UNIVERSITY OF CALIFORNIA SAN DIEGO

Helping Students Write More Effectively; An Investigation of Student Test-Taking Strategies

A Thesis submitted in partial satisfaction of the requirements for the Master of Science

in Biology

by

Matthew Arash Nedjat-Haiem

Committee in charge:
James Cooke, Chair
Thomas Bussey
Sarah Stockwell

2019

The Thesis of Mathew Arash Nedjat-Haiem is approved, and is acceptable in quality and form for publication on microfilm and electronically

_____

_____

_____
Chair

University of California San Diego

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. James Cooke for his unwavering support throughout the duration of my project, as the chair of my committee. Having in my corner was instrumental in my journey as a master's student. Dr. Cooke's support as an advisor and mentor could not be replaced and I am truly grateful.

I would also like to acknowledge Nanea Dominguez. Nanea was instrumental in obtaining much of the sorted data used throughout this thesis paper. Her contributions were greatly appreciated.

This thesis is currently being prepared for submission for publication of the material. Dominguez, Nanea and Cooke, James. The thesis author was the primary investigator and author of this material.

ABSTRACT OF THE THESIS

Helping Students Write More Effectively; An Investigation of Student Test-Taking Strategies

by

Matthew Arash Nedjat-Haiem

Master of Science in Biology

University of California San Diego

Dr. James Cooke, Chair

When students have to generate their own responses, subsequent retention of information is greater than when students have to recognize a correct response, as is the case with multiple-choice questions (Kang, et al., 2007; McDaniel et al., 2007). While previous studies have looked into the strategies students use when preparing for a test (Hartwig, et al., 2012; Ross, et al., 2006) no studies have assessed the strategies that students use when taking the test itself. Through the use of a post exam survey, which asked students what strategies they used during free response exams, we investigated "what strategies do students use during open-ended exam questions?" In our treatment quarter we utilized an exam wrapper to see if strategies changed in response to metacognitive reflection. Next, we investigated if the change in student strategies affected student's use of verbosity, which we defined as extraneous details that does not contribute to a student's answer, when answering free response questions. We found that students report using a wide variety of strategies during free response exams. The strategies used by students change in response to the metacognitive reflection assignment, which led to a decrease of extraneous

information in their answers. By investigating the strategies students used on free response exams as well as how student-employed meta-cognitive reflection may impact their use of certain strategies, we hoped to improve student's communication of science.

Introduction:

The literature demonstrates the benefits of using free response exams over multiple choice exams to increase retention of information and student learning outcomes. When students have to generate their own responses, subsequent retention of information is greater than when students have to recognize a correct response, as is the case multiple-choice questions (Kang, et al., 2007; McDaniel et al., 2007). Previous studies have investigated what strategies students use to prepare for exams, and strategies utilized during multiple choice exams (Towns et. al 1993; Ross et.al 2006; Hartwig & Dunosky 2011; Prevost & Lemons 2016). Though free response exams have been noted to lead to increased learning outcomes when used instead of multiple choice (Kang et.al 2007), no study has assessed what strategies students use during free response exams. This study aimed to address the lack of literature and investigate what student's report doing when answering free response questions.

Through the implementation of an exam survey, we asked students of a class in a large research university about what strategies they used during their free response examinations in a upper-division biology course. Using the theoretical framework of phenomenography, which involves the interpretation of the human experience in an effort to understand ones perspective or point of view (Marton 1986; Malterud, 2001;Sin 2010), we hoped to understand the breadth and scope of the types of things students felt they were doing during free response exams. By looking into student's habits and approaches to learning and testing, which is done here by looking at student's behaviors during free response exams we increase our pedagogical opportunities with the ultimate goal of improving student outcomes and learning. (Marton 1986).

After the initial stages of our study, we found that students report using a wide variety of strategies during free response exams. One particular strategy was the use of extraneous information, or verbosity, in their answer responses. When analyzing student exams, we found this to be true, that many students, were being overly verbose when answering exam questions, often including information that is irrelevant to the question stem. Verbosity within communication has often been suggested to hinder the effectiveness of delivering the desired message (Hsia 1977; Kuo 1995; Arbuckle etl.al 2000). Within science writing, verbosity is also admonished (Aaronson 1977; Kirkman & Turk 2002.). To combat this, we implemented a treatment through the use of an exam wrapper, which we call here as our post-exam reflection assignment. Our exam wrapper was designed to employ metacognitive conditioning so students can focus on beneficial habits for their next exam, as exam wrappers and metacognition have been shown to change students habits, and improve test scores and learning outcomes (Baird 1984; Ross et.al 2006; Mazumder 2010; Siegsmund 2016; Gezer-Templeton et.al 2017). Using this treatment, we hoped to shift student's perceptions on what constitutes a good answer, and ensure that they focus on being concise rather than verbose; to ultimately offer a platform for students to practice and learn about effective communication and writing in science.

Our early results also indicated that students felt pressured for time. Time pressures have been linked negative emotions leading to emotional exhaustion, pessimistic academic outlook; and has been found to lead to overall burnout. (Ackerman & Gross 2003; Williams et.al 2013). We hoped that by reducing the amount of excess information students wrote during exams, then they would feel less pressured for time since students wouldn't spend as much time writing non-essential information in their answers. We hoped to find a reduction in time pressures to lead to better outcomes, and more positive emotions associated with the course.

In the later stages of our studies, we were curious as to what caused students to write excess information; was it based on student performance? Or was the excess information due to the level of cognitive demand expected of the students depending on the question type. To analyze question type we used Bloom's taxonomy to classify questions based on cognitive load (Bloom 1956).The first 4 levels of Blooms analysis which are knowledge, comprehension, application and analysis have been considered to be a hierarchy of cognitive load. Suggesting that each subsequent Bloom level requires integration of the previous level, suggesting the increase in cognitive load (Zheng et.al 2008). In our exam setting, only bloom levels 3 and 4 are considered higher-order thinking (Bissell 2006). It has been suggested that students have trouble performing when cognitive loads are higher (Zoller, 1993, Bransford et.al.m 200; Bailin, 2002, Crowe et.al 208). We were curious to see if this cognitive load may affect the amount of excess information included in exams. Below we discuss our important findings and conclusions, and the implications of these findings on biology education.

Methods:

Setting and Participants

The participants in this study were from an upper division human reproduction course for biology majors, BIPN 134, at UCSD. The study was performed over two quarters: the summer, which served as our control, and the winter, which served as our experimental group. Enrolment in the summer quarter was 83, and enrolment in the winter quarter was 251.

Data Collection

Students in each quarter were given two midterms that consisted almost entirely of open-ended questions. The midterms were essentially isomorphic to one another from one quarter to the next. Immediately after the midterms, students were asked to complete a survey using the course's online platform. Students were given roughly forty-eight hours to complete the survey, which accounted for 2% of their course grade (1% each). During the treatment quarter, students' exams were marked for extraneous information by the course's instructional assistants and they were given a post-exam reflection assignment. In the control quarter, no marks for extraneous information were made on student exams, and no reflection assignment was given. In the control quarter, extraneous information was assessed by a research associate. This research associate also re-evaluated the extraneous information listed on the treatment students' exams, to ensure that extraneousness was evaluated consistently.

Post Exam Survey

The post exam survey consisted of six Likert-scale questions and one open-ended question that asked: "Did you have any specific strategies when answering the free-response questions? If so, please describe any strategies you used." Answering this question was optional for the students. The Likert-scale questions asked students how they felt about: the exam representing material from class; question clarity; time pressure; including as much information as possible in answers; exam difficulty; and feeling prepared for the exam before and after taking it. Of the 83 students registered in the control section, 49 completed the survey after midterm 1, and 35 those provided an answer to the open-ended question of strategies used during free-response questions. Sixty-eight of 83 students in the control section completed the survey after midterm 2, and 48 of those provided an answer to the open-ended question of strategies used

during free-response questions. In the treatment group, of the 251 students registered, 194 completed the survey after midterm 1 and 164 of those provided an answer to the open-ended question of strategies used during free-response questions. Two hundred and thirteen of the 251 students in the treatment condition completed the survey after midterm 2, and 169 of those provided an answer to the open-ended question of strategies used during free-response questions.

Quantification of Student Implemented Strategies

The student responses were gathered, binned and coded into categories from each of the two post exam surveys for each group. Using a grounded theory approach, strategy categories were created based on students' self-reported strategies. After the first post-exam survey, two people worked together to develop categories from roughly 20 students. Those same two people then independently coded another twenty responses. Interrater reliability was measured using the percent agreement method, where the number of times both individuals agreed was divided by the total number of 'scores' made by both individuals (agreed / total number of responses). With this method, the interrater reliability score was 83.3%. After which point, one person scored all remaining responses. The list of strategies was adjusted as new categories arose in subsequent post exam surveys.

Student Perceptions

To assess student's perception of the exam, the post exam survey given to students featured a five-point Likert-scale, with the options, strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree. Questions featured on the survey asked about student preparedness, exam difficulty, student's willingness to write as much information as possible, and time pressures.

Extraneous Information

To assess extraneous information students used on exams, exam questions that contained extraneous information were marked on students' exams by an instructional assistant during the grading period. Instructional Assistants were asked to underline any part of the answer that did not contribute anything of value to the students' answers. That is: any irrelevant information. During the control section, exams were photocopied and returned to students and the photocopies were then marked for extraneous data by a student researcher trained to identify irrelevant information in student answers. In this way, the students in the control section were not aware of the experiment. The exams from the treatment were also photocopied so that their extraneous information scores could be calculated. In both cases, the extraneous information was tallied for midterm 1 and midterm 2, to compare the changes in proportion of questions students used extraneous information.

What is extraneous?

Answers that included information that was not essential to answering the question were marked as extraneous. Extraneous information was any information that was irrelevant for the particular question in which that information was provided, even if the content was factually accurate, if it did not relate to the question it was marked as extraneous (see Figure 1). In BIPN 134, partial marks are given out for answers that contain any information that was relevant. That is: a student can get 0.5 out of 3 for a question because they did not correctly answer the question, but they did provide a statement that was relevant in some way. As a result, questions that received zero marks that still had writing were considered to be "extraneous". To verify this approach, 50 scores of 'zero' marks were assessed for any relevant content. Of the fifty assessed, only two answers had any information that was relevant. Since this rate (4%) is quite

low, we felt justified in continuing with our model of ‘zero points with anything written down is “extraneous”.

Extraneous information by question

To assess whether certain questions were leading to more extraneous information than others, questions from midterms one and two from the control were paired with their isomorphic counterpart from treatment quarter. The proportion of extraneous information for each question pair was then quantified. Each question pair from the two midterms were then categorized based on Blooms Taxonomy (Crowe et.al 2008). In accordance to Bloom’s Taxonomy, questions that asked students to describe a process were scored as a “2” which correlates a comprehension question, which is Bloom’s second level. Questions that asked students to predict an outcome using contextual information were scored as a “3”, corresponding to application-based questions. While questions that asked students to infer how different components were associated were scored as a “4”, which is considered analysis-based questions in Blooms Taxonomy.

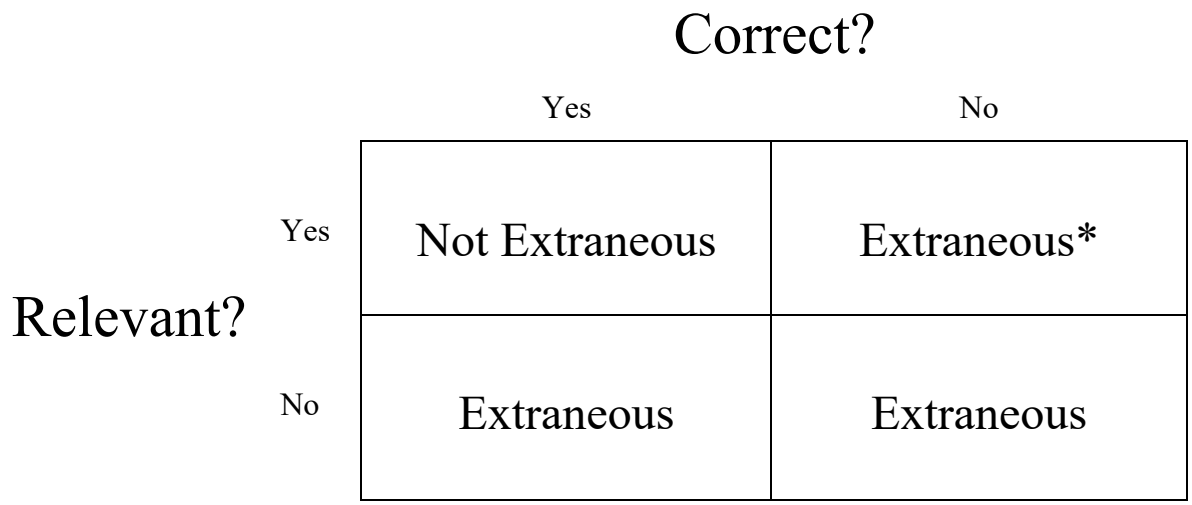

Figure 1. Identifying Extraneous Information.
The diagram above explores the methods involved in identifying what this study depicted as extraneous. Extraneousness was determined based on relevance and accuracy. *Answers receiving zero marks were deemed irrelevant, as responses receiving zero marks indicated that the answer was not geared towards the question while answers with some level of thoughtfulness to the true question received partial marks.

Post Exam Reflection

The treatment group was given the post exam reflection after their first midterm in order to induce metacognition within the treatment group about their habits leading up to, and during the exam. After receiving the assignment, the day after the first exam, the students had five days to complete the assignment. This assignment asked students about how many hours they studied for the exam, how they studied, and asked whether or not they included information that didn't belong in an answer. If students marked yes on that question, they were asked to rewrite that question without the unnecessary information and write why they included the extra information in the first place. The use of metacognitive reflections has improved exam preparation strategies (and subsequent performance) in other studies (Baird 1984; Mazumder 2010; Montague 1992). Here, the primary function of the post exam reflection was to have the students look at their test, and rewrite one of their questions that was marked for extraneous information, but this time to write it more concisely.

Results:

Students Use a Variety of Self-Reported Strategies Revolving Around Four General Approaches

Students self-reported a variety of different strategies when answering free response questions on their midterms. A total of sixteen different strategies were reported and were sorted within four overarching categories. The four categories we observed were: 1) *formulating an answer, 2) writing [their] response(s), 3) time saving measures* and 4) *question details*. Table 1 outlines the strategies used and includes representative quotes used by students.

Table 1. Student Self-Reported Free Response Test Taking Strategies

The strategy list was created using students answers to the post-exam survey which asked students about strategies they implemented during the exam. The quotes corresponding to each strategy serve as examples of quotes representative of each strategy. Quotes were obtained from answers to the surveys administered during the treatment quarter. Percentage reflects the total percentage of responses for each category across all midterms across both treatment and control sections.

Strategy/Category	***Quote Example***	***%***
Formulating an Answer		***36.2***
Brain Storm in Writing	"For some questions, I wrote notes on the side of question to organize my thoughts before writing the answer." (MT1 153)	9.7
Drawing	"Usually I begin answering by drawing a picture or an arrow sequence of what is happening before I put it into words" (MT1 #76)	5.9
Write normal, then deviation from normal	"I would explain the way the correct pathway should work, and then address the change in the pathway that was mentioned in the test question." (MT1 174)	3.9
Write as much info before answering question	" I write everything that I know down before I write my response"(MT1 #8)	2.4
Mental/Visual Processing	"I try to visualize the pathways in my head." (MT1 #9)	12.3
Writing a Response		***32.9***
Include as much info. as possible	"Just writing as much as possible with the amount of time we had for the exam." (MT1 #75)	11.8
Include only important information	"I tried putting down key points because if I tried putting in too much info I would run out of time." MT1 #25	17.9
Answer the question**	"I tried to answer only what was being asked. If it was yes or no question, I answered and elaborated on it." (MT1 #61)	3.1
Answered Similar to Discussion Section**	"Answered questions in the way they were answered on discussion sheets" (MT1 #154)	0.2
Time Saving Measures		***17.3***
Transcription Methods to improve timing	"...I used bullet format and shorthand (arrows, abbreviations, etc.) instead of writing in complete words and sentences." (MT1 153)	8.1
Order of questions answered	"When answering the free response questions, I focused on the set of FR questions I knew the answers to first and then went back to the others to work them out." (MT 1 #30)	8.1
Point total suggests amount of information	"I look at how many points the question is worth and try to provide enough information that correlates to how many points are available." (MT1 #71)	1.1
Question Details		**13.6**
Analyze what the question is asking	"Make sure to understand the question before trying to answer it." (MT1 #21)	4.6
Focus on question vocabulary	I look for key words (vocabulary) and use the vocabulary to guide my answers. (MT1 #130)	2.4
Annotate key points in the question	"First I underlined the key concepts that the question was asking. The, I tried to write the keywords and diagrams that are related first before answering." (MT 1 131)	5.2
Ask for clarification**	"The questions [asked were] not so clear, so [I] took lot of time just to understand the question, so my strategy was to ask the TA for clarification" (MT1 #56)	1.5

Students Focus on Formulating Their Answers

When students were asked to report what strategies they used during free response exams, most reported focusing on formulating the answer they wanted to write, with 36.2% of strategies recorded belonging to this category. Students also reported strategies focusing on writing their actual responses, which accounted for 32.9% of the total strategies. Strategies involved in saving time and focusing on the question details accounted for 17.3% and 13.6% respectively. A larger breakdown of the individual strategies within each category is outlined in Table 1.

Students Feel Pressured for Time During Midterm Exams

During data collection, our Likert survey result indicated that students felt pressured for time during the first midterm. Over 78% of students said they felt pressured for time, while 8.51% suggested they did not feel pressured for time. The results are outlined in table 2 below. The table was created by binning the "strongly agree" and "agree" respondents and binning the "strongly disagree" and "disagree" respondents.

Table 2. *Time Pressures on Midterm 1 of Control*
The following data was obtained from the post-midterm surveys after the first exam of the control quarter. From the five-point Likert-scale, 'Strongly Agree' and 'Agree' were collapsed into 'Agree' featured in the table. 'Strongly Disagree' and 'Disagree' were collapsed and featured on the table as 'Disagree'.

I Felt Pressured For Time			
	Agree	Neither Agree/Disagree	Disagree
n	37	6	4
%	78.72%	12.77%	8.51%

Metacognition Affects What Strategies Students Use to Formulate Their Answers

During strategy categorization, we noticed that a number of students indicated that they felt it was important to include as much information as possible; meanwhile students were also feeling pressured for time. We wondered if students would change their strategies in response to our reflection assignment. Our results suggest that the use of metacognition in a post-exam surveys changed student behaviors. When looking at the reported strategies for writing a response, we found that during the control quarter, 3.7% of students thought it was important to include as much information as possible and it was 12.25% after midterm two of the control. Contrastingly after midterm 1 of the treatment quarter, 16.6% of students reported that they tried to include as much information as possible, following treatment, when asked again for strategies after the second midterm this number dropped to 10%. More striking was how the number of students who felt it important to include only important information increased almost 2-fold, 13.6% to 25.1 between midterm 1 and 2 during the treatment quarter. While in the control quarter the proportion of students that reporting only including important information had a slight increase from 9.43% to 14.29%. The data mentioned is depicted in Figure 2 below. Though strategies for writing a response changed in both groups, we found that in the control more students acknowledged that they strived to be more concise on the second midterm, and the number of students who claimed they tried to be verbose in their writing responses decreased in the treatment group but increased in the control. These findings suggest that metacognition aimed at answer conciseness may change student's writing habits during exams.

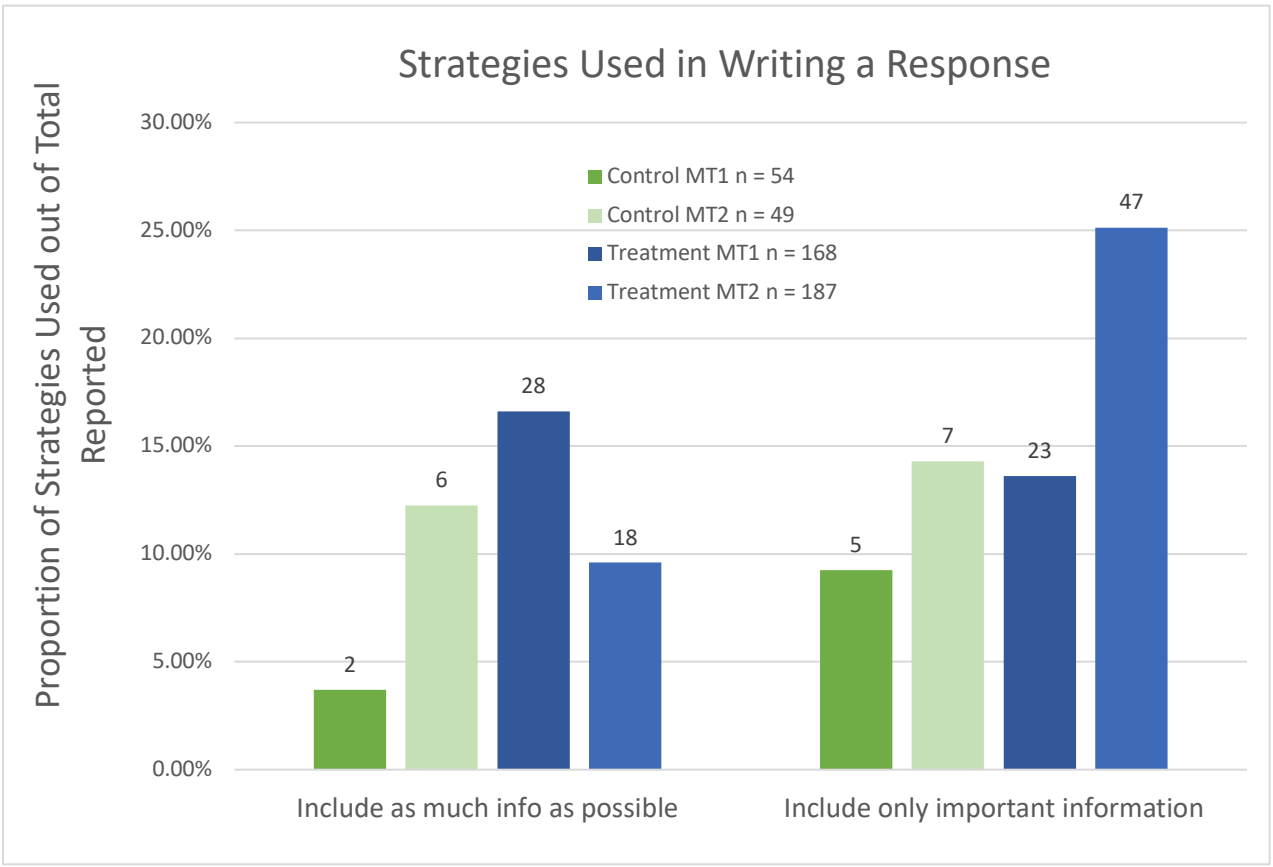

Figure 2. *Strategies Students Used in Writing Their Answers*
Strategies from the most midterm survey were grouped across midterms for our treatment and control group. Here we look at strategies studied used in writing a response.

Students Perceived Use of Writing Strategies Matches Extraneous Use

After quantifying the amount of excess information included on students' exams (Figure 3), we found a statistically significant difference in the proportion of questions containing excess information amongst all students between the treatment group (M=12.08%) and the control group (M=16.35%) on Midterm 2, $p = 0.037$. We also performed paired t-tests to compare test performance from midterm 1 to midterm 2 for both the treatment and control group, separately. We found a statistically significant difference between midterm one and midterm two for both the control and experimental group. To account for missing exams in midterm 2 of the treatment, we re-analyzed the data comparing the 46 remaining students whose data were present in midterm 2 to themselves during midterm 1. The removal of the remaining students did not

change any or trend or statistical significance of the analysis. Overall, our data indicates that use of excess information decreases from midterm 1 to 2, for both the control and treatment, but the use of a reflection assignment had a statistically significant impact in decreasing extraneous information compared to control.; suggesting the possible efficacy of our treatment in reducing excess information and changing student writing habits.

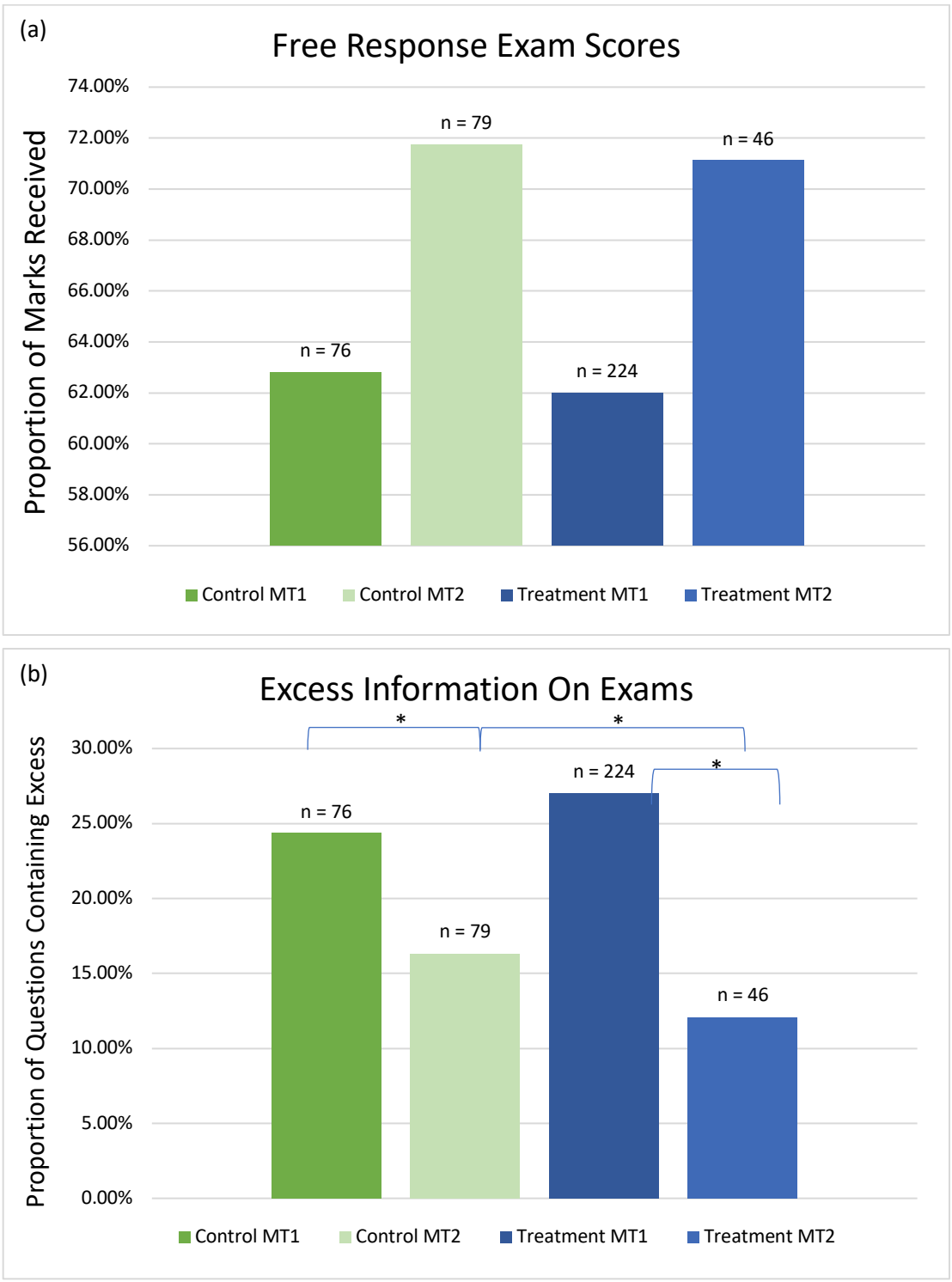

Figure 3. *Exam Scores and Excess Information Statistics*
(a) Exam Scores were recorded from the two midterms for both control and treatments. (Bottom) Data depicted in the figure details the number of questions including excess information across all of the student's midterms within their group, control and treatment groups respectively.
T-tests were performed to check for statistical significance, * denotes $p < 0.05$.

Students in the Treatment Group Feel Less Pressured for Time

The survey data suggests that compared to midterm one, on midterm two, students feel less pressured for time. For the control group, on midterms 1 78.73% said that they felt pressured for time, moving into the second midterm; this dropped by 16.96%. In the treatment group, 66.98% of students suggested that they felt pressured for time on the first midterm, while only 54.17% of students felt pressured for time on midterm two. When looking at the proportion of who students who "strongly agreed" that they were pressured for time this group of students increased by 1.40% in the control group, while decreasing by 16.80% in the treatment group. Our survey data demonstrates a trend that students feel less pressure for time on subsequent exams, see Table 3 below for data overview.

Table 3. The following data was obtained from the post-midterm surveys after each exam in both the control and treatment quarters. From the five-point Likert-scale, 'Strongly Agree' and 'Agree' were collapsed into 'Agree' featured in the table. 'Strongly Disagree' and 'Disagree' were collapsed and featured on the table as 'Disagree'

Time Pressures Across Groups

	Strongly Agree	Agree	Neither	Disagree	N
Control Midterm 1	38.30%	40.43%	12.77%	8.51%	47
Control Midterm 2	39.70%	22.06%	20.59%	17.35%	68
Δ Control	+1.40%	−18.37%	+7.82%	+8.84%	
Treatment Midterm 1	37.63%	29.38%	16.49%	16.49%	194
Treatment Midterm 2	20.83%	33.33%	20.83%	25.00%	192
Δ Treatment	−16.80%	+3.95%	+4.34%	+8.51%	

Stronger Students Seem to Include Less Excess Verbiage on Exams

To determine whether or not there was an association between student's exam scores on the free response portion, and the amount of excess information included on exams, a Pearson Correlation analysis was performed. Across all tests and groups, student test performance has a negative correlation with proportion of questions containing excess information (MT1 Control R= -.527, MT2 Control, R= .-611, MT1 Treatment R = -.515, MT2 Treatment R = -.381. All R

values were significant at a $p > 0.05$ level. Figure 4). Suggesting that as student test performance increases, we can expect a lower proportion of questions containing excess information in student responses. The treatment groups' second midterm had the lowest correlation out of any of the exams with a R value of -.381. This lower value may be due either to high-performing student's including more information, or low-performing students including less information. To determine which of these is occurring, we used a one-tailed paired t-test where equal variance was not assumed, we found that the treatment quarter had significantly higher extraneous scores for lower performing students (i.e.: students who scored lower than 65%), included excess information on their first midterm exam. After the treatment for midterm two, there was a statistically significant difference between the extraneous scores for the control group and treatment group for midterm 2; here we found that the treatment group had a significantly less extraneous information compared to control (Table 4). There was no statistically significant change when a independent t-test was used to compare the amount of excess information for students who scored better than 80% when comparing midterm one control to treatment, or comparing midterm two control to treatment. To demonstrate the robust effect of the treatment without having vastly different "N" values, the students exams lost after midterm 1 of the treatment were not considered in the data used to create this table (Table 4), specifically for the data on treatment MT1; indicating that the treatment group were the same group of students from midterm one to midterm two. When all available midterm 1 treatment tests were used, there was no changes in statistical significance when comparing groups, or trends observed.

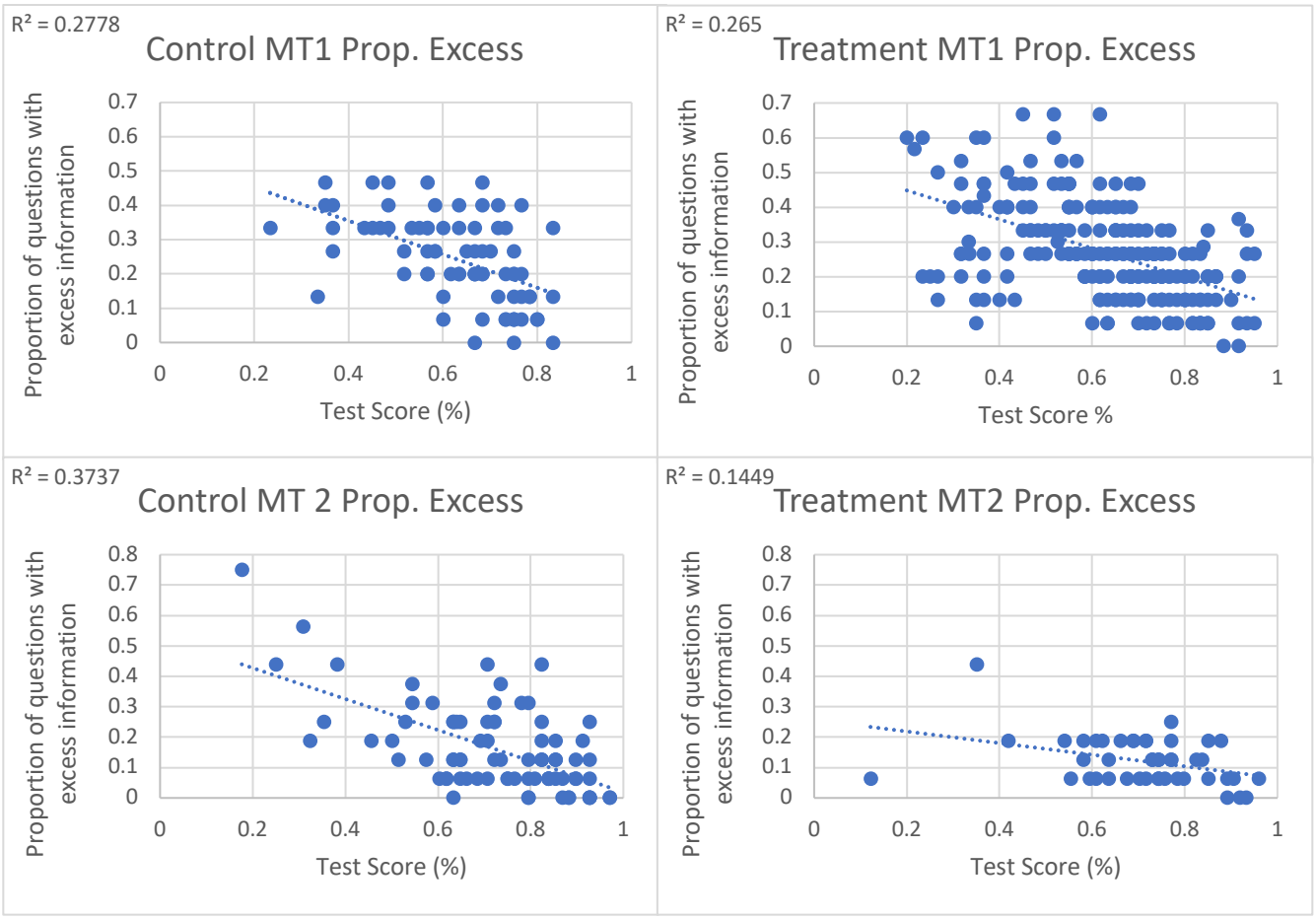

Figure 4. *Correlating Exam Scores and the Use of Excess Information*
Each student's individual exam score, as well as their own proportion of excess information included on their exam was recorded and plotted on a scatterplot; a scatterplot was created for each midterm from both groups. The top left (a) demonstrates the scatterplot for the control midterm 1 ($R^2 = 0.28$; $R = -0.527$). (b) The top right scatterplot was produced for the treatment quarters midterm 1 ($R^2 = 0.27$; $R = -0.515$). (c) The bottom left was produced for the control groups midterm 2 ($R^2 = 0.37$; $R = -0.611$). (d) The bottom right graph was produced to demonstrate the data from midterm two of the treatment group ($R^2 = 0.14$; $R = -0.381$)

Table 4. The following data was obtained by grouping students who performed above 80% and below 65% from midterm two of the control and treatment. Subsequent tests were then averaged out for scores as well as the proportion of questions containing excess on the exams.
*Indicates statistical significance at $p < 0.05$ level

Amount of Excess Included for High and Lower Performing Students

	Proportion of Questions Containing Excess	Exam Score	n	N
Students Scoring > 80%				
Control MT1	0.19	86%	5	80
Treatment MT1	0.17	89%	6	42
Control MT2	0.097	88%	30	76
Treatment MT2	0.080	89%	12	46
Students Scoring < 65%				
Control MT1	0.31*	51%	37	80
Treatment MT1	0.39*	53%	20	42
Control MT2	0.23*	53%	27	76
Treatment MT2	0.15*	56%	12	46

Examining Question Type Using Bloom Taxonomy

To determine whether or not question type affected the amount of extraneous information included on exams, we classified all exam questions based on Blooms Taxonomy (Crowe et.al 2008). As demonstrated in figure 5, all free response questions from midterm one and midterm two of the control were paired with their isomorphic counterpart from the treatment quarter to assess the amount of times students used extraneous information on each question. The proportion of excess for each question was characterized by bloom level which resulted in figure table. The excess values by bloom level were then compared via ANOVA, resulting in a p value of 0.57. We found that the amount of extraneous information varied by question, however there was no difference in proportion of excess included in questions between blooms levels.

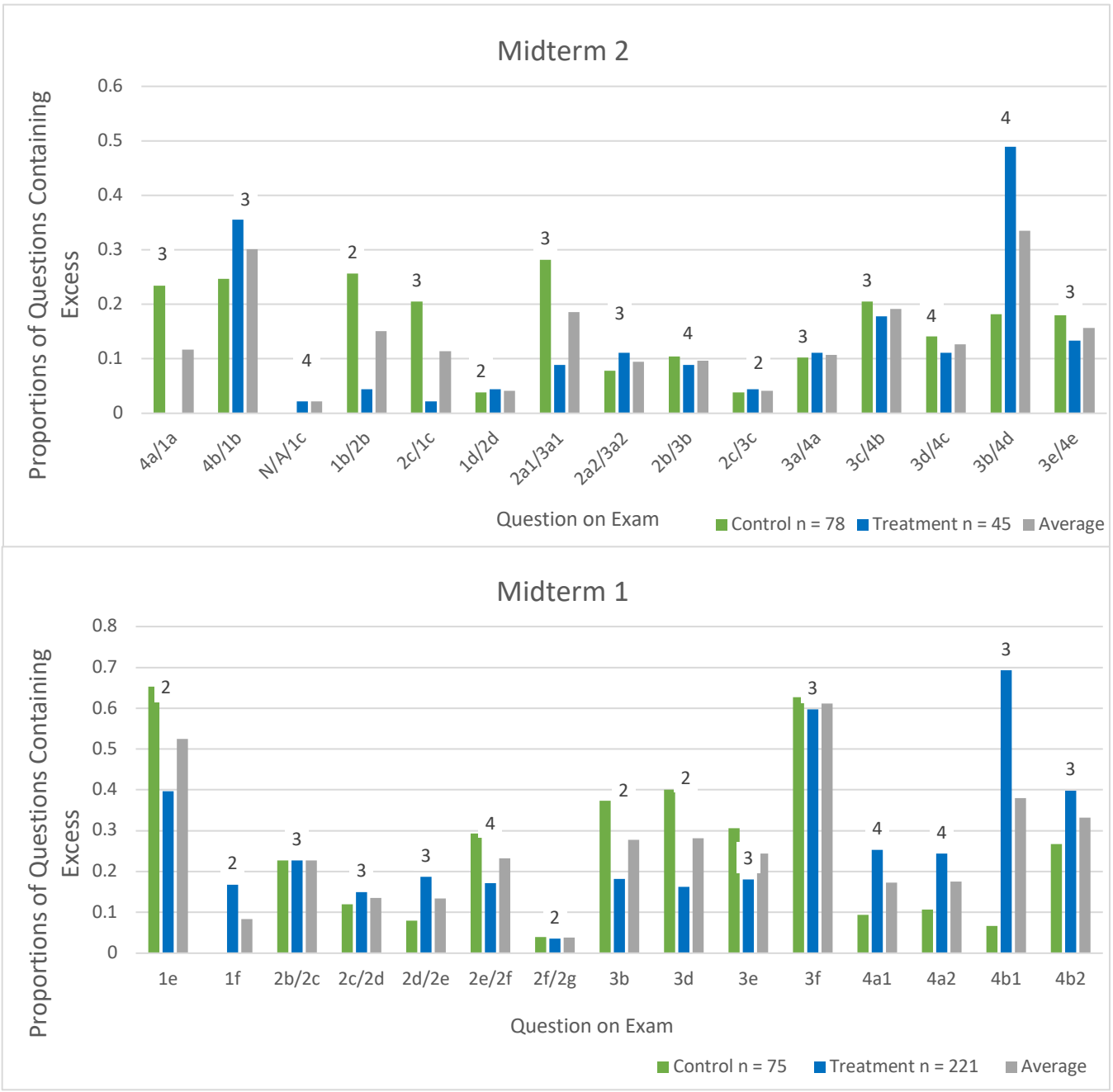

Figure 5. *Bloom Classification and Extraneous Quantification of Exam Questions*
The graphs above demonstrate extraneous information by Bloom's level from midterm one and two for the control and treatment. Certain questions were isomorphic to each other from the control to treatment quarters and were described together as the proportion of students using extraneous for each question was quantified. The numbers above the bars for each question represents the Bloom level for each question, 1 = knowledge, 2 = comprehension, 3 = Application, 4 = Analysis, 5 = Synthesis.

Table 5. All questions were binned based on bloom taxonomy level to assess how often students included excess information on each bloom level. We found 7 question pairs that were classified as Bloom level 2, 16 pairs level 3 questions, and 7 level 4 pairs. The pairs were based off each question and their isomorphic counterpart from the other quarter. Proportion of excess refers to the average proportion of excess included in each question type.

Excess information by Bloom Level

	Comprehension	Application	Analysis
Bloom Level	2	3	4
Proportion of Excess	16.55%	22.57%	16.59%

Discussion:

In this study we investigated the strategies students use during free response exams, a field that has yet to be studied. We found that students use a wide variety of strategies, centered on developing context for their answers, writing their responses, saving time, and obtaining details from the question stems they are given. During our initial findings we noticed that students stated that they were being overly verbose in their writing by admitting to "writing as much information as possible". We also noticed that students suggested that they were feeling pressured for time. Using a metacognitive treatment, we were able to reduce the amount of extraneous information they included in their responses. Subsequent analysis revealed that the low-performing students were preferentially benefitting from our reflection treatment. Finally, while we found that the amount of extraneousness varied by question, there was no relationship between the level of Blooms taxonomy and the amount of extraneous information.

Students Report a Wide Variety of Strategies

As we looked into the strategy's students used during free response exams, we found that students implemented a vast variety of different categories. We noticed that of the 16 categories students reported using the ones used most often were strategies involved in formulating an answer or strategies involved in writing a response. We categorized strategies involved in

developing the context in a solution or involved in the retrieval of information to answer a question as strategies involved in “formulating an answer”. “The formulating an answer” group contained strategies such as written brain storming, drawing, writing a pathway then a deviation from the pathway, mental and visual processing and writing as much information before answering the question. We categorized responses involved in actually transcribing one’s answers as strategies involved in “writing a response”. Strategies that fell into the “writing a response category” were ‘including as much information as possible’, ‘including only important information’, ‘answering similar to discussion section’ which describes a type of short hand writing style, and interestingly enough several students proclaimed they just ‘answered the question’. Though the two major categories reported were involved in formulating answers and writing their responses, students also suggested ‘time saving measures’ and strategies involved with ‘question details’. We categorized any strategy focusing on time saving methods, basically ways to maximize available time during exams, as a ‘time saving measure’, such strategies included, “transcription methods to improve timing”, answering questions in a specific order based on knowledge ‘order of questions answered’, and writing an amount of information based on how many points the question was worth ‘point total suggest amount of information’. We also found that students used strategies that were involved with drawing information out of the question stem these strategies were categorized under ‘question details. Strategies involved with question details were, ‘analyzing what the question is asking’, ‘focus on question vocabulary, ‘annotating the key points in the question’, and asking for clarification. Though some strategies had some overlap and likely could have been binned together, we chose to do these separate distinctions based on student’s own descriptions of their strategies (see Table 1). We limited the number of bins that each self-reported strategy was placed into, by only placing students into bins for behaviors they explicitly stated in their responses.

Developing Answers During Free Response Exams

Our results of student strategies suggested that most students are focusing on figuring out ways to come to the answer ("formulating an answer"), which is similar to approaches taken during multiple choice exams. In multiple choice exams students tend to use deductive reasoning to eliminate answer choices (Towns & Robinson 1993), but such practices would not be effective in a free response setting where no possible choices are given. Instead we find that students are forced to do tasks such as brain storming, or suggested more specific types of brain storming, such as writing a healthy pathway, then deviation from healthy. Writing the "healthy, then deviation from healthy" for a pathway would be considered a domain-specific strategy, considering this type of approach may be more appropriate when answering physiology questions when there is something wrong. A previous study (Provost & Lemons 2016) demonstrates that use of domain specific strategies increased the likelihood of a student answering questions correctly.

Students self-report performing tasks such as brainstorming and drawing to generate their own responses. Such processes involved in recollection lead to greater retention of information post-exam compared to recognizing a correct response during multiple choice exams (Kang et al 2007). Our findings suggest that some of the ways students report retrieving the information needed to answer exams. It may be beneficial in the future to stress the use of domain-specific strategies in the classroom to promote academic success.

Student's Utilize Strategies to Save Time

Time saving strategies are a ubiquitous technique that have been detailed previously for their usefulness in maximizing time during exams (Boyd 1988; Koenigs 1987, Phillips, 1983). Previous studies had investigated the frequency with which these types of strategies were used

during multiple choice exams (Towns & Robinson 1993) and found that students did not frequently use such strategies during exams. The omnipresence of time-saving strategies may suggest that they are often used without consideration, which would suggest their low level of reporting in previous studies, as well as in our current study.

Of the time saving measures mentioned, one involved writing with abbreviations in order to save time, 'transcription methods to improve timing', a strategy specific to free response exams. The strategy 'point total suggests amount of information' could only be used in the free response setting. Students employing this strategy thought it was beneficial to write more for questions worth more points, while writing less on questions worth less; this approach intended to maximize points and increase the effective allocation of time during the exam; surprisingly only two students reported this strategy. There is some overlap with the categorization of strategies used, as one could also suggest writing an answer concisely would also fit in to strategies used to save time, but this was categorized in strategies involved when writing a response due to the fact that the major objective of this strategy was to answer the question. In an upcoming section we outlined the use of a strategy that we found was actually detrimental to student time allocation during exams, the use of verbosity in their responses.

Student's Use the Question Stem To Guide Their Answers

During multiple choice exams students have been noted in using the question stem to further guide details to look for in answers (Kim & Goetz 1993; Towns et.al 1993). Here we find that students take a similar approach when dealing with the question stems of free response questions. Students reported breaking down what the question is asking, annotating the question stem and focusing on vocabulary to help guide their responses. Students claimed that using these strategies allowed them to use the details within the question to help guide their answers and

provide guidance in answering the questions so that they are on point when writing their details. Student's also noted that if they were unsure about what a specific question was suggesting, they would ask for clarification in order to better answer the question. Free response questions are different in their nature considering one must construct a factual response based on details in a question, this allows for highlighting gaps in students' subject-knowledge (Bennett 1991). Due to the variability and the need to construct a written response during free response exams, students used these strategies were implemented by students to make sure that they were not missing points due to misunderstanding of the question.

Student's Approaches to Writing Their Answers

When we took a look at what strategies students report using during exams, we saw that there was essentially a dichotomous approach in the way students assessed that they wrote their answers. The two main approaches students took in writing their answers were including as much information as possible, and including only the important information. The idea of writing as much as possible has been studied in literary context, as well as its detriment to the quality of writing. Verbosity in writing is considered redundancy, which information theorists define as the amount of relative information (Shannon & Weaver, 1949; Fano 1949; Cherry 1957; Hsia 1977). Verbosity in science, as well as other fields, takes away from the essential message trying to be conveyed (Krifka 2002). Ideally students would stick to answering the question by only including information that is essential for their answer. We hypothesized that students enlisted in this approach for several reasons. As detailed in some of the student's own responses, they often included as much information as possible because they were unsure of what the answer truly was and wanted to maximize points by utilizing a 'dump of information'. Other reasons include: they were unsure of what the Professor was asking, or were unsure of their expectations. These could

arise due to a lack of mastery of material, or an unfamiliarity in in a professor's testing style. Since verbosity in communicating science may lead to a degradation of the message presented, it would be ideal to lower the amount of redundancy in student's responses. Previous studies have suggested that metacognition can improve students' habits and test-taking methods, and subsequent scores (Siegesmund 2006, Baird 1986). In part two of our study we hoped to utilize metacognition to decrease the amount redundancy in student's answers.

Changes in Student Self-Reported Strategies Through Metacognitive Reflection

Noticing a trend of students using excess information, we used a metacognitive approach to have students reflect on their writing habits, to try and reduce any excess verbiage. Metacognitive approaches have been used throughout educational settings with success in changing student behaviors and outcomes (Baird 1986; Sternberg 1998; Mazumder 2010). By getting students to reflect on their writing habits, we found a significant decrease in the amount of excess information they included on their exams. In our control group, we found a significant decrease in the amount of excess information included from midterm one, to midterm two, even in the absence of treatment. This suggests that exam style familiarity may play a role in the amount of excess information students include, as they could be trying to write as much possible since they are unfamiliar with the professor's expectations. Yet with this trend in mind, the treatment had a more significant impact, determined via t-test, in reducing the amount of excess information included when comparing midterm two of the control and treatment group. Since our results demonstrated that exam familiarity also helps reduce the amount of excess information on exams, modelling answers to exam style questions could be beneficial to students, as they would get accustomed to proper ways to answer exam questions, while also getting exposure to the Professors expectations during exams (Butler 2010; Mayer et.al 2012).

Reduction of Excess Does Not Necessarily Improve Time Constrains

As discussed earlier, verbosity in writing hampers the effectiveness of communication one tries to achieve in their writing (Hsia 1977). Verbosity also branches away from ideal scientific writing (Kirkman & Turk 2002). Our results indicated that the use of a metacognitive treatment was effective in reducing the amount of extraneous information within students' answers.

Though we found a change in self-reported strategies toward being more concise, it was difficult to assess whether or not time pressures were alleviated since both control and treatment quarters saw a substantial decline in student's perceptions that they felt pressured for time during the exams. Though the second midterm from the treatment group had a lower percentage compared to, the percentage of students who felt pressured for time during midterm one of the treatment group was already lower than that of the control group. We did however find a drastic decline in students who strongly agreed that they were pressured for time during the treatment group. While this data suggests that students may be less polarized about their time pressures, the overall data trend prevents us from making any determinations on the efficacy of the metacognitive shaping of student's habits on time pressures during the exam.

Metacognitive Reflection Appears to Help Lower-Performing Students' Use of Verbosity

As we investigated extraneous information, we correlated student performance with extraneous information. By running the correlation analysis, we were curious to see if there was any association between the amount of extraneous information students included and their exam scores. We found that both midterm one of the control and treatment and similar "r" values showing statistically significant negative associations, at -.527 and -.515, respectively. These results indicated that at least after initial testing, there was a mild negative correlation between

exam scores and excess information which would suggest that as student's grades went up, the amount of excess information included would decrease. We performed the analysis once again after midterm two, and found that for the control group, the strength of association between excess information on their exam questions and grades increased from -.527 to -.611. While in the treatment group, the strength of association between the two variables decreased from -.515 on midterm 1 to -.381 on midterm 2. We were curious to see which types of students were responsible for the shift within each group from midterm one to midterm two so we separated the two ends of the grade spectrum, students who exam scores below 65%, and who students who had scores better than 80%. The results demonstrated that the the control and treatment group students who scored above 80% on their exam did not have significant differences the in amount of excess information they included for midterm one, or two. When comparing students who performed below 65% on their exams, we observed a statistically significant difference between the control group and treatment group both when comparing midterm one of the control and treatment, as well as when comparing midterm two extraneous scores. When comparing the use of excess information on midterm two, the treatment group was found to use a significantly smaller amount than control. The data suggests that at first the group of students in the treatment group were using more extraneous than the control group, but after the treatment, their use of extraneous information was significantly less than that of the control group, suggesting a robust treatment effect. These results may offer insight in the decrease correlation strength in the treatment quarter as lower performing students began to change their habits due to the treatment and subsequent changes resulted in a noticeable drop in use of excess information. The robust effect of the treatment on extraneous use by lower-performing students suggests its efficacy in helping low performing students. These results suggest the effectiveness of metacognitive

treatment in improving learning and test taking outcomes for lower performing students, as seen in other studies (Chi & VanLehn 2008).

While the treatment decreased the amount of extraneous use across the overall class, there was no effect on exam scores suggesting that reducing extraneous information does not impact exam performance. Though exam scores do not increase, such decreases in a students' use of extraneous information is still beneficial as it will work to improve their science communication skills (Arbuckle etl.al 2000) and overall writing habits.

Bloom Level Does Not Appear to Predict Verbosity

After analyzing trends in student scores to see if exam scores and student knowledge was related to use of excess information, we were interested to see if the questions implemented by the professor may be including terminology or tasks that motivate students into using more extraneous information. To separate question styles and requirements, we classified our exam questions using Blooms Taxonomy (Bloom 1956); (Crowe et.al 2008). After looking into the Blooms levels of our questions, we found no distinct trend in the amount of excess information included by question type. Our results suggested that there was no distinct trend when looking across all question types (Figure 5). When questions were then grouped based on Blooms level (Table 5), there seemed to an identical amount of excess included on comprehension questions and analysis questions, yet a slightly higher amount included on application questions. Without a distinct and glaring difference, and with no apparent trend with use of extraneousness and Bloom level, we determined that Bloom's level does not play a role in indicating the amount of verbosity included in exam responses; suggesting that the amount of excess information is not based on cognitive load demanded, at least according to the hierarchy of Bloom's taxonomy.

Conclusion:

Overall Findings

Through this study we were able to gather information about what students perceive they are doing during exams, specifically free response exams. We found that students use a wide variety of strategies during free response exams, revolving around four larger domains of strategies, formulating answers, writing responses, saving time, and examining the question. Within these strategies we also found a strategy we believed hindered students overall writing quality, which was the use of excess information during exams, as some students felt it was important to include as much information as possible when answering a question. Through the use of a metacognitive reflection, we were able to reduce student's use of extraneous information in their answers, while promoting awareness and a conscious change in writing habits for students. We found that though the amount of extraneous information students used during exams declined during exams, promoting better writing habits; the use of extraneous information did not necessarily impede exam performance, or lead to a greater time pressure for students. When trying to find the root cause of extraneous information, we correlated the use of extraneous information to exam scores and found a negative correlation. The strength of the correlation decreased after implementation of our treatment, we found the strength of the correlation decreased due to the fact that we appeared to specifically be helping lower-scoring students to have less verbosity within their answers. When analyzing question types, we found that Bloom level, as mentioned in Blooms taxonomy (Bloom 1956) had to apparent trend with extraneous information in students' answers.

Future Direction

In future studies we hope to further investigate what within-question stems promote verbosity within students' answers. For this investigation we plan to group students for exam style questions, and through the implementation of a think-aloud we hoped to gain insight on what types of questions, or what words within questions incentivize students to write more redundant answers.

As another future treatment to follow this study, we hope to implement scaffolding within the classroom so that students can see effective ways to answer exam style questions through the eyes of the professor or teacher's assistants. Scaffolding exam style questions has been shown to promote learning and improve students test scores (Butler 2010; Mayer et.al 2012). Through the use of scaffolding we hoped not only to improve student test scores, but we also hoped to improve their conciseness within their exam answers to further promote effective writing practices as emphasized in this study.

Limitations

In any situation where humans are responsible for quantifying data, there is a possible source for human error. This idea of human error is a flaw within the framework of phenomenography as people are responsible for contextualizing other people's experiences; this problem was countered through the use of inter-rater-reliability to try to assure quality of classification of strategies, yet the chance for human error still exists. The extraneous information in our exams was also quantified by humans, though experts in the matter, such subjectivity may play a role; suggesting that one expert may consider different things extraneous. We attempted to minimize the variability within this set of quantifications by having by structuring an extraneous "rubric" which is outlined in figure 1. When comparing exam data

from control-treatment quarters it is important to note that these tests are essentially isomorphic yet could still potentially hold some differences that may affect the results of this study. In any study when examining human test subjects there is always some variability that cannot be accounted for, this should also be kept in mind when attempting to generalize the findings of this study.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. James Cooke for his unwavering support throughout the duration of my project, as the chair of my committee. Having in my corner was instrumental in my journey as a master's student. Dr. Cooke's support as an advisor and mentor could not be replaced and I am truly grateful.

I would also like to acknowledge Nanea Dominguez. Nanea was instrumental in obtaining much of the sorted data used throughout this thesis paper. Her contributions were greatly appreciated.

This thesis is currently being prepared for submission for publication of the material. Dominguez, Nanea and Cooke, James. The thesis author was the primary investigator and author of this material.

References

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational research*, 58(4), 375-404.

Ackerman, D. S., & Gross, B. L. (2003). Is time pressure all bad? Measuring the relationship between free time availability and student performance and perceptions. *Marketing Education Review*, *13*(2), 21-32.

Amonson, S. (1977). Style in scientific writing. *Essays of an Information scientist*, *3*, 4-13.

Arbuckle, T. Y., Nohara-LeClair, M., & Pushkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. *Psychology and Aging, 15*(1), 65-77

Bailin, S. (2002). Critical thinking and science education. *Science & Education*, *11*(4), 361-375.

Baird, J. R. (1986). Improving learning through enhanced metacognition: A classroom study. *European Journal of Science Education*, *8*(3), 263-282.

Bennett, R. E. (1991). On the meanings of constructed response. *ETS Research Report Series*, *1991*(2), i-46.

Bissell, A. N., & Lemons, P. P. (2006). A new method for assessing critical thinking in the classroom. *BioScience*, *56*(1), 66-72.

Bloom, B. S. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay*, 20-24.

Boyd, R.T. (1988). *Improving your test taking skills*. (ERIC Document Reproduction Service No. ED 302 558)

Bransford J., Brown A. L., Cocking R. (2000). How People Learn: Brain, Mind, Experience, and School, Washington, DC: *National Academies Press*

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118.

Chi, M., & VanLehn, K. (2008, June). Eliminating the gap between the high and low students through meta-cognitive strategy instruction. In *International Conference on Intelligent Tutoring Systems* (pp. 603-613). Springer, Berlin, Heidelberg.

Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, *7*(4), 368-381.

Entwistle, N. (1997). Introduction: Phenomenography in higher education. *Higher Education Research & Development*, *16*(2), 127-134.

Fano, R. M. The transmission of information. MIT Research Lab. Electronics. Tech Report 65, 1949.

Gezer-Templeton, P. G., Mayhew, E. J., Korte, D. S., & Schmidt, S. J. (2017). Use of exam wrappers to enhance students' metacognitive skills in a large introductory food science and human nutrition course. *Journal of Food Science Education*, *16*(1), 28-36.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement*? Psychonomic Bulletin & Review*, 19(1), 126-134.

Hsia, H. (1977). Redundancy: Is it the lost key to better communication?. *AV communication review*, 25(1), 63-85.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. European Journal of Cognitive Psychology, 19(4-5), 528-558.

Kim, Y. H., & Goetz, E. T. (1993). Strategic processing of test questions: The test marking responses of college students. *Learning and individual differences*, *5*(3), 211-218.

Kirkman, J., & Turk, C. (2002). *Effective writing: improving scientific, technical and business communication*. Routledge.

Koenigs, S. (1987). *Test scores count! A handbook for teaching test-taking skills*. (ERIC Document Reproduction Service No. ED 301 572)

Krifka, M. (2002). *Be brief and vague! And how bidirectional optimality theory allows for verbosity and precision*. na.

Kuo, C.-H. (1995). Cohesion and Coherence in Academic Writing: From Lexical Choice To Organization. RELC Journal, 26(1), 47–62. https://doi.org/10.1177/003368829502600103

Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. Qualitative Research Series, 358, 483–488.

Marton, F. (1986). Phenomenography—a research approach to investigating different understandings of reality. *Journal of thought*, 28-49.

Mayer, R., Moeller, B., Kaliwata, V., Zweber, B., Stone, R., & Frank, M. (2012, September). Educating engineering undergraduates: effects of scaffolding in a problem-based learning environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*(Vol. 56, No. 1, pp. 2507-2511). Sage CA: Los Angeles, CA: SAGE Publications.

Mazumder, Q. H. (2010). Metacognition approaches to enhance student learning in mechanical engineering classroom. In *Proceedings of the World Congress on Engineering* (Vol. 2).

McDaniel, Mark A., Henry L. Roediger, and Kathleen B. McDermott. "Generalizing test-enhanced learning from the laboratory to the classroom." Psychonomic bulletin & review14.2 (2007): 200-206.

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. *Journal of learning disabilities*, *25*(4), 230-248.

Prevost, L. B., & Lemons, P. P. (2016). Step by step: Biology undergraduates'problem solving procedures during multiple-choice assessment. CBE—*Life Sciences Education*, 15(4), ar71.

Ross, M. E., Green, S. B., Salisbury-Glennon, J. D., & Tollefson, N. (2006). College students' study strategies as a function of testing: An investigation into metacognitive self-regulation. *Innovative Higher Education*, *30*(5), 361-375.

Siegesmund, A. (2016). Increasing student metacognition and learning through classroom-based learning communities and self-assessment. *Journal of microbiology & biology education*, *17*(2), 204.

Sin, S. (2010). Considerations of quality in phenomenographic research. *International journal of qualitative methods*, *9*(4), 305-319.

Shannon, C. E., & Weaver, W.*The mathematical theory of communication*. Urbana: University of Illinois Press, 1949

Towns, M. H., & Robinson, W. R. (1993). Student use of test-wiseness strategies in solving multiple-choice chemistry examinations. Journal of research in science teaching, 30(7), 709-722.

Williams, K. H., Childers, C., & Kemp, E. (2013). Stimulating and enhancing student learning through positive emotions. *Journal of Teaching in Travel & Tourism*, *13*(3), 209-227.

Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's taxonomy debunks the" MCAT myth". *Science*, *319*(5862), 414-415.

Zoller, U. (1993). Are lecture and learning compatible? Maybe for LOCS: Unlikely for HOCS. *ACS Publications*, 70(3).