

# CHANGING ANSWERS IN EXAMS — FOR THE BETTER OR FOR THE WORSE?

Jürgen Symanzik\*, Utah State University,  
Natascha Vukasinovic, Utah State University.

\*Department of Mathematics and Statistics, Logan, UT 84322–3900,  
e-mail: symanzik@math.usu.edu

**Key Words:** CyberStats, Electronic Textbook, Multiple Answers, Teaching.

## Abstract

During the year 2004, Utah State University offered three sessions of an introductory long-distance statistics course (Stat 2000) for students in its International Program in Hong Kong. These sessions were based on the electronic textbook CyberStats. The exams were given electronically and all student answers were stored in the CyberStats data base. *All* means that if a student changed the answer to question multiple times, all previous answers were still accessible and not only the final answer. In this paper, we investigate how those changes in answers affected students' scores. 98.6% of the multiple-choice questions and 84.4% of text-based questions got answered at least once. Conditional on the fact that a question got answered at least once, text-based answers got changed at a rate of 10.8% that almost doubles the rate at which multiple-choice answers got changed (5.7%). Text-based answers got changed for the second time at a higher rate (1.3%) than multiple-choice answers (0.8%). Conditional on at least one change, the second response resulted in significantly more points for multiple-choice answers as well as for text-based answers. For text-based answers, this trend continued as the third response also resulted in significantly more points than the second response.

## 1. Introduction

During the year 2004, the Utah State University International Program offered three sessions of "Introduction to Statistical Methods" (Stat 2000) as a long-distance course for Utah State University students enrolled in Hong Kong. The Spring 2004 session lasted 15 weeks (1/12/04 through 4/19/04) and was attended by 60 students, the Summer/1 2004 session lasted 12 weeks (2/23/04 through 5/10/04) and was attended by 30 students, and the Summer/2 2004 session lasted 15 weeks (4/26/04 through 8/2/04) and was attended by 89 students. Overall, a total of 179 students went through these three

sessions.

These sessions were based on the electronic textbook CyberStats (<http://www.cyberk.com>) that allowed us to provide exams in electronic form and to record *all* student answers, including all changes, corrections, and additions in a previously given answer. A review of CyberStats Version 2.0 can be found in Dear (2001). Comparisons of CyberStats with two other popular teachware packages, ActivStats and MM\*Stat, can be found in Symanzik & Vukasinovic (2002) and Symanzik & Vukasinovic (2003). A description of a "hybrid" course that makes use of CyberStats can be found in Utts, Sommer, Acredolo, Maher & Matthews (2003). A recent review of six online instructional materials, including CyberStats, can be found in Larreamendy-Joerns, Leinhardt & Corredor (2005). A detailed description of the course organization and teaching experiences with this Web-based long-distance course can be found in Symanzik, Vukasinovic & Wun (2005).

Briefly, two types of questions are supported by CyberStats: multiple-choice questions, where students have to select one correct answer, and essay (text-based) questions, where students have to type their answers in a provided text box. The questions can be chosen from the existing test banks in CyberStats, or instructors can create their own questions.

The main objective of this paper is to investigate what happened to students' scores when they changed their answers. Did they get more points in their second response to a multiple-choice question than with their first response, did they lose points, or were point changes basically random with no significant point gain/loss? Also of interest was the question whether students improved their point score when they corrected or added content to a text-based answer or whether most changes were minor such as the correction of typos.

In Section 2 of this paper, we describe the setup of the electronic exams and their grading. In Section 3, we present our analysis of the recorded answer changes. Section 4 contains our conclusions and suggestions for future work.

## 2. Electronic Exams and Their Grading

The electronic exams in CyberStats consisted of two midterms and one final exam in the Spring 2004 session and one midterm and one final exam in the Summer/1 2004 and Summer/2 2004 sessions. Students took the exams in an electronic classroom in the Institute of Advanced Learning (<http://www.in-learning.edu.hk>) in Hong Kong. While all students of the Summer/1 2004 session could take the exams simultaneously, students from the Spring 2004 and Summer/2 2004 sessions took the exams consecutively in two or three groups. Students from a later group were asked to meet 15 minutes prior to the scheduled start of their exam in a different classroom to minimize the possibility of direct verbal communication.

Exams were a mixture between multiple-choice questions with one correct answer and essay questions that required a text-based answer. Multiple-choice questions had exactly one correct answer and typically three incorrect answers. A small proportion of multiple-choice questions had a different number of incorrect answers, ranging from just one to five incorrect answers. Overall, we can assume that a score of 25% correct answers for a set of multiple-choice questions basically represents the fact that the students were selecting their answers completely at random, i.e., just by guessing.

Students had full access to CyberStats and all its features during the exams. Answers to multiple-choice questions had to be marked via radio buttons and text-based answers had to be filled in into text boxes. To ensure the recording of student answers on the CyberStats server, students explicitly had to click on a "Submit" button after each answer. Answers that were not confirmed by "Submit" were lost in case the Web browser window with the answers was closed or refreshed as well as in case of any computer crashes. Students were frequently reminded that they had to click on "Submit" for each answer. Throughout the three sessions and seven exams, it appeared that four students lost a major part of their answers in an exam either due to a Web browser failure or by unintentionally closing the window with their answers before clicking on "Submit". In such cases, another exam was used to adjust the overall grade for these students.

Students could access the exam questions only during their assigned exam period via an additional password that was provided at the start of the exam. The question order was randomized for each student, but all students within each of the three sessions had

identical exam questions.

The number of points awarded for each answer differed based on difficulty level, total number of questions in the exam, and amount of work a student had to provide to answer the question. Typically, multiple-choice questions were worth 6 to 11 points. Points were awarded on the "all or nothing" basis, i.e., the full number of points was given for a correct answer and zero points for an incorrect one. Text-based questions were worth 10 to 13 points. Points were usually awarded for partial answers, even for an incorrect one, if a student tried reasonably hard to answer the question.

Grading of the exams was done via a html Web page particularly developed for these sessions by Ms. Palyne Gaenir. Multiple-choice questions could be graded automatically by providing the correct answer key. Questions that required a text-based answer had grading forms where points and personal comments could be added. Grading was done by showing all answers to a question at the same time, i.e., all  $n_1$  answers to question 1, then all  $n_2$  answers to question 2, etc. Usually, the number of answers  $n_i$  for each question  $i$  was bigger than the number of participating students because all answers, including duplicated ones, were recorded. However, only the last submitted answer was used to determine the students' grades.

The electronic recording of students answers, including answering times and duplicated answers, provided some interesting data to analyze. A preliminary analysis of these data with respect to the answering order can be found in Symanzik, Erbacher, Gaenir, Vukasinovic & Wun (2004).

## 3. Analysis of Answer Changes

To be able to compare point gains and losses for questions that initially did not have the same number of points, we rescaled all points from 0 (completely wrong) to 1 (completely right). Thus, a multiple choice question that was answered correctly always resulted in 1 point for our analysis, no matter whether the students were awarded with 6, 8, or 11 points. A text-based answer that was answered in half resulted in 0.5 points for our analysis, no matter whether the students were awarded 5 out of 10 points or 6 out of 12 points.

Within this paper, we are not primarily interested how individual students changed their answers, e.g., which percentage never changed an answer, nor whether answers to a particular question (or question type), e.g. doing a calculation related to the normal distribution, got changed most frequently. Instead, our main observational

		Multiple-Choice		Text	
		#Students	#Questions	#Students	#Questions
Spring	Midterm 1	59	13 (15)	59	10
	Midterm 2	57	15	57	9 (10)
	Final	51 (58)	25	51 (58)	25
Summer 1	Midterm	25 (27)	20	25 (27)	15
	Final	27	24 (25)	27	25
Summer 2	Midterm	85 (87)	20	85 (87)	15
	Final	81	25	81	25
Total Students*Questions		7770		6728	

Table 1: Number of participating students and questions to be answered in the seven exams across the study period. The numbers in parentheses indicate the original number of students who participated in an exam and the number of questions that were given. The numbers before the parentheses indicate the number of students and number of question that were actually included in our analysis. The total Students\*Questions indicates the possible maximum number of unique answers in case each participating student had answered each question exactly once.

	Multiple-Choice	Text
Total Students*Questions	7770	6728
1 <sup>st</sup> Response	7659 (98.6%)	5680 (84.4%)
2 <sup>nd</sup> Response	435 (cond. 5.7%)	617 (cond. 10.8%)
3 <sup>rd</sup> Response	60 (cond. 0.8%)	73 (cond. 1.3%)
4 <sup>th</sup> Response	6	15
5 <sup>th</sup> Response	—	8
6 <sup>th</sup> Response	—	2
7 <sup>th</sup> Response	—	1
8 <sup>th</sup> Response	—	1

Table 2: Number of responses received for the first through the eighth time a question was answered. Due to the few students who answered a question four or more times, our analysis dealt with one to three responses only. The percentage for the 1<sup>st</sup> response relates to the ratio of the number of answers received for each question type and the possible maximum number of unique answers for this question type. The conditional probability (cond) relates the number of 2<sup>nd</sup> and 3<sup>rd</sup> responses to the number of 1<sup>st</sup> responses that were received, and not to the theoretically possible maximum number of unique answers.

units are students\*questions, i.e., the questions that could have been answered by the students participating in a given exam. Of interest is how the first response given to students\*questions differs from later responses and whether there are significant point gains/losses that are associated with the given answers. For this paper, we have also pooled the data collected over seven exams in three sessions. As more than one change to students\*questions rarely occurred (only 60 multiple-choice students\*questions and 73 text-based students\*questions received three or more responses), a breakdown into individual sessions or even individual exams would hardly provide a sample size big enough to obtain any statistically signif-

icant result.

Not all of the 179 students participated in each exam. Table 1 summarizes the number of students who participated in a particular exam and the number of multiple-choice and text-based questions that had to be answered in each exam. As explained in more detail in Symanzik et al. (2005), seven students were caught cheating in the Spring 2004 final and were excluded from our analysis for that particular exam. Two students had computer-related problems in the Summer/1 2004 midterm and two had such problems in the Summer/2 2004 midterm. Data for these students for these exams were also excluded from our analysis. Two multiple-choice questions in the Spring 2004 midterm 1 and

Response	Time	Answer	Points
1	6:50:02 AM	find the real corelation between Height and weight. No outlter found.	1/11
2	6:55:33 AM	find the real corelation between Height and weight. No outlier found.	1/11
3	6:57:00 AM	find the real corelation between Height and weight. No outlier found.	1/11
4	7:06:46 AM	find the +ve corelation between Height and weight. No outlier found.	1/11
5	7:06:52 AM	find the +ve correlation between Height and weight. No outlier found.	1/11
6	7:20:58 AM	Find the +ve correlation between Height and weight. No outlier found.	1/11
7	7:24:14 AM	No trend can be find inside it and only few outlier found.	8/11
8	7:32:58 AM	No trend can be found inside it and only few outlier found.	8/11

Table 3: Eight responses received by one student for question 17 in the Summer/1 2004 midterm.

Response	Multiple-Choice		Text	
	Average	Number	Average	Number
1	0.596	7659	0.658	5680
2	0.421	435	0.667	617
3	0.417	60	0.635	73

Table 4: Average number of points awarded for the first through the third response and the number of students\*questions that are contributing to this average.

one multiple-choice question in the Summer/1 2004 final as well as one question with a text-based answer in the Spring 2004 midterm 2 were ambiguously worded with no obvious correct answer and therefore were also excluded from our analysis (while students were awarded full points for these questions, regardless which answer they had chosen). Assuming that each student left in the study for a particular exam had answered each remaining question exactly once, there would have been a total of 7770 unique answers to multiple-choice students\*questions and 6728 unique answers to text-based students\*questions.

In Table 2, we summarize how many answers were received overall and how many questions were answered more than once. From the possible 7770 unique multiple-choice students\*questions, an impressive 98.6% got answered. From the possible 6728 unique text-based students\*questions, only 84.4% got answered. This outcome matches previous results where students tended to answer multiple-choice questions earlier than questions with text-based answers and possibly ran out of time to answer all text-based questions (Symanzik et al. 2004). Also, students indicated in a survey that their favorite exam questions were multiple-choice questions and their least-favorite questions were questions with text-based answers (Symanzik et al. 2005). From Table 2, it also becomes apparent that a much higher percentage of students answered a text-based question for a second (or even a third)

time than a second (or third) answer was given for a multiple-choice question. As hardly any student answered a question more than three times, we only dealt with one to three responses in our analysis.

Nevertheless, Table 3 shows the eight text-based responses to question 17 that were submitted by one student in the Summer/1 2004 midterm. The background information for this question was: “The next interactivity creates scatter plots using data from a survey of a class of 120 students in a college statistics course. Create the plot with X = Age (in years) and Y = Height.” The students had to use an interactive tool in CyberStats and they had to make sure to select the variables referred to in the question. The question asked was: “Is there an evident trend? What outliers are there, if any?” The correct answer was: “There does not appear to be any trend. There are a few outliers to the right. They are people 25 and over: P74, P93, P96, P105.” This student’s responses 1 through 6 were far off the correct answer and only one point (out of 11) was awarded for *some* answer. Note the minimal corrections in spellings, such as “find” to “Find” from response 5 to response 6. Most likely, the student looked at the wrong variables during the first six responses. In response 7, the student correctly identified that there was no trend and that there were some outliers, resulting in eight points (out of 11). To obtain all possible 11 points, students had to provide some additional information on the outliers, e.g., how many outliers were visible or some additional characteriza-

Response	Multiple-Choice			Text		
	Average	Number	P-value	Average	Number	P-value
1	0.260 <sup>-</sup>	435		0.545	617	
2	0.421	435	<0.0001	0.667	617	<0.0001

Table 5: Average number of points awarded for the first and the second response, conditional on one change, and the number of students\*questions that are contributing to this average. <sup>-</sup> indicates that this value itself is not significantly different from 0.25 (p = 0.6428).

Response	Multiple-Choice			Text		
	Average	Number	P-value	Average	Number	P-value
1	0.300 <sup>-</sup>	60		0.436	73	
2	0.133 <sup>+</sup>	60	0.0490	0.537	73	0.0016
3	0.417	60	0.0024	0.635	73	0.0029

Table 6: Average number of points awarded for the first, second, and third response, conditional on two changes, and the number of students\*questions that are contributing to this average. <sup>-</sup> indicates that this value itself is not significantly different from 0.25 (p = 0.4054). <sup>+</sup> indicates that this value itself is significantly different from 0.25, i.e., below 0.25 (p = 0.0107).

tion where these outliers were located. In response 8, the student fixed a typo, but did not really improve the answer, so that the same number of points was awarded.

Table 4 summarizes the average number of points for the first through the third response for all students\*questions that got answered at least so many times. It appears that the highest average for multiple-choice questions was obtained with the first response (about 60%) while consecutive responses (for those students\*questions that got changed) only resulted in about 42%. As stated before, an average of about 25% would be expected if all answers were randomly guessed. For text-based answers, the average remained at about 65% from the first through the third response. As this comparison was not our main goal, no formal tests were conducted. Instead, we moved on to the points awarded, conditionally on one change or two changes.

Table 5 summarizes the average number of points for the first and the second response for those students who answered a particular question at least two times. For the 435 multiple-choice students\*questions that got answered more than once, the first average of about 26% is not significantly different from 25%, i.e., the result that would have been obtained by randomly guessing all answers (t = 0.4642, df = 434, p-value = 0.6428). When comparing the second multiple-choice answer with the first multiple-choice answer for these 435 students\*questions, we observe a significant point gain from about 26% to about 42%. A two-sided

paired-t-test whether the difference between the points obtained for the two multiple-choice answers for students\*questions is different from zero or not resulted in t = -4.1436, df = 434, p-value < 0.0001, and (-0.237, -0.085) as the 95% confidence interval for the difference.

For text-based answers, there is no guaranteed point minimum for guessing, and thus no test whether the observed average is different from a minimum score. When comparing the second text-based answer with the first text-based answer for these 617 students\*questions, we observed a significant point gain from about 55% to about 67%. A two-sided paired-t-test whether the difference between the points obtained for the two text-based answers for students\*questions is different from zero or not resulted in t = -10.0662, df = 616, p-value < 0.0001, and (-0.145, -0.098) as the 95% confidence interval for the difference.

Table 6 summarizes the average number of points for the first through the third response for those students\*questions that got answered at least three times. For the 60 multiple-choice students\*questions that got answered more than two times, the first average of about 30% is not significantly different from 25%, i.e., the result that would have been obtained by randomly guessing all answers (t = 0.8381, df = 59, p-value = 0.4054). The second average of about 13% is significantly different from 25%, i.e., below 25% (t = -2.6362, df = 59, p-value = 0.0107). This means, students are performing worse than if just guessing. When comparing the sec-

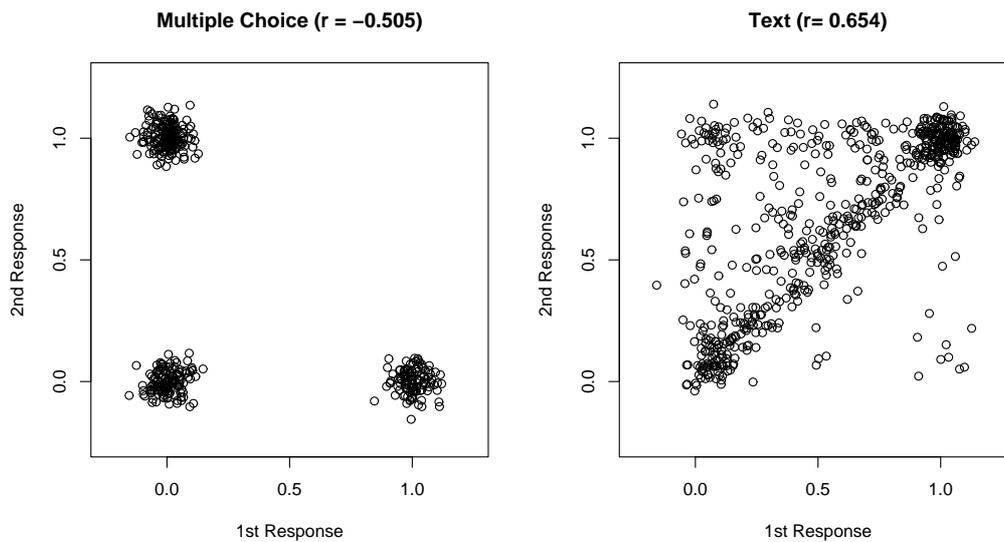


Figure 1: Change patterns for multiple-choice questions (left) and text-based questions (right), conditional on one change.

ond multiple-choice answer with the first multiple-choice answer for these 60 students\*questions, we observed a marginally significant point loss from about 30% to about 13%. A two-sided paired-t-test whether the difference between the points obtained for the first two multiple-choice answers for students\*questions is different from zero or not resulted in  $t = 2.0102$ ,  $df = 59$ ,  $p\text{-value} = 0.04900$ , and  $(0.0008, 0.3326)$  as the 95% confidence interval for the difference. When comparing the third multiple-choice answer with the second multiple-choice answer for these 60 students\*questions, we observed a significant point gain from about 13% to about 42%. A two-sided paired-t-test whether the difference between the points obtained for the second and third multiple-choice answers for students\*questions is different from zero or not resulted in  $t = -3.1754$ ,  $df = 59$ ,  $p\text{-value} = 0.0024$ , and  $(-0.4619, -0.1048)$  as the 95% confidence interval for the difference.

When comparing the second text-based answer with the first text-based answer for the 73 students\*questions that got answered more than two times, we observed a significant point gain from about 44% to about 54%. A two-sided paired-t-test whether the difference between the points obtained for the first two text-based answers for students\*questions is different from zero or not resulted in  $t = -3.2765$ ,  $df = 72$ ,  $p\text{-value} = 0.0016$ , and  $(-0.1618, -0.0394)$  as the 95% confidence interval for the difference. When comparing the third text-based answer with the second text-based an-

swer for these 73 students\*questions, we observed a significant point gain from about 54% to about 64%. A two-sided paired-t-test whether the difference between the points obtained for the second and third text-based answers for students\*questions is different from zero or not resulted in  $t = -3.0827$ ,  $df = 72$ ,  $p\text{-value} = 0.0029$ , and  $(-0.1625, -0.0349)$  as the 95% confidence interval for the difference.

A set of figures will help to further understand these point change patterns. As many values frequently reoccur among students\*questions (for multiple-choice answers, the only existing options are false-false (0, 0), false-correct (0, 1), and correct-false (1, 0) — the option correct-correct (1, 1) does not exist with exactly one correct answer), we have jittered the original values, i.e., we have added a small normally distributed random error to each point score. This effectively helps to address the problem of overplotting in our plots.

Figure 1 shows the change patterns for multiple-choice questions and text-based questions, conditional on one change. As already confirmed by our paired-t-test, more students have changed their multiple-choice answer from incorrect to correct (0, 1) than from correct to incorrect (1, 0). Apparently, a large number of students have simply changed an incorrect answer to another incorrect one (0, 0), or, even worse, changed a correct answer to an incorrect one (1, 0). As the option correct-correct (1, 1) does not exist, the overall correlation between points for consecutive multiple-choice an-

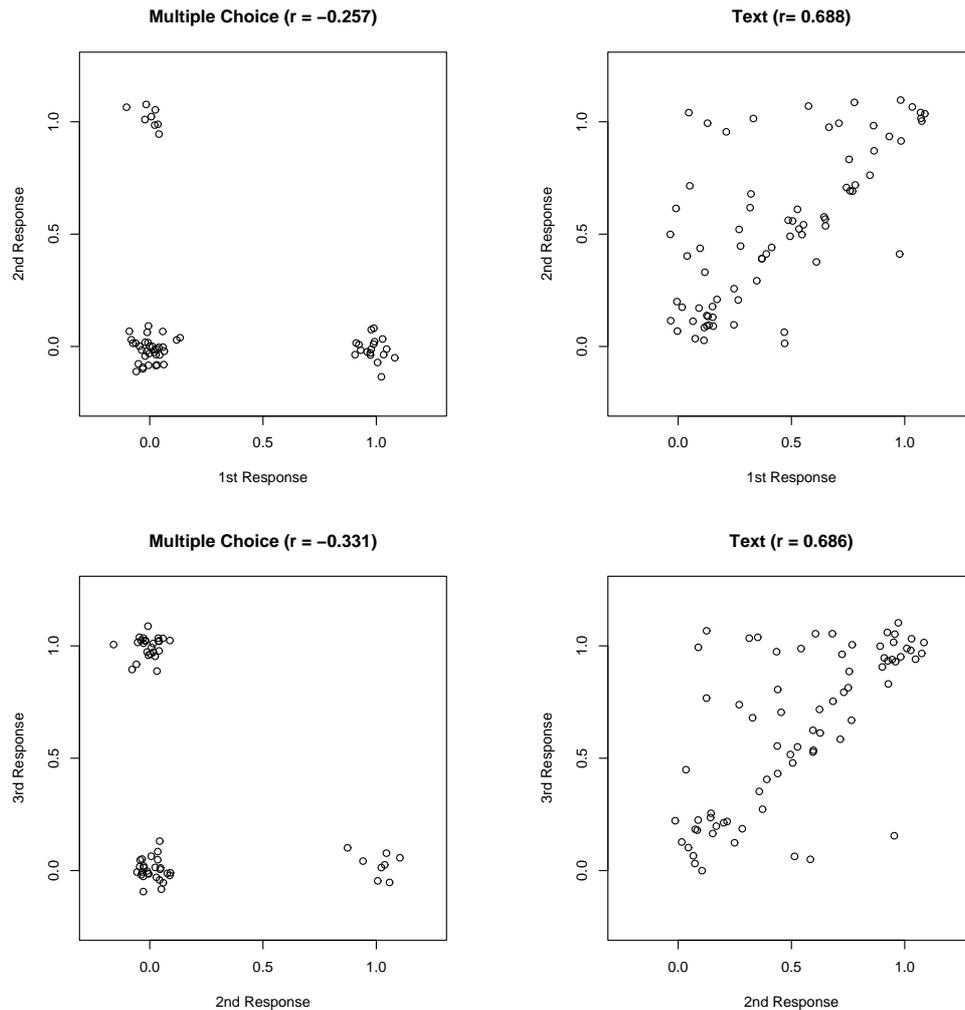


Figure 2: Change patterns for multiple-choice questions (left) and text-based questions (right), conditional on two changes. The effect of the change from the first to the second response is shown at the top, the effect of the change from the second to the third response is shown at the bottom.

swers is negative ( $r = -0.51$ ). For text-based answers, several of the features found in Table 3 can be found in Figure 1 as well. Many changes in text-based answers are minor, often the correction of typos, that do not result in any point gains or point losses. These are the dots on or nearby (due to the jittering) the diagonal line from (0, 0) to (1, 1). The number of students who gained points (the dots above the diagonal), often by answering an additional component of a question, is much bigger than the number of students who lost points (the dots below the diagonal). The overall correlation between points for consecutive text-based answers is positive ( $r = 0.65$ ).

Figure 2 shows the change patterns for multiple-choice questions and text-based questions, conditional on two changes. Similar patterns as in Fig-

ure 1 can be seen. Also, the marginally significant result that students lose points from their first to their second multiple-choice answer can be seen in the top left plot where we have more changes from correct to incorrect (1, 0) than from incorrect to correct (0, 1). The overall correlation between points for consecutive multiple-choice answers is negative both for the change from response one to response two ( $r = -0.26$ ) as well as from response two to response three ( $r = -0.33$ ). For text-based answers, we got a correlation of  $r = 0.69$  in both cases.

#### 4. Conclusions and Future Work

In summary, the exam data collected from three sessions of an introductory long-distance statistics course offered for Utah State University students enrolled in Hong Kong revealed several interest-

ing results. While an overwhelming percentage of multiple-choice questions got answered at least once (98.6%), students omitted text-based answers far more frequently and only answered about 84.4% of such questions. However, conditional on the fact that a question got answered at least once, text-based answers got changed at a rate of 10.8% that almost doubles the rate at which multiple-choice answers got changed (5.7%). However, many of the changes in text-based answers are minor and do not result in any point gain/loss. Also, text-based answers got changed for the second time at a higher rate (1.3%) than multiple-choice answers (0.8%).

Conditional on at least one change, the second response resulted in significantly more points for multiple-choice answers as well as for text-based answers. For text-based answers, this trend continued as the third response also resulted in significantly more points than the second response. This can be explained by the fact that students answered an additional component of a text-based question or just corrected a typo, but rarely changed the entire answer to a text-based question.

For multiple-choice questions that got answered at least three times, the overall trend is not clear, in particular as there is a marginally significant point loss from the first response to the second response. As no other good explanation has been found so far, it would be easiest to contribute this point loss indeed to chance. The change pattern of multiple-choice answers also indicates that students got a multiple-choice answer right with the first response at a much higher rate (about 60%) than after any changes (about 42%). Likely, those students that answered most of the questions just once and selected the right answer were better prepared for the exam than students that did not get the right answer right away. However, those students who changed a multiple-choice answer at least improved when compared to their previous answers.

During our presentation at the Joint Statistical Meetings in Seattle, Washington, we presented a preliminary analysis of our results. For this paper, we have removed the data from students who have been caught cheating and from students who experienced computer problems during an exam as well as we deleted ambiguous multiple-choice questions where students were awarded full points no matter which answer they chose. Although about 400 multiple-choice students\*questions and about 300 text-based students\*questions were removed from the data in various data cleaning steps, we did not see any major change in the results compared to our preliminary results. It is therefore safe to assume

that any other minor problems that may exist in the data also will not result in any major change of our results reported here. In particular, one should admit that grading errors are such a possible problem. For example, another student provided the following three responses to the previously discussed question 17 in the Summer/1 2004 midterm: (1) “no, there isn’t an evident trend and no outliers are there.”, (2) “No, there isn’t an evident trend and no outliers are there.”, and (3) “No, there isn’t an evident trend but with outliers at P74, P93, P96, P105.” For these three responses, the student was awarded with (1) 6, (2) 1, and (3) 11 out of 11 points, respectively. Obviously, a grading error occurred for response 2 where also 6 points should have been awarded. Fortunately, the student got the correct number of points for the final response that counted towards the overall grade.

Finally, it should be emphasized that these data were collected in long-distance courses where the participating students in Hong Kong did not speak English as their first language. While we would expect a similar pattern of changes for multiple-choice answers for English-speaking students, there probably would be less changes of text-based answers, in particular less changes of the form where simple typos got corrected. Unfortunately, CyberStats, the electronic textbook we have used for this course, now is owned by Thomson Publishing and has been totally restructured. Thus, it will not be possible to conduct a similar study among English-speaking students, unless the present owner of CyberStats allows for customized exams and provides required technical support.

## References

- Dear, K. (2001), ‘Review of Cyberstats’, *MSOR Connections* 1(3), 57–60. <http://ltsn.mathstore.ac.uk/newsletter/aug2001/pdf/cyberstats.pdf>.
- Larreamendy-Joerns, J., Leinhardt, G. & Corredor, J. (2005), ‘Six Online Statistics Courses: Examination and Review’, *The American Statistician* 59(3), 240–251.
- Symanzik, J., Erbacher, R., Gaenir, P., Vukasinovic, N. & Wun, A. C. K. (2004), On the Effect of the Ordering of Questions in Exams — A Visual Analysis, in ‘2004 Proceedings’, American Statistical Association, Alexandria, VA. (CD).
- 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/activstatscyberstatismmstat.pdf>.
- Symanzik, J., Vukasinovic, N. & Wun, A. C. K. (2005), Experiences with International Web-Based Introductory Long-Distance Statistics Courses, in ‘2005 Proceedings’, American Statistical Association, Alexandria, VA. (CD).
- 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>.