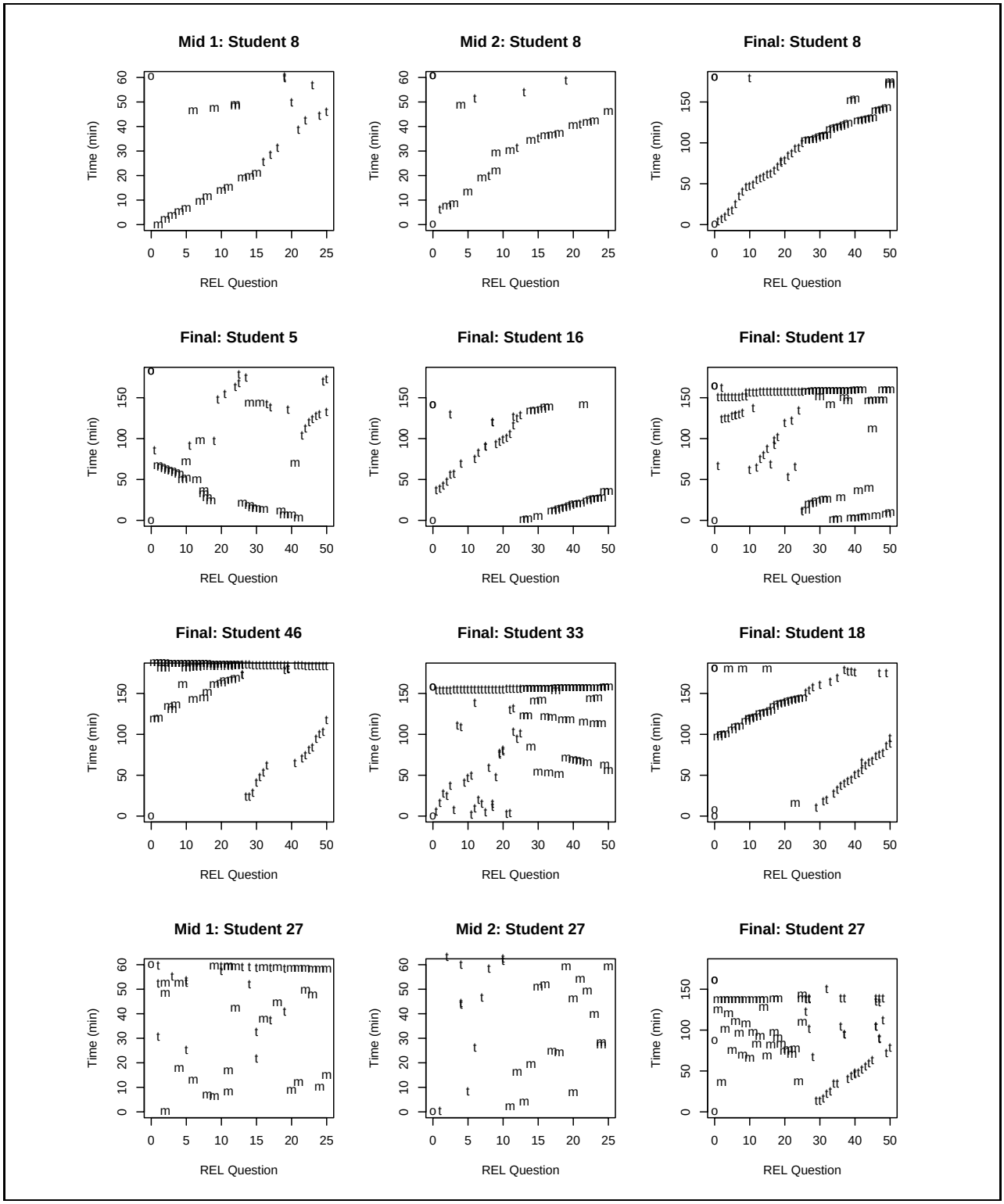

Figure 2: Answer times of selected students for Midterm 1, Midterm 2, and the Final. m represents a MC question, t a TA question, and o any other entry (such as begin/end exam). REL Question represents the order of questions as seen by the students.

Figure 2 shows the answer times for each question for selected students for Midterm 1, Midterm 2, and the Final. Questions are displayed in relative order (REL Question) as they were visible to these stu-

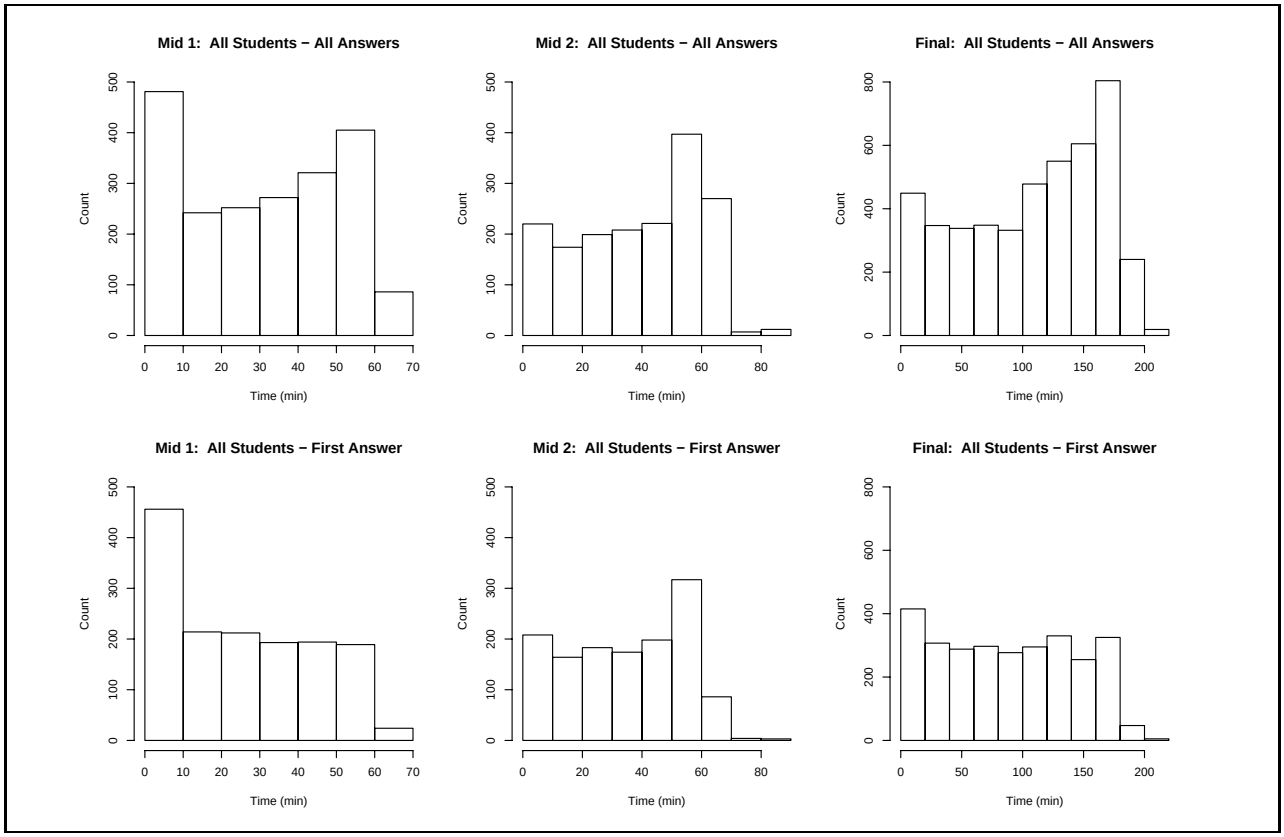

Figure 3: Number of answers per 10 minutes (Midterm 1 and Midterm 2) and 20 minutes (the Final) time interval of all answer submissions (top) and first answer submissions (bottom) for all students.

dents. The first row shows a sequential answering behavior for Student 8 across all three exams — although this student had the MC questions first in Midterm 1 (i.e., sequence type `[1]`, a fully randomized question order in Midterm 2 (i.e., type `[3]`), and the TA questions first in the Final (i.e., type `[2]`). The second row shows three students (5, 16, and 17) who chose to answer MC questions first in the Final — most notabely, none of these students had sequence type `[1]`. The third row shows three students (46, 33, and 18) who chose to answer TA questions first in the Final. The last row shows a random answering behavior for Student 27 across the three exams.

Figure 3 shows the number of all answers per 10 minutes (for Midterm 1 and Midterm 2) and 20 minutes (for the Final) time interval. While the total number of submission increases towards the end of the exam period, the number of first answer submissions to a question over a time interval remains almost constant in Midterm 1 and the Final (although there are more submissions towards the end of Midterm 2). All students were actually allowed to work a few minutes beyond the 60 (180) minutes time limits. Due to technical problems, a few students were allowed up to 100 (220) minutes in Midterm 2 (the Final). The high peak in the first 10 minutes time interval for Midterm 1 are the answers of those students that ignored to click on *Submit* after each question. In Midterm 2 and the Final, the first 10 minutes time interval contains about 60 *Begin Exam* entries, i.e., one for each student.

Figure 4 shows boxplots of the first answer submission times for each of the 25 questions ($X2$ through $X26$) and the *End Exam* control entry ($X0$) for Midterm 1. Apparently, about seven students who submitted an *End Exam* within the first 20 minutes seemed to ignore the instructions to click on *Submit* after each question. Nevertheless, similar plots for Midterm 2 and the Final (now with a *Begin Exam* entry recorded at 0 seconds of each exam) showed a similar pattern.

Figure 5 shows boxplots of the median first answer submission times, split by MC and TA questions for Midterm 1 (left), Midterm 2 (center), and the Final (right). In each exam, median times of first submissions to MC questions are observed much earlier than median times of first submissions to TA questions.

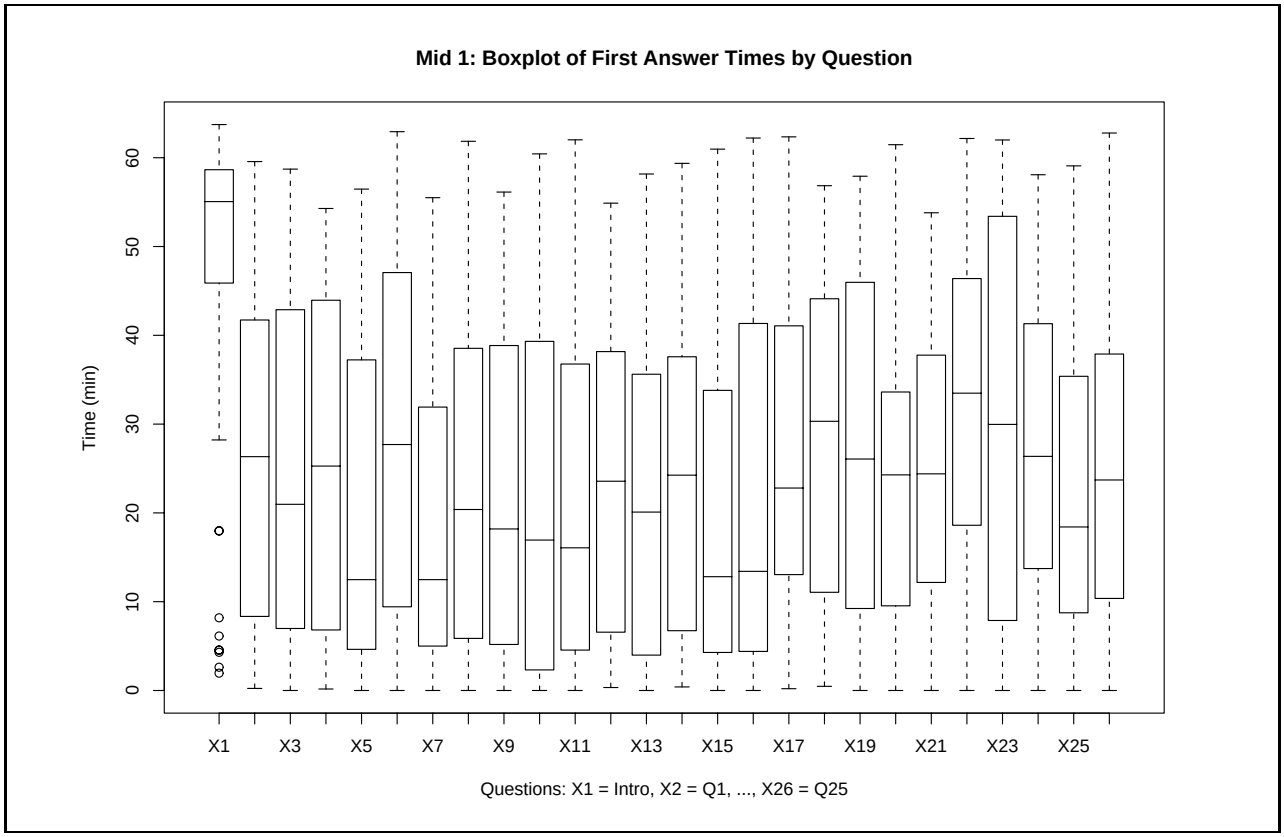

Figure 4: Boxplots of first answer submission times for each of the 25 questions ($X2$ through $X26$) of Midterm 1 and control entries ($X1$). For example, half of the students had submitted an answer to MC question 4 ($X5$) after about 12 minutes, while it took about 33 minutes before half of the students had submitted an answer to TA question 21 ($X22$).

5. Conclusions and Future Work

The results of this study may be beneficial for all instructors who are setting up exam questions: Should an exam start with simple questions or does the order of questions have no effect on the exam outcome? Let us revisit our earlier research questions.

Question (i) asked "In which order do students answer the questions ...?" Based on Figure 2, the answer is that there are different student "types", i.e., students who answer questions mostly sequentially, students who answer MC questions first, students who answer TA questions first, and students where the answer order seems to be random for an outside observer.

Question (ii) asked "Do students speed up towards the end of the exam period?" The data showed (see Figure 3) that students frequently change (or resubmit) answers towards the end of an exam. However, there was no noticeable increase in the number of first answers submitted during the last 10 minutes in Midterm 1 and the last 20 minutes in the Final, but there was some increase in the number of first answers submitted during the last 10 minutes in Midterm 2.

Question (iii) asked "Do students have a preference when to answer MC and TA questions?" The data for all three exams showed (see Figure 5) that the medians of the times MC questions got answered first are considerably smaller than the medians of the times TA questions got answered first. This indicates a clear preference to answer MC questions earlier in an exam and TA questions later in an exam.

What are the implications for an instructor when determining the order of questions in an exam? Many students will pick their preferred questions (most frequently MC questions) first or randomly answer questions through the exam. For those students, the given question order will have no major effect on the order in which those students answer the exam questions. Otherwise, students that answer questions mostly sequentially may never reach the questions at the end of an exam when there is a major time constraint for an exam. To better help those students to pass a single exam or an entire

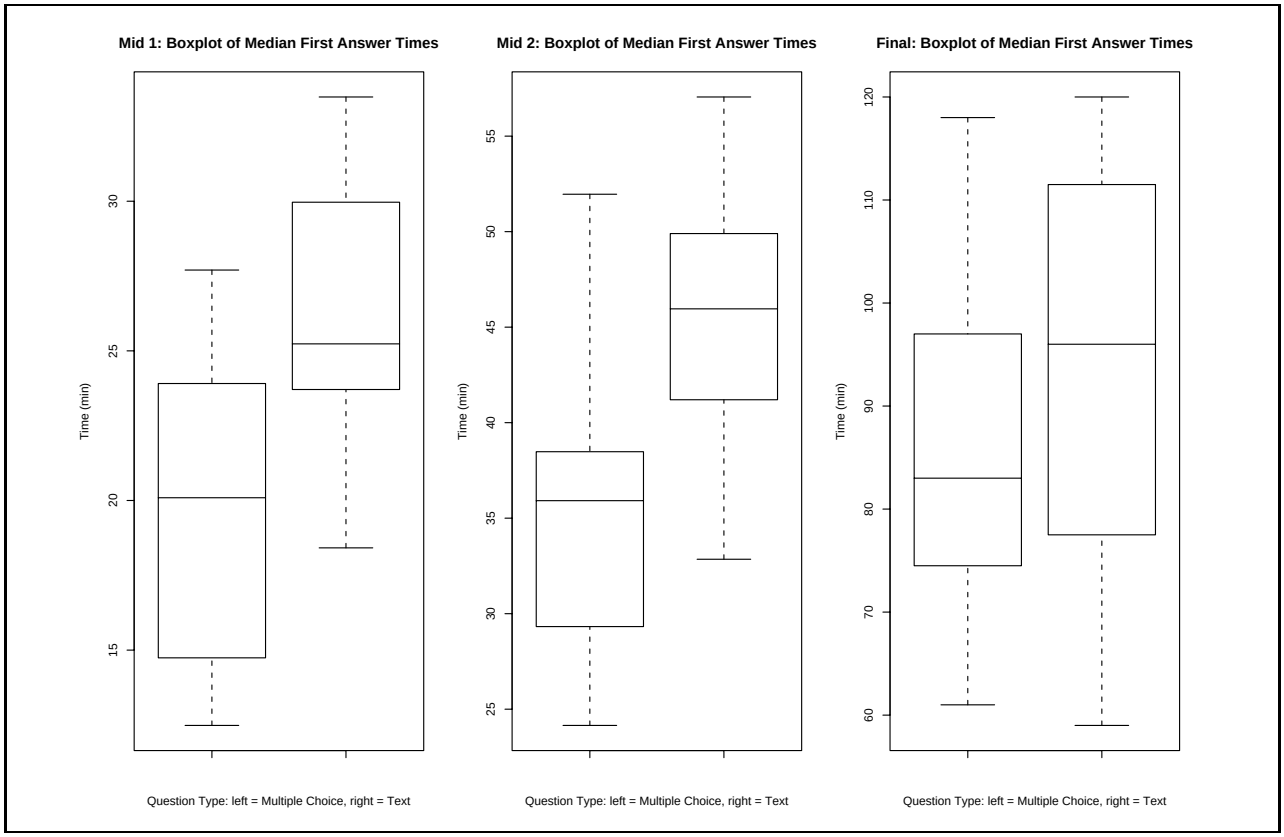

Figure 5: Boxplots of median first answer submission times for MC (left boxplots) and TA (right boxplots) questions for Midterm 1 (left), Midterm 2 (center), and the Final (right).

course, it would be best to provide easier questions at the start of the exam and more difficult questions at the end of the exam. Thus, these students can still demonstrate their basic understanding of the material (and therefore may still pass the exam or entire course) and only miss points in questions at the end of an exam that are given to distinguish between good and average students.

We want to continue our visual analysis of the Spring 2004 data to generate further hypotheses, e.g., to determine whether there is an association of point scores with question order type, i.e., whether students obtain more points on average when MC questions are given first for example. Also, we want to address the question whether there is an association of point scores with student "type", i.e., whether a random answering order result in more or less points than a sequential answering order or a selective answering order. In addition, we have collected additional data for two more offerings of this course (Summer 2004/1: 29 students, Summer 2004/2: 87 students) and plan to use this additional data to conduct statistical inferences based on the hypotheses created from our visual analysis of the Spring 2004 data.

References

Crespie, I. & Morris, D. (1984), 'Question Order Effect and the Measurement of Candidate Preference in the 1982 Connecticut Elections', *Public Opinion Quarterly* **48**(3), 578–591.

Dear, K. (2001), 'Review of Cyberstats', *MSOR Connections* **1**(3), 57–60. `http://ltsn.mathstore.ac.uk/newsletter/aug2001/pdf/appliedstats_cyberstats.pdf`.

Drew, J. H. (1991), Modeling the Effects of Question Order and Form on Survey Responses, *in* '1991 Proceedings of the Survey Research Methods Section', American Statistical Association, Alexandria, VA, pp. 419–424.

McClendon, M. J. & O'Brien, D. J. (1988), 'Question–Order Effects on the Determinants of Subjective Well–Being', *Public Opinion Quarterly* **52**(3), 351–364.

McFarland, S. G. (1981), 'Effects of Question Order on Survey Responses', *Public Opinion Quarterly* **45**(2), 208–215.

Perreault Jr., W. D. (1975), 'Controlling Order–Effect Bias', *Public Opinion Quarterly* **39**(4), 544–551.

Sigelman, L. (1981), 'Question–Order Effects on Presidential Popularity', *Public Opinion Quarterly* **45**(2), 199–207.

Symanzik, J. & Vukasinovic, N. (2002), Teaching Statistics with Electronic Textbooks, *in* W. Härdle & B. Rönz, eds, 'COMPSTAT 2002: Proceedings in Computational Statistics', Physica–Verlag, Heidelberg, pp. 79–90.

Symanzik, J. & Vukasinovic, N. (2003), 'Comparative Review of ActivStats, CyberStats, and MM*Stat', *MSOR Connections* **3**(1), 37–42. `http://ltsn.mathstore.ac.uk/newsletter/feb2003/pdf/activstatscyberstatsmmstat.pdf`.

Utts, J., Sommer, B., Acredolo, C., Maher, M. W. & Matthews, H. R. (2003), 'A Study Comparing Traditional and Hybrid Internet–Based Instruction in Introductory Statistics Classes', *Journal of Statistics Education* **11**(3), . `http://www.amstat.org/publications/jse/v11n3/utts.html`.