UNIVERSITY OF CALIFORNIA SAN DIEGO

Why do Students add Extraneous Information on Free Response Exams and Can we Reduce it?

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Biology

by

Nanea Dominguez

Committee in charge:

James Cooke, Chair
Lisa McDonnell
Melinda Owens

2020

The Thesis of Nanea Dominguez is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

I'd like to dedicate my thesis to:

My Mom and Dad, whose endless love and belief in me allowed me to be who I am today

All my friends who loved me throughout everything and encouraged me constantly

Special thanks to Jared and Charlie for being my Sun and my Sestra

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to first acknowledge my mentor, Dr. James Cooke for his wonderful guidance, support, and counsel. His knowledge and expertise were paramount in helping me design and analyze my research.

I would also like to acknowledge my committee members, Dr. Melinda Owens and Dr. Lisa McDonnell for their input and constructive feedback I received.

Lastly, I would like to acknowledge Journal club (Chem 251) for the opportunity to mingle with some amazing graduate students along with having a deeper understanding of what education means.

Material from this thesis is coauthored with Nanea Dominguez, Matthew Nedjat-Haiem and James Cooke. The thesis author was the primary author of this chapter.

ABSTRACT OF THE THESIS

Why do Students add Extraneous Information on Free Response Exams and Can we Reduce it?

by

Nanea Dominguez

Master of Science in Biology

University of California San Diego, 2020

Professor James Cooke, Chair

Previous research indicates that one of the most popular strategies utilized by undergraduate biology students on open-ended exams is "include as much as possible." While this strategy may have possible benefits, it could also have drawbacks such as including extraneous information. My thesis sought to answer two questions: 1) why do students include extraneous

information in their responses, and 2) could we design an intervention to help reduce the amount of extraneous information included in students' answers?

To answer the first question, we conducted think-aloud interviews of 26 students enrolled in an upper-division physiology class and asked them to explain their rationale for adding extraneous information in their exams. We identified 6 different categories from the student's responses. To answer the second question, we implemented a modeling worksheet during discussion sections, based on the theory of cognitive apprenticeship as a potential treatment for extraneousness.

The worksheets significantly reduced the amount of extraneous information on both Midterm 1 and Midterm 2 for the treatment section compared to control sections. Our modeling treatment assisted low performing, mid performing, and high performing students lessen the amount of extraneous information on their exams. The treatment section was also significantly better at identifying what extraneous information was. Modeling concise answers for students could be a potential treatment that other professors can use to aid students in their ability to write succinctly on free-response open-ended exams.

INTRODUCTION

General Overview

My research focuses on student's responses, specifically to free-response exams for an undergraduate human physiology class. There have been reported benefits to using free response questions as a way to assess a student's comprehension of the material. However, due to the open-ended option, students can often write anything they deem relevant to the answer, even if it's not. These can be rooted in issues in problem-solving from time constraints or a lack of understanding of what the grader is looking for. I attempt to illuminate why students add extraneous information to their open-ended exams by employing think alouds in order to find out what was their rationale is. I designed an intervention, a potential solution to reduce the amount of extraneous information present in their exams. The invention, modeling worksheets, were based on metacognition and aimed to teach novices to start to think like experts by utilizing cognitive apprenticeship. The goal is that this can be beneficial for the students to not only identify extraneous information but also how to write concisely. Once they develop proficiency, the students can further apply this skill in their future endeavors.

Different Ways of Administering Tests

Teachers can evaluate students' understanding of the material ranging from multiple-choice, fill in the blank and free-response questions. Each strategy has its own benefits and teachers can employ a variety of each of them on their exam. Multiple choice and fill-in-the-blank exams are the most efficient to grade because of the ease of distribution and no grader being required to read each response (Bennett et al. 1991). However, students can often get the right answer without

fully understanding the concept by employing a range of strategies to find their answer from the options provided. These strategies can include the process of elimination, ranking, and even just recognizing keywords (Lissitz et al. 2012; Blackley 2009). Free response, open-ended questions can offer partial credit and allow the grader to follow the thought process of the students as they write their answers ((Ackerman & Smith, 1988). However open-ended questions may be more time-intensive to grade since a grader must read through every sentence and assign points for the student's justifications. Methods such as multiple choice are more effective at the retention of memorization type concepts while essay and free-response questions are better at formulating more complex ideas and responding to modified problems (Ackerman & Smith, 1988).

Instructors have shown interest in free-response exams since it's been indicated that they are an effective way of learning and retention of material (Haynie 1994). Free-response questions are associated with deeper conceptual understanding (Jensen et al., 2014) and better long-term retention of information (Kang et al., 2007). My thesis explores a problematic aspect of free-response questions: students providing extraneous information.

The Problem->Verbosity/Extraneous

A difficulty that instructors can run into when grading these free-response exams is the use of extraneous information on their test responses. Earlier work indicated that "writing as much as possible" is one of the more common test-taking strategies students employ during free-response exams (Nedjat-Haim et al. 2019). But not all the information written down is applicable to the question and can be classified as extraneous. My definition of extraneous information is concepts that are not relevant to the answer to the problem and regardless if they are right or wrong, will not contribute to the point total. In other fields such as psychology, the term "verbosity" is used to

identify when people communicate in a way that lacks clarity or coherence (Arbucke 2000; Krifka 2011). It unfortunately has been found to be a rising trend in scientific writing that is published in journals (Grais 2008). I plan on continuing a research path that links verbosity to free-response exams in biology.

A previous study that I was involved in was the first to quantify extraneous information on open-ended exams and utilized a post-exam reflection exam to change student behavior. (Nedjat-Haim et al. 2019). It was found to change their strategy that they employed on free-response exams from "include as much information as possible" to "only include important information" (Nedjat-Haim et al. 2019). I go one step further on this and **t**he goal of my thesis is to 1) understand why biology students use extraneous information in their open-ended responses, and 2) determine whether we can reduce the amount of extraneous information students include on open-ended questions by using modeling worksheets. With an in-depth understanding of what extraneous information is and how to reduce is, hopefully, students can learn to become better, more concise writers.

Why do Students add Extraneous Information/Think Alouds

One method researchers have used to gain an understanding of student's thought processes is by using think-aloud interviews (Someren 1994; McDonnell et al. 2016). Think-alouds have been an effective way for the interviewers to gain a glimpse into the student's mind and to understand why they do what they do. Think-alouds are a methodology that asks participants to work through an example and to narrate their cognitive processes and reasonings to gain insight into their problem solving (Someren 1994). The interviewer asks prompting questions to allow the student to verbalize their thought processes but takes careful considerations in order to not bias the

student. The main point of the exercise is to understand the process of problem-solving and not necessarily the outcome of it. My thesis uses think-alouds to better understand why students add extraneous information when answering free-response exam questions.

Steps of Problem Solving/Test Pressure->Can lead to Extraneous Information

During free-response questions, students might have to come up with a novel solution that can involve parts from many different concepts, pathways, and the solution can be different than what has been previously mentioned in class (Glaser 1991). Understanding what a question is asking for involves many steps: reading comprehension, word association, organizing ideas, problem-solving, and then articulating the response (Collins 1982). If the students falter at any of these steps and are unsure of how to progress, they could rely on more familiar methods to answer questions such as "knowledge telling strategies" (Scardamalia 1984). They can be unsure of what is relevant and can often over-explain and add material that isn't needed in order to maximize points on their answer (Bereiter 1988).

Time constraints can furthermore add pressure and the effects of being stressed for time can be detrimental to the student's exam performance. Research has shown that it can interfere with their ability to recall important information and this can lead to important steps being skipped over in the problem-solving process (Veenman et al. 2004). In my research, I wanted to see if there was a correlation between being stressed for time and the amount of extraneous information on the student's exams. Since issues in problem-solving can lead to unnecessary material being written down in their answers, my research used metacognitive strategies to show students how to efficiently get their point across.

Solutions to Verbosity/ Metacognition

A potential solution to reducing the quantity of extraneous information would be involving metacognition. Metacognition can be defined as learning about the process of learning and utilizing problem-solving strategies to perform at a higher level on examinations (Veenman & Beisuizen 2004; Lazar 2000). It allows participants to think not just about the problem but also the process of answering the problem (Bruer 1993). Metacognition skills have been shown to be lacking in novices. A novice learner is - someone who lacks experience and well-organized knowledge of the material at hand (Veenman et al. 2004). In contrast, expert learners have a strong grasp of the material and often engage in deliberate comprehensive analysis of knowledge networks before answering a problem (Chi et al. 1981; Bransford 2004). Novices often attempt to answer the problem without first coming up with a systemic methodology and jump right into explanations, which can lead to errors in reasoning and the addition of extraneous information that is not relevant (Veenman & Beishuizen 2004). Experts often think through all the possibilities first before delving into answering a problem and usually only include significant information (Chi et al. 1981; Bransford 2004). Although expert-level thinking can take years to build, there is evidence that novices that quickly pick up these metacognitive strategies and implement them into their thinking to gain competence in an area of knowledge (Daley 1999). Utilizing a theory of learning called Cognitive Apprenticeship, I aimed to teach novice level students to think more like experts by using modeling worksheets during weekly discussion sections.

Solutions to Extraneous Information-> Cognitive Apprenticeship

A theory of learning called cognitive apprenticeship instructs students to verbalize their thought processes and to learn from an expert and as an "apprentice" start to apply these steps to express more complex ideas. (Collins et al.1987; Collins et al.1988). There are 6 steps to cognitive apprenticeship: modeling, coaching, scaffolding, articulation, reflection, and exploration. Through these steps, students can begin to apply expert-level techniques and through repeated practice, eventually become independent and self-sufficient (Scardamalia 1984).

Normally the methodology of answering free-response questions is very individualized with little instruction shown in class. Students are left to learn for themselves what to write and through a process of trial and error learn what works for them. Students can only apply skill sets that they know and are comfortable with and if they don't encounter many free-response open-ended questions in their education, they could be at a loss of how to respond to it. In the literature, there's been a lack of methodologies that can instruct students on how to respond to free-response questions using expert-level thinking. I utilize cognitive apprenticeship in modeling worksheets to instruct students on how to effectively answer free-response questions in a clean, crisp approach. Students can learn how to write down only what's relevant and still receive full credit.

Further Application

For these skills to be honed and cemented into the student's psyche, they need to be refined by constant repetition and application of these skills. There are 3 steps to skill acquisition: the cognitive stage, the associative stage, and the autonomous stage (Anderson 1993). In the first stage, the students are taught the material and learn about the skill set. In the second stage, students can

apply the skill to situations and their mistakes and misunderstandings can be clarified so they could have a thorough understanding of the material. The last step is to become independent and have a solid grasp of the skill so they can perfect it until they reach the expert's level. Through the repeated application of my modeling worksheets, students can start to internalize and retain the training of being concise in order to apply these skills on their free-response exams.

MATERIALS AND METHODS

Overall Methodology

My research focused on why students add extraneous information on their free-response exams and if I could develop techniques to help students identify and eliminate unnecessary information. We compared 3 BIPN 100 Human Physiology classes in total, 1 section in Fall 2018, and 2 sections taught in Spring 2019. All these classes were instructed by the same professor and taught the same material. The Midterm 1 (MT1) and Midterm 2 (MT2) tests from each section were isomorphic with similar difficulty and complexity. The first section was Fall 2018, and we tested out a method where the Instructional Assistant (IA's) would underline the extraneous information on the exams, and then the tests were handed back to the students. This was a potential intervention to see if bringing attention to the extraneous information would have any impact on the student's behavior. We quantified the extraneous information on the Fall 2018 exams and used it as a comparison to the spring treatment and control and it was called "control for underlining." It helped us confirm that the control section in the Spring was comparable to the Fall section and that the underlining had no effect on the student's inclusion of extraneous material.

The control group was the morning section of Spring 2019, Section A. There were 151 students enrolled in the class and they were taught normally, with no mention of reducing extraneous information or underlining of extraneous information by IAs when grading. The treatment section was the other Spring 2019 class, section B with 211 students enrolled. These students had the intervention of modeling treatments that were held during the discussion section and the underlining of extraneous information in their test responses. I provided modeling

worksheets to the IAs of the treatment section and instructed the IAs to use the modeling worksheets in discussion sessions to help students create more concise answers.

After each midterm, the student's responses were quantified to see if there was extraneous material present and their test scores were recorded. A post-midterm Likert survey was given to all students asking them whether they felt pressed for time comparing the control section to the treatment section. I then interviewed 26 students (that had extraneous material present on their exams) using "Think Aloud" interviews to understand the thought process/cognitive rationale that contributes to why they include extraneous information on free-response exams. At the end of the quarter, students were given a summative assessment to test how well they were able to identify extraneous information.

Think Alouds

To understand why students include extraneous information in free-response exams, I conducted think-aloud interviews. After Midterm 1 (MT1) and Midterm 2 (MT2) were finished, a post-midterm survey was administered offering 1% extra credit to students who consented to an interview with me. Over 300 students responded to the survey in total and 26 students were picked. The students who were picked had to have extraneous information on their exams. Students from both the Spring control and treatment section were selected. Fourteen students were interviewed after MT1, 7 from the control group, and 7 from the treatment group. Twelve students were interviewed after MT2: 6 were from the control group and 6 from the treatment group.

They were asked whether they knew what extraneous information was, if they could identify it on their own test responses and why they included it. Their entire interview was recorded and then later transcribed. Their answers to why they included extraneous info on the exam were

pulled from the transcript and then binned into categories. There were over 70 responses total from the 26 students.

The process of coding the think-aloud responses was done in an iterative process. We started by separating the student's answers to why they included extraneous information into groups and further refined them into 6 main categories.

To generate IRR (Inter-Rater-Reliability), 13 interviews were analyzed individually by myself and my Principal Investigator (PI) separately. We would code and bin their responses individually into the categories and then come together to generate discussion and agreement. After 2 rounds of this, we had 42 instances of agreement, and only 1 instance of disagreement, a 97.6% IRR. These results were graphed in two ways, the proportion of students who mentioned each category and the total number of responses for each category. Examples of student's responses for each category are shown in a table.

Quantify the Extraneous Information on the Exams

To quantify the amount of extraneous material on each exam, Instructional Assistants (IA) were trained to underline extraneous information using the following definition: concepts that are not relevant to the answer to the problem and regardless if they are right or wrong, won't contribute to the point total. The control for underlining, Fall 2018, and the Spring treatment had the IA's underline, the Spring control had no underlining present. Only free-response questions were considered, multiple-choice, and fill in the blank questions were ignored. The exams were scanned and then handed back to the students.

Following this, I re-analyzed and coded each question for all sections, giving each question a score of 1 or 0 depending on whether that question contained any extraneous information. In this

way, the quantity of extraneous information was binary: it was present, or it was not present, it did not matter the amount. For example, a student might have 3 sentences that were classified as extraneous in a response and receive a score of 1 while another student might only have 1 sentence of extraneous and receive the same score of 1. This was used to gauge the proportion of extraneous information that was present on each question and for each student.

To reach a consensus on the definition of extraneous information, 20 student responses were evaluated individually by myself and my Principal investigator. Once we discussed and agreed which responses would constitute a score of 1, we had a 95% IRR. I continued grading the rest of the exams on my own.

As I went through the exams and identified what extraneous information looked like, I noticed that they fell into 3 main types (answered a question that wasn't there, information too far from what was relevant and repeated themselves), and the scores varied for each response. To determine if there was a pattern for the extraneous responses, I took 100 examples of extraneous information, 25 from MT1 and MT2 for both the control and treatment sections. This was then graphed into a table by type of extraneous and their score: mostly correct (more than 50%) or mostly incorrect (50% and below).

For Fall 2018 (control for underlining), there were 48 exams for MT1 and 45 for MT2. For Spring control there were 127 exams for MT1 and 127 exams for MT2. For Spring Treatment there were 186 exams for MT1 and 180 exams for MT2.

To evaluate the amount of extraneous information, the proportion of extraneous information was calculated. This is found by dividing the number of questions that had extraneous information in it by the total number of free-response questions on each exam, Midterm 1 and Midterm 2. The control for underlining had a total of 14 free-response questions for MT1 and 16

for MT2. For Spring control and treatment, there were 14 free-response questions on MT1 and 21 free-response questions for MT2.

The standard error of the mean was calculated for the error bars and Cohen's d for the effect size. A 2 tailed t-test was analyzed to calculate if there were any significant differences between the 3 groups.

Modeling Treatments

To see if an invention designed to reduce the amount of extraneous information could be effective, mandatory once-weekly discussions were held by the Instructional assistants (IAs) for the Spring control and treatment groups. Both sections would go over isomorphic test questions and review that week's material to help prepare the students for the midterms. In both sections, students were asked to first answer individually to the isomorphic test question and then talk about their responses in their discussion groups of 3-4 students. However, in the treatment section, the IAs (Instructional Assistants) stressed the importance of having concise and correct answers and explained ways to write succinctly. The students were encouraged to use shorthand notation and to think before they write. In the treatment section, the IAs also administered an additional worksheet with modeled responses written by the professor. The treatment section had 4 of these types of sessions, once before MT1 and 3 before MT2

On the modeling worksheet, there were 3-5 different model answers to the isomorphic question. The answers ranged from correct and concise to having varying amounts of extraneous information within them. Students were asked to assign a grade to each response and underline any extraneous information for each modeled response. They would compare their thinking to the professor's succinct answer and see what expert thinking looked like and what the rubric was

looking for. The students were asked to analyze their own responses to the exam questions (done at the beginning of the section), to see if all the information that they wrote was relevant to the question. They would then again discuss in their groups about their answers and how they could edit them to become more concise. The IA's would go over the correct answers, explain what parts were extraneous and why it wasn't relevant to answer the question. At the end of the section, they could keep the worksheets so they could reflect on it later. The IAs were trained and coached on how to become effective facilitators in weekly IA meetings. Over 80% of the students attended at least 3 of the sessions. A modeling worksheet is provided in the appendix below.

Did it improve Test Scores?

To determine whether the treatment had any effect on MT1 and MT2 test scores, the average was calculated for each section: the control for underlining, Spring 2019 control, and treatment. The average score was graphed, and the standard error of the mean was used for the error bars and Cohen's D was used for the effect size. A 2 tailed t-test was used to see if there were any significant differences between the exam scores.

Likert Survey

To assess whether students felt like they had enough time to complete the midterms, a post-exam survey after MT1 and MT2 was given. The midterms both had 4 free response questions with multiple sub-questions within them. They were given 1 hour to complete the test. The post-exam surveys were based on a 5-point Likert scale ranging from strongly agree to strongly disagree. The question that was asked: "Did you feel pressed for time on the exam?". Students' responses were subsequently further condensed into 3 categories: strongly agree/agree, neither agree nor

disagree, strongly disagree/disagree. From the control section after MT1, 69 students responded and from the treatment section, 119 students responded. From the control section after MT2, 78 students responded and 100 students from the treatment section.

Summative Assessment for Identifying Extraneous Information

To see if the modeling treatment was effective in helping students identify extraneous information in an explanation, a summative assessment was given at the end of the quarter during the discussion section. For both the control and treatment section, the students were tasked to identify and underline any/all extraneous information in 4 different modeled responses to an exam question. They had to do this individually and not discuss their answers with other students. Once they were finished, the IAs collected the worksheets and answers were discussed. The worksheets were then analyzed for extraneous statements that were correctly identified. For example, if the sentence **was** extraneous and they underlined it, that was correctly identified as extraneous. Overall there were 6 statements total in the worksheet, across the four answers. I also analyzed the number of non-extraneous statements that were incorrectly identified as extraneous. For example, if they underlined a statement that was needed to answer the problem, that was considered incorrectly identified. Overall there were 11 of these types of statements across the four answers. The proportion of correctly identified and incorrectly identified was calculated by the number of statements the students underlined divided by the total number of statements for each category. Differences between treatment and control sections were analyzed with a 2-tailed t-test, and Cohen's d-scores were calculated to measure effect sizes. This is included in the appendix

RESULTS

Why Do Students Include Extraneous Information?

To understand why students added extraneous information on their free-response exams, 26 students were interviewed using think alouds, and they verbalized 70 different responses that were classified into 6 categories. The 6 categories were: *1) unsure of the question, 2) unsure of the answer, 3) knows the answer but is trying to maximize points 4) anxious/stressed/rushing 5) taught to write a lot 6) inner monologue*. 2 examples from each category, transcribed from the interviews are shown below (**Table 1**). More than half (58%) of the students specified that they knew the answer and were trying to maximize points. The category that the least number of students mentioned was *taught to write a lot* (23%). (**Figure 1**)

Table 1: Examples of student's responses from the think aloud interviews for each category. 2 examples were featured for each category along with how many students mentioned it and how often it was mentioned

Reason Why They Included Extraneous Information	Coding Examples (verbatim from students)	Number of Students who mentioned it (n=26)	How many times it was mentioned (n=70)
Unsure of what the question was asking	Got confused about what the question wants I only write extraneously when I don't know what he wants from me, when the question is ambiguous and open-ended	10	11
Unsure of the answer	I word vomited because I didn't know if my answer was correct I don't know what he's looking for or what the answer would be so I just keep writing	8	9
Knows the answer/Maximize points	All or nothing, I write down as much as I think of so nothing gets cut off and I get all the points possible The answer doesn't come out the way I want it too so I keep writing hoping the grader would understand me, don't want to erase and potentially miss something	15	23
Anxious/Stressed/Rushing	I'm really nervous and scared of free response, I don't know what to do I was rushing so I wasn't really thinking about how to write concisely	11	12
Taught to write a lot	My professor at my community college taught me to write whatever I could to get partial credit I had professors in the past who took off points if I didn't write enough	6	6
Inner Monologue	It's difficult to put into words the thoughts I have in my head, so I just write as the answer comes to me When you're writing your thought process, sometimes the answer comes to you	7	9

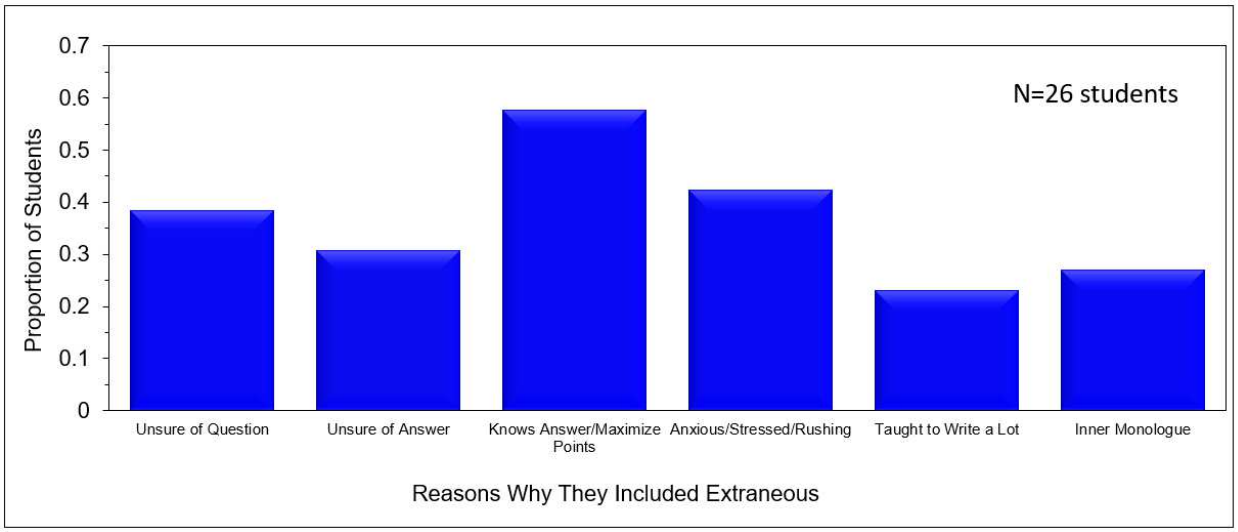


Figure 1: Proportion of the students that mentioned a category in their answers to the question "Why did you include extraneous information in this response?" For example, out of the 26 students interviewed, 10 (38%) indicated that they were unsure what the question was asking (far left).

What was the Most Frequently Mentioned Rationale for Including Extraneous Information?

Since they could repeat reasons and explain themselves, students often mentioned more than 1 reason why they added extraneous material. Out of all the 70 responses, the most prominent category (33%) was that they knew the answer and were trying to maximize points. The least prominent category was that they were taught to write a lot (8%). Knowing the answer and trying to maximize points was the most common reason for students to include extraneous and the most restated response. Overall the least popular response for both figures was that they were taught to write a lot. (**Figure 2)**

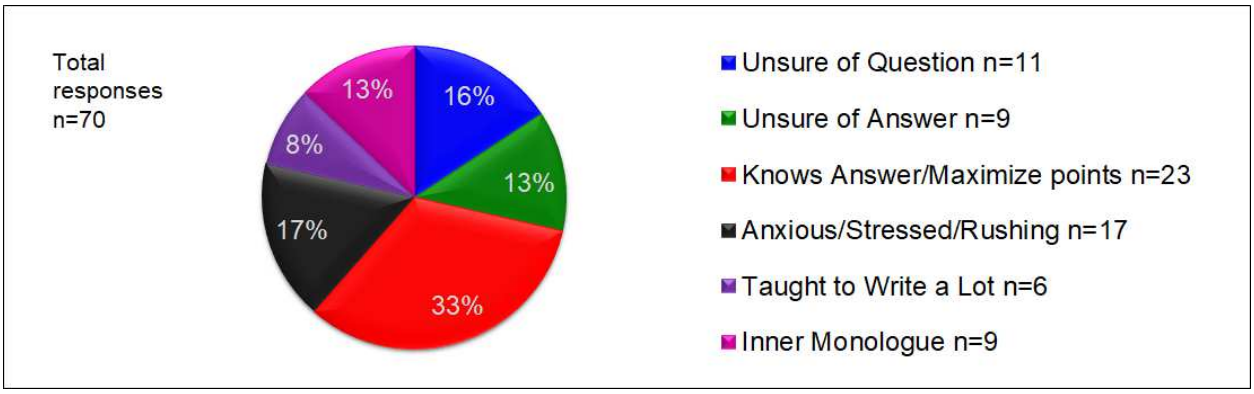


Figure 2: The percentage of responses categorized by the total number of responses given. Since there were a total of 70 responses from our 26 students, and there were 23 mentions of "knowing the answer but wanting to maximize points" here we present 23 out of 70 responses (33%).

Effects of Treatment on Proportion of Extraneous Information

To assess whether the modeling treatment was effective at reducing the amount of extraneous information present on the midterm responses, the proportion of extraneous information - out of the number of opportunities to include extraneous information - was calculated and compared across all 3 sections. For the "control for underlining" section, the proportion of responses with extraneous information for MT1 was 0.1205 (± 0.0102) and for MT2 it was 0.1264 (± 0.0123) (**Figure 3**). For control MT1 was 0.1085 (± 0.0077) and for MT2 it was 0.1059 (± 0.0071) (**Figure 3**). For treatment MT1 it was 0.0783 (± 0.0049) and for MT2 it was 0.00602 (± 0.0041) (**Figure 3**). Between the "control for underlining" and the control section, there was no significant difference for MT1 ($p = 0.3516$, Cohen's $d = 0.1512$) and MT2 ($p = 0.1565$, Cohen's $d = 0.2537$), indicating that underlining had no effect on student's inclusion of extraneous information. There was the most significant reduction between the "control for underlining" and the treatment section for MT1 ($p = 0.00041$, Cohen's $d = 0.6106$) and MT2 ($p = 0.00054$, Cohen's $d = 0.456$). There was also reduction in extraneous information between the control and treatment

for MT1 (p =0.00120, Cohen's d =0.3632) and MT2 (p =0.00010, Cohen's d =0.6576). See (**Figure 3**)

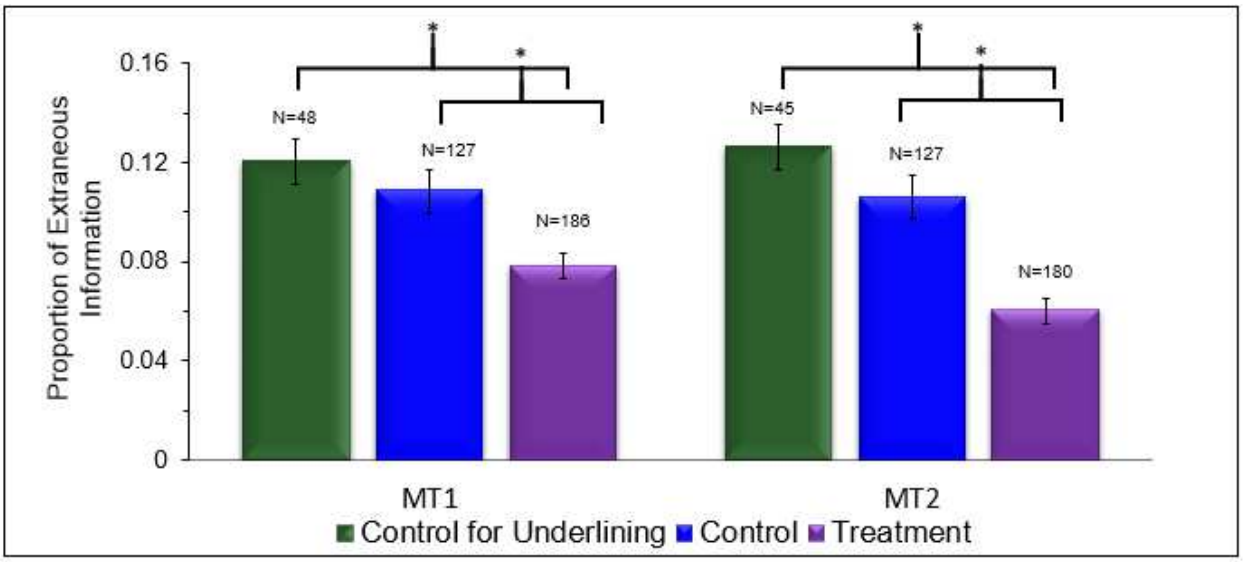


Figure 3: Quantification of extraneous information across the three sections. Proportion of extraneous information was expressed as the number of open-ended questions that had extraneous information divided by total number of open-ended questions. The proportion of extraneous in MT1 for the "control for underlining", control and treatment section was 0.1205 (±0.0102), 0.1085 (± 0.0077), and 0.0783(± 0.0049). For MT2 the proportion of extraneous information was 0.1264 (±0.0123), 0.1059 (± 0.0071), 0.0602 (± 0.0041). There was a significant difference between the "control for underlining" and treatment for both midterms (Cohen's d =0.6106 for MT1 and 0.4560 for MT2). There was also a significant *(p <0.05) reduction in the amount of extraneous information between the control and treatment section for both midterms (Cohen's d =0.3632 for MT1 and 0.6576 for MT2). This indicates that the modeling treatment was enough for lessening the quantity of extraneous information on both midterms.

Effect of Modeling Treatment on Test Scores

To see whether the modeling treatment would impact midterm (MT1, MT2) exam scores, the average scores were calculated for all 3 sections. In the "control for underlining" section, the scores for MT1 and MT2 were 69.83% (± 1.35) and 73.8% (± 2.67) respectfully (**Figure 4**). The control's section score for MT1 and MT2 was 71.68% (± 1.24) and 70.27% (± 1.32) (**Figure 4**). For the treatment section, it was 66.42% (± 1.07) for MT1 and 75.11% (± 0 .925) for MT2 (**Figure 4**). We performed a 2-tailed t-test as a statistical analysis and Cohen's d and found significant differences between control and treatment sections for MT1 ($p = 0.001543$; Cohen's $d = 0.367$) and MT2 ($p = 0.00310$; Cohen's $d = 0.352$). The treatment section did significantly worse on MT1 and significantly better on MT2. There was no difference between "control for underlining" and control for MT1 ($p = 0.3151$, Cohen's $d = 0.1565$) and MT2 ($p = 0.2404$, Cohen's $d = 0.2155$). Or between "control for underlining" and treatment for MT1 ($p = 0.0513$, Cohen's $d = 0.277$) and MT2 ($p = .6483$, Cohen's $d = .084$) conditions. See (**Figure 4**)

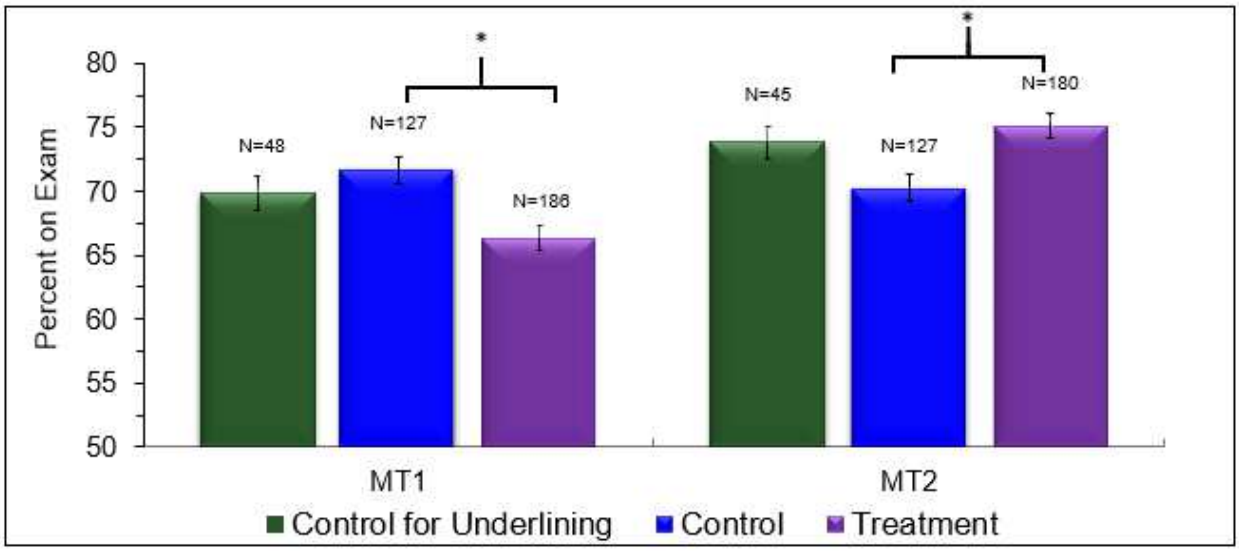


Figure 4: Mean exam scores recorded from the 2 midterms for all three sections. MT1 scores for "control for underlining", control and treatment sections were 69.83% (± 1.35), 71.68% (± 1.24), 66.42% (± 1.07) respectively. MT2 scores were 73.81% (± 2.67), 70.27% (± 1.32), 75.11% (± 0.925) accordingly. '*' denotes $p < 0.05$, two-tailed t-test. The treatment section did significantly worse on MT1 and significantly better on MT2.

Is there a Correlation between Exam Score and Proportion of Extraneous Information?

To see if there was a correlation between low (<50%), mid (50%-80%), and high (>80%) performing students and the amount of extraneous present on their exams, we plotted the amount of extraneous information for each student over their midterm scores. A line of regression was used to identify a positive or negative correlation. If there was a strong positive correlation, this would have indicated that high performing students had higher amounts of extraneous than mid or low students. If there was a strong negative correlation, this would have identified the opposite effect. The data shows that there was a weak correlation for all the sections (R^2= 0.0084, 0.0071, 0.0021, 0.008), suggesting that no one group (ie: low, mid or high performers) had much more extraneous information than the other (**Figure 5**)

Who Benefits from the Treatment?

To see whether the treatment helped the low, mid and high performing students reduce the quantity of extraneous information on their exams, we compared control and treatment conditions for MT1 (**Figure 5A, Figure 5B)** and MT2 (**Figure 5C, Figure 5D**). Overall there is a decrease in the quantity of extraneous information in both treatment sections when compared to the control. This difference is even more evident for Midterm 2 (MT2). The lower amounts of extraneous information for the treatment group are spread across all 'performance' levels, with low, mid and high performing students all having less extraneous information in the treatment condition compared to control. The data illustrates the effectiveness of the modeling treatment to reduce the amount of extraneous information in the treatment group for low, mid, and high performing students. Shown in (**Figure 5**)

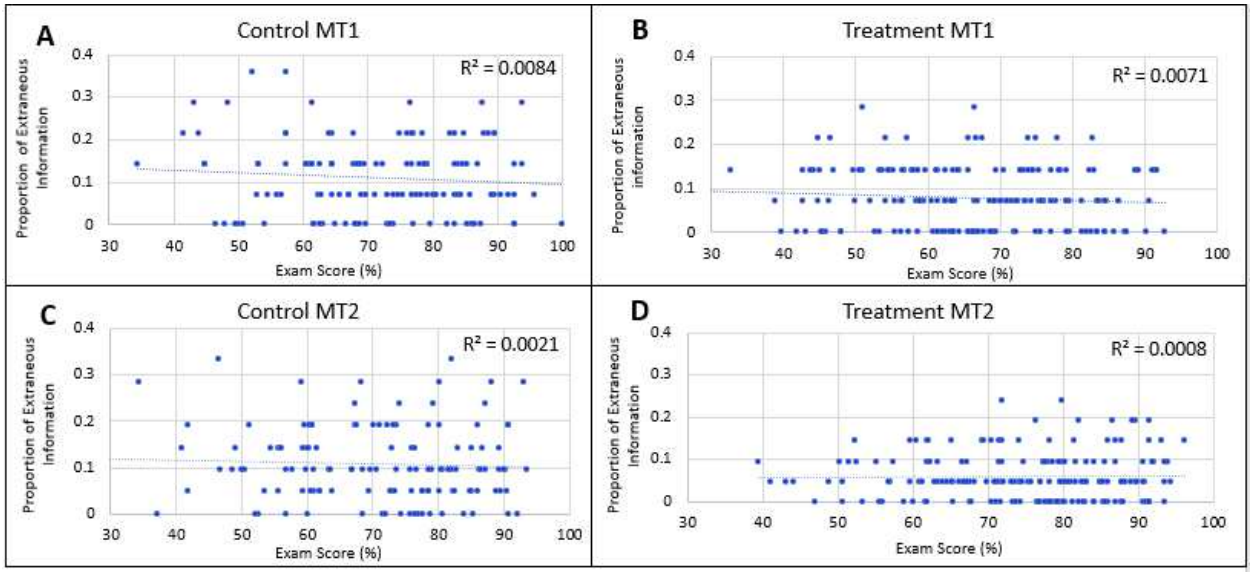


Figure 5: The graphs above (A-D) show the correlations of the proportion of extraneous information for students and their midterm exam scores by using a scatterplot. (a) Control MT1 scores, (R^2 = 0.0084) (b) Treatment MT1 (R^2 =0.0071) (c) Control MT2 (R^2 =0.0021) (d) Treatment MT2 (R^2 = 0.0008). Overall there is a reduction in the proportion of extraneous information for the low (<50%), mid (50%-80%), and high performing students (>80%) in the treatment sections for both midterms.

Did the Treatment Help Students Identify what is Extraneous and what is Relevant?

The proportion of extraneous statements correctly identified by the control section was 0.577 (± 0.0322) and the number of extraneous statements correctly identified by the treatment section was 0.767 (± 0.0197) (**Figure 6**). This difference was significant (p = 0.000234, two-tailed t-test; Cohen's d=0.135). The proportion of important statements incorrectly labeled as "extraneous" was 0.198 (± 0.0196) for the control group and 0.132 (±0.0107) for the treatment group (**Figure 6**). This difference was significant (p =0.0398, two-tailed t-test; Cohen's d = 0.257). The data shows that the treatment group was significantly better at not only underlining extraneous sentences but also correctly identifying important/relevant information. The modeling treatment helped students in the treatment group more accurately recognize extraneous and important material. This is shown in (**Figure 6**).

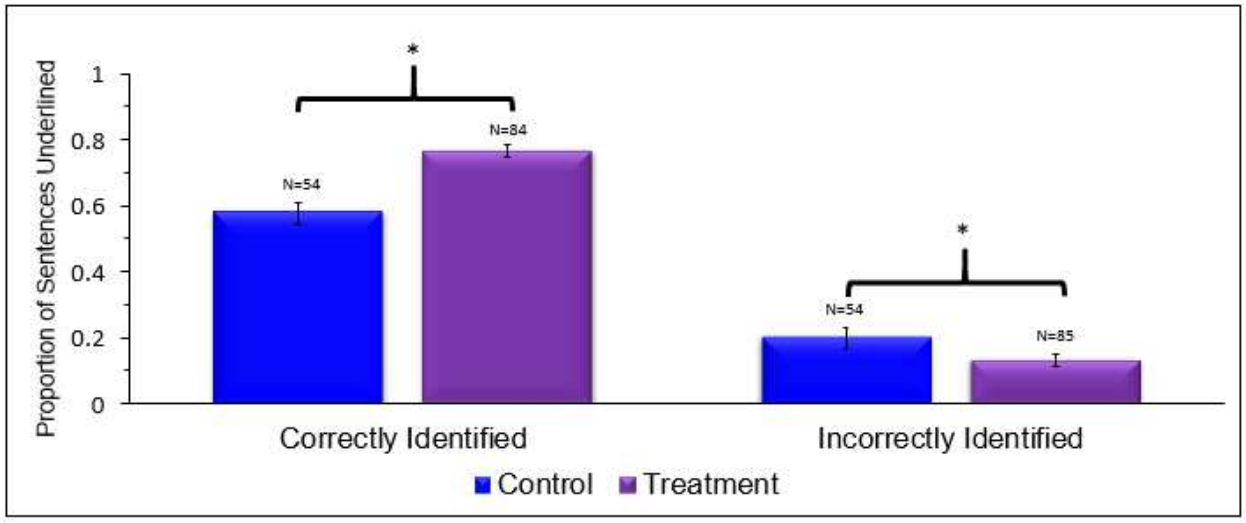


Figure 6: Student's ability to correctly identify extraneous information. Proportion of extraneous statements correctly identified for control and treatment conditions (left) were 0.577 (± 0.0322), and 0.767 (± 0.0197), respectively. Proportion of extraneous statements incorrectly identified (that is: important sentences misidentified as extraneous) was 0.198 (±0.0196) for control and 0.132 (± 0.0107) for treatment (right). The treatment group was significantly ($p < 0.05$, two-tailed t-test; Cohen's d = 0.135) better at correctly identifying extraneous statements and was significantly* ($p < 0.05$, two-tailed t-test; Cohen's d = 0.257) better at not underlining relevant sentences (incorrectly identified) as extraneous.

Categorization of Extraneous Responses

When identifying extraneous information on the student's free response answers, 3 main categories emerged: *1) answering a question that wasn't there 2) logic that was too far from what's needed 3) repeating themselves* (**Table 2).** To further determine whether the extraneous responses were mostly correct (high scores) or mostly incorrect (low scores), a distinction of more than 50% and 50% and below was made (**Table 3**)

Table 2 indicates representative extraneous examples from students and Table 3 is the categorization of 100 extraneous responses based on type and score. Out of the 100 extraneous responses that were categorized, 59% had scores that were higher than 50% (mostly correct) and more than half (54%) of them had logic that was too far than what was needed (**Table 3**). It was noted that extraneous responses taken from students were longer and had more information than was needed to answer the question. The concise answer would be worth the same amount of points and would take up half the space

Table 2: Examples of student's responses containing extraneous information compared to concise answers. Each exemplifies cases where the students either answered a question that wasn't there, logic was too far from what was needed, or they repeated themselves.

Type of Extraneous	Question on Exam	Student's Answer (with extraneous underlined)	Concise Answer
Answers a question that wasn't there	Lateral inhibition allows your sensory system to do what? (1 mark)	It allows you to discriminate where the stimulus is coming from and is found on the hands. It's due to neurons inhibiting their neighboring neurons.	Locate the stimulus with great accuracy.
Logic was too far from what's relevant	How has the absolute relative refractory period changed if we doubled the # of VG K^+ channels and why? (2 marks)	In an action potential, it starts when the threshold is reached, has a rising phase and a falling phase. The threshold and rising phase would be unaffected since we didn't change the number of VG Na^+ channels but the falling phase would be steeper because of the higher electromotive force->ends shorter than normal	The falling phase of the action potential would be steeper due to more electromotive force and the inactivation gates would open faster than before-> shorter than normal
Repeated themselves	What would happen to the # of cross bridges formed if we add a drug that slows down the closing rate of RyR and why? (2 marks)	If we slowed down the closing rate of RyR, then Ca^{2+} would flow through the slow closing rate of the RyR gates, the drug would keep the gates open longer to allow Ca^{2+} to flow out even more and all of this calcium could lead to more crossbridge which needs Ca^{2+}	More Ca^{2+} would be able to leave the sarcolemma since it's open longer and can bind to form more cross bridges

Table 3: Number of extraneous responses that were categorized based on the type of extraneous and score given. 100 examples were analyzed.

Extraneous Categories	High Scores (Higher than 50%)	Low Scores (50% and below)
Answers a question that wasn't there	19	15
Logic was too far from what's relevant	35	19
Repeated themselves	5	7

Did the Treatment Make Students Feel Less Pressed for Time?

To uncover if students felt pressed for time due to having too much extraneous material on their exam, a 5-point Likert scale was given after MT1 and MT2. The expectation was that the treatment group would feel less pressed for time since they were more concise and wrote relevant concepts on their exams. However, the data shows no significant difference between the control and treatment group (**Figure 7A and 7B**).

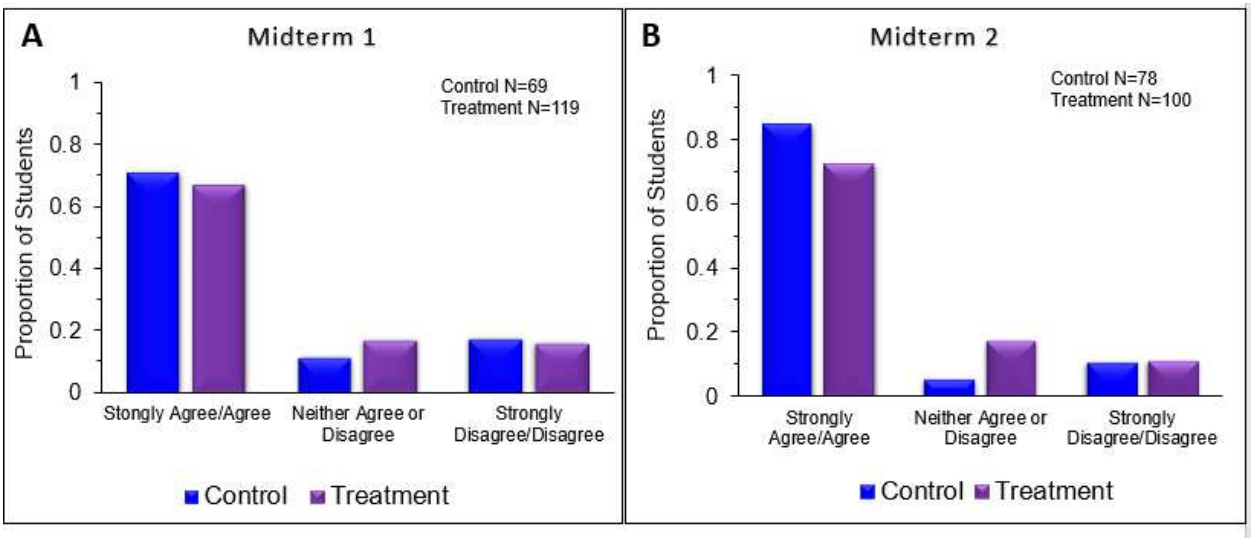


Figure 7. Results of a Likert survey administered after Midterm 1 and Midterm 2 responding to the question "Did you feel pressed for time?" (a) Student's answers after Midterm 1 (b) Student's answer after Midterm 2. "Strongly agree" and "Agree" was condensed into one group as was "Strongly disagree" and "Disagree."

DISCUSSION

Summary of findings

In this thesis, I first characterized why students included extraneous information in their answers, using think alouds to develop six categories. My data found that the majority of all the students cited that they knew the answer and were trying to maximize the points.

We then used the cognitive apprenticeship approach of modeling to train students to recognize extraneous information and to reduce its incorporation into their responses. I measured multiple variables such as *1) the proportion of extraneous information 2) exam scores 3) feeling pressed for time 4) who benefits most from the treatment 5) the ability to identify extraneous information.* Since the Fall 2018 "control for underlining" and the Spring control had no significant differences, this indicates that both of these classes are comparable, and underlining was not effective at reducing the quantity of extraneous material in the student's responses. So, I focused my comparison between the Spring control section (no underlining or modeling worksheet) and Spring treatment section (modeling worksheet).

I sought to see if there was a connection between exam scores, feeling pressed for time, and the inclusion of extraneous material. I hypothesized that students felt that time was limited because they were writing too much extraneous information and as a result, did poorer on the exams. My findings show that the treatment section had lower exam scores on midterm 1 and higher exam scores on midterm 2. Consequently, no conclusive evidence can be drawn for the modeling treatment helping exam scores. Similarly, there was no significant difference between the control and the treatment group feeling less pressed for time after answering a post-exam Likert scale survey.

On the other hand, my hypothesis that my modeling treatment would be effective at showing students how to write more concisely was supported by a significant reduction in the proportion of extraneous information on MT1 and MT2 for the treatment section.

Additionally, I assumed that higher-performing students would have greater amounts of extraneous information in their responses. The findings established that there was no correlation between the proportion of extraneous information and their exam score, which indicates that all groups of students equally include extraneous information. Furthermore, when analyzing who benefited the most from the treatment, the data showed overall that students in each performance group showed a decline in the proportion of extraneous in their exams.

At the end of the quarter, the data showed that the treatment section was significantly better at identifying what extraneous information was when it was presented to them. This reinforces my hypothesis that the modeling worksheets not only taught students to write more concisely but also trained them on how to recognize information as either relevant or extraneous.

Finally, I was one of the first studies to group extraneous information into categories (by what type and by what score) on free-response exams. My findings showed that the majority of the responses, regardless of the score on the problem, was due to answering with logic that was too far from what's relevant. As well, the bulk of the extraneous scores were found on responses that were given a score of over 50% (mostly correct). These findings allowed me to categorize what extraneous information is.

Extraneous Information

While a predecessor had used the term "extraneous information" in their studies (Nedjat-Haiem, 2019), other fields have defined communication that hinders the clarity of the message

by adding too many words as "verbosity" (Hsia 1977; Kuo 1995; Arbuckle et.al 2000). My study adds onto Nedjat-Haim's research who first identified and quantified extraneous information on free-response exams. I continue by coding extraneous information and classifying it into different categories based on type and score. Coding is used by many researchers to find patterns and to extract data from qualitative data (Cambell et al. 2013, Ongena et al. 2006).

Since previous research found that adding as much information as possible was a strategy to answer free-response questions (Nedjat-Haim et al. 2019) I wanted to understand why by utilizing think alouds. Other researchers have used think alouds to understand student difficulties while solving structure-function questions (Halmo et al. 2018) and to have students verbalize their thought process to answer genetics problems (McDonnell & Mullally 2016). I used think-alouds to help understand the student's rationale for including extraneous material in their responses.

Cognitive Apprenticeship

My treatment focused on the use of modeling worksheets that were based on the theory of cognitive apprenticeship. Apprenticeship itself is an age-old technique that has been used for centuries for experts to impart the knowledge and practice needed for their apprentices to pick up their trade (Lave and Wenger 1991). Cognitive apprenticeship is a stepwise procedure where experts first model and instruct students on techniques and skillsets. Then they move into scaffolding exercises while offering assistance and practice for the novices, also known as the apprentice. (Collins & Brown 1989). Using this method, researchers were able to help students who were poor readers improve their reading comprehension scores from 15% to 85% accuracy (Palincsar & Brown 1984). It's an effective and commonly used technique, especially in the

medical industry where residents shadow doctors and learn hands-on experience from them (Lajoie 2009). Our results suggest that cognitive apprenticeship can also be used to help students write more effectively by reducing extraneous information.

Think Alouds and Categorization of Extraneous Information

The results that my think alouds generated were novel, my study one of the first to identify the rationales that motivate students and divide them into subgroups. In addition, the findings that I got from categorizing the types of extraneous information were corroborated by my think-aloud data. For example, most of the extraneous information was found on responses that got over 50% (mostly correct). Likewise, the most frequently stated reason from the think alouds was that they knew the answer and were trying to maximize points. Both of these findings could be explained by students knowing the answer and writing everything that they thought could be relevant to get the highest possible score. They know something about the concepts, but not enough at an expert's level to be able to write concisely.

Correlation between Exam Scores and Proportion of Extraneous

Since most extraneous data was found on responses that received scores over 50%, I presumed that higher-scoring students would also have a greater proportion of extraneous material. Contrary to expectation, the scatter plot showed that all subgroups of students, low, mid, and high performing had extraneous information present, with no group showing a preference. This demonstrates the pervasive issue of extraneous material; it can affect all types of students.

My findings for BIPN 100 (Human Physiology) contradict what another class BIPN 134 (Human Reproduction) found. Their data showed a negative correlation between test scores and the proportion of extraneous information, suggesting that students who have higher test scores have a lower proportion of extraneous information (Nedjat-Haim et al. 2019). This implies that the correlation between exam score and proportion is extraneous is variable. Furthermore, BIPN 134 had two-fold the amount of extraneous information on their midterms which indicates that extraneous information is contingent on the subject matter, how the test questions are worded and even the students who attend the class. This demonstrates that further experiments in other fields would need to be tested to see if this trend holds true.

There were also differences between how my study and previous research approached an intervention for extraneous information. In order to help students, reduce their quantity of extraneous information, I built my research on pre-existing literature that the use of metacognitive strategies could change student behaviors (Nedjat-Haim et al. 2019). The use of exam wrappers which are post-exam reflection assignments, altered students' free-response test-taking strategies from "writing as much as possible" to "only including what was important" (Nedjat-Haim et al. 2019). I wanted to try another metacognitive treatment that could help students not only understand what is important to include, but also lessen the frequency of extraneous information in their answers. I used a modeling treatment that focused on cognitive apprenticeship that could help progress students from novice to expert level thinking. The exam wrappers only helped higher-performing students reduce their extraneous information while my modeling treatment assisted ALL students achieve concise answers, regardless of their exam score. A potential explanation is that the lower performing students need something more

explicit, training designed to walk them through how to write a concise answer rather than just going through it by themselves.

Modeling Worksheets

My data supported my hypothesis that the modeling worksheets were a useful tool in guiding students to write more succinctly because the treatment section had significantly less extraneous material on their midterms. This reduction was more pronounced on MT2 than on MT1. The number of treatments that had occurred before the student took MT1 was only 1 and the students had 4 treatments before MT2. This indicates the importance of repetition and reiteration for students to successfully learn a new skill. Not only that but when the summative assessment worksheet was handed out to the students, the treatment group did far better at underlining and identifying the extraneous statements than the control group.

Novice to Expert-Like Thinking

A possible explanation for both findings could be that the students have learned how to progress from novice to expert-like thinking. Experts are able to classify and "chunk" related concepts to each other and see the bigger picture (Daley 1999 et. al, Bereiter et al.1988, Lajoie 2009). Additionally, experts often use self-checking strategies to fine-tune their solution (Daley 1999 et. al, Bereiter et al.1988). Both are incredibly valuable skills that allow experts to write out concise answers. These skills take time and practice to develop properly and cognitive apprenticeship provides a pathway and the support needed to instill new skills. (Lajoie 2009). I designed my modeling worksheets that were handed out during discussion around the 5 steps of cognitive apprenticeship: *1) Modeling 2) Coaching 3) Scaffolding 4) Articulation 5) Reflection*

(Collins et al. 1987, Collins et al. 1988). Students were shown what concise "modeled" answers looked like, along with examples of poorly written answers, generated directly by the professor. During the discussion, they had access to previous rubrics and were coached on techniques on how to write concisely such as planning out the answer before they write and focusing on what the question was asking. The knowledge was scaffolded further by the students working independently to practice writing out concise answers to free-response questions. Finally, to further cement the skills, they would discuss their answers in small groups and would have to articulate why a concept was relevant and why they included it. All of this helped students evolve from novice level thinking to expert level thinking. Even though both the treatment and control sections were both taught the same class concepts, the addition of the modeling worksheet helped the treatment section truly grasp what extraneous information really is.

Summative Assessment

In order to include less extraneous information, students must first understand WHAT extraneous information is. To assess whether our treatment helped students understand what extraneous information was, we used the summative assessment at the end of the quarter. Summative assessments are a method of evaluation usually at the end of the course that measures a student's understanding of a topic and is graded using a rubric (Haber and Mitchell 2017). Similarly, the IA's gave out a worksheet that the students in both the treatment and control section would have to complete individually. The students were tested on their ability to correctly recognize and underline extraneous information that was present in modeled responses to a free-response question. This supports and complements why I saw a reduction of extraneous material on the treatment's MT1 and MT2 free response answers. For my treatment to be

successful, first the students must have a better understanding of what extraneous material is. The students how to identify it when presented in front of them but most importantly they need to be able to identify it in their own writing. They then must make the conscious choice to not include it in their writing. My findings affirmed that the modeling treatment accomplished its goal, the treatment section was better at recognizing and underlining extraneous information than the control section.

Time Pressure and Exam Scores

My final hypothesis was that if I could reduce the amount of extraneous present on their exams, there would be an improvement in test scores and a corresponding decrease in feeling pressured for time as a result. Contrary to the expectation, even with a significant reduction of extraneous material in their responses, there was no improvement of test scores in the treatment section. This could be because the treatment aimed to show students how to write only the relevant concepts, not how to retain the material better or test any better. In this study, students do not get penalized for having extraneous information present, so even though they are diminishing the amount, it's not being reflected in their test score. While our treatment lowered extraneous information in students' responses, The Likert scale survey showed that it did not alleviate their perception of perceived time pressure. A potential reason why the treatment did not help the students feel less pressed for time is that regardless of how much or how little the students write, the feeling of being pressed for time can stem from other reasons. These can include test anxiety, the difficulty of the questions, and even the length of time given to take the test. These are all independent factors that my research did not focus on.

Alternative Hypothesis

The data supports my hypothesis that my modeling worksheets rooted in cognitive apprenticeship, would improve student's ability to not only write concise answers but also identify what extraneous information is. However, previous interventions such as exam wrappers (post-exam reflection assignments) also resulted in students becoming more concise. (Nedjat-Haim et al. 2019).

Since both involve metacognition, this suggests there's an alternative hypothesis, that it's metacognition that has the potential to help students communicate in writing. Arguably, anything that allows students to contemplate and reflect on their answers to become more concise will be effective. Possibly students don't take much time to ruminate on their answer before they write them out, and this could be causing the inclusion of extraneous information. Or even after they write out their responses, they might go back to check their answers for correctness but not conciseness. Metacognitive activities are unique in the way that they teach individuals how to gain an awareness and an understanding of their thought process (Lazar 2000). It allows students the ability to shift perspective and change from just answering a question to being critically aware of what's needed to answer the question (Veenman 2004).

My modeling worksheets could be just one of many potential treatments to help students learn how to write more concisely. A way to test this alternative hypothesis would be to try out multiple metacognitive treatments that prompt students to focus on what the question is asking and compare which one gives the best results. Other interventions could be points taken out if irrelevant information is present, remind students in the question that the answer should only be 1-2 sentences, give keywords to mention in the answer, etc. This could lead to the creation of a

new study that is focused on instructing students on how to interpret, understand, and respond concisely to free-response questions.

Limitations of my Study

A possible factor limiting this study is the sample size of the think-aloud interviews, only 26 students were interviewed, and 70 responses were transcribed. There could be other potential rationales for why students include extraneous information on free-response exams outside of the 6 categories I got.

One of the possible critiques of my study is how I define extraneous. My definition of extraneous information are concepts that are not relevant to the answer to the problem and regardless if they are right or wrong, will not contribute to the point total. Another instructor might define it using a different definition; however, I would argue that something that contributes nothing of value to an answer is extraneous in all fields.

An unintentional byproduct of being too concise could lead to students losing points in another class because their professor deemed, they didn't write enough to get full credit. This is a difficult problem to fix due to the variability of instructors and how they want their class to be run and what they are looking for in an answer. Nonetheless, this did not seem to be a problem in our study. Students in the treatment section showed the most significant reduction in extraneous information while at the same time, their performance on Midterm 2 was higher than the control. It remains to be seen if this is the case for other classes so a follow-up study would have to be done to evaluate the effect of being concise on student's exam scores.

Given that my study only focused on an undergraduate biology class, the types of extraneous categories that emerged might just be unique to this class. For example, the most

common type of extraneous information that appeared was “answering a question with logic that is too far from what’s needed.” That data could be skewed to just this class since many of the questions are pathway-related and the students often started too early in the pathway. As a result, their responses were classified as extraneous. A future follow-up experiment that could be explored is to repeat this in another class or even a different field and see if the same results emerge.

Conclusion/Future follow up

Free-response questions are a powerful tool in the arsenal that professors can employ in their ways to assess a student's comprehension of the material. However, since students can struggle in understanding what and how much to write to get full credit, they can often over-explain and waste time writing out those concepts. My modeling treatment is effective at not only reducing the amount of extraneous information on the student’s answers but also helped them identify what extraneous material is.

I propose this could be an effective approach that professors could use in their classes to teach students what is important to articulate in their answers to receive full credit. Even in classes that lack IA’s or discussion sections, the instructor can set aside time to do a brief exercise where the students identify what relevant concepts would be important to add and learn what the instructor expects to see on their exams. This can be beneficial in large classroom settings, small group discussions, or even one on one sessions where students could learn how to assess their own writing and learn how to respond efficiently to free-response questions. It would make both parties satisfied, the students wouldn’t be floundering in the dark, wondering what to write while the graders of the question wouldn’t have to wade through unnecessary sentences.

One of the reasons why I sought to reduce the amount of extraneous material on free-response exams, was so that grading can be less of a labor-intensive process if the answers were more succinct. I did not record the time it took to grade the student's responses in the control and the treatment section, but a follow-up study can measure if there are any significant differences. Another unanswered question that could be followed up is a longitudinal study to see if the students in the treatment section will continue to apply the skills of being concise to other classes and future endeavors. The idea of being concise is not just important for classes and getting good grades but has practical uses. It is applicable for scientific writing, submitting to journals, and even writing grants and proposals. Extraneous information is a problem that instructors are quite familiar with, but my research has shown promise and hope that it can be mitigated with just a little bit of training.

Material from this thesis is coauthored with Nanea Dominguez, Matthew Nedjat-Haiem and James Cooke. The thesis author was the primary author of this chapter.

REFERENCES

1. Ackerman, T., & Smith, P. (1988). A Comparison of the Information Provided by Essay, Multiple-Choice and Free-Response Writing Tests. The American Journal of Nursing, 17(6), 539. https://doi.org/10.2307/3405502

2. Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35–44. https://doi.org/10.1037/0003-066X.48.1.35

3. 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. https://doi.org/10.1037/0882-7974.15.1.65

4. Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of free□response and multiple□choice items. *Journal of Educational Measurement*, *28*(1), 77-92

5. Bereiter, C., Burtis, P. J., & Scardamalia, M. (1988). Cognitive operations in constructing main points in written composition. Journal of Memory and Language, 27(3), 261–278. https://doi.org/10.1016/0749-596X(88)90054-X

6. Blackey, R. (2009). Society for History Education So Many Choices, so Little Time: Strategies for Understanding and Taking Multiple-Choice Exams in History. In Source: The History Teacher (Vol. 43, Issue 1). https://www.jstor.org/stable/40543353

7. Bransford, John D. *How People Learn: Brain, Mind, Experience, and School*. National Acad. Press, 2004.

8. Brockriede, W., & Ehninger, D. (1960). Toulmin on argument: An interpretation and application. Quarterly Journal of Speech, 46(1), 44–53. https://doi.org/10.1080/00335636009382390

9. Bruer, J. T. Schools Of Thought: A Science of Learning in the Classroom. Cambridge, Mass.: MIT Press, 1993.

10. Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. Sociological Methods and Research, 42(3), 294–320. https://doi.org/10.1177/0049124113500475

11. Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

12. Collins A., and Smith, E.E. (1982). "Teaching the Process of Reading Comprehension." In D.K. Detterman and R.J. Sternberg (Eds.), How Much and How Can Intelligence Be Increased? Norwood, NJ: Ablex.

13. Collins, A., Brown, J. S., & Newman, S. E. (1987). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics (Technical Report No. 403). BBN Laboratories, Cambridge, MA. Centre for the Study of Reading, University of Illinois. January 1987.

14. Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship. *Thinking: The Journal of Philosophy for Children*, *8*(1), 2-10.

15. Daley, B. (1999). Novice to Expert: An Exploration of How Professionals Learn (pp. 133–147). Adult education Quarterly.

16. Glaser, R. "The Maturing of the Relationship Between the Science of Learning and Cognition and Educational Practice." Learning and Instruction, 1991, 1, 129-144

17. Grais, R. (2008). The Verbosity Epidemic. Science, 320(June), 1718–1720.

18. Haber, N., & Mitchell, T. N. (2017). Using formative & summative assessment to evaluate library instruction in an online first year writing course. *Journal of Library & Information Services in Distance Learning*, *11*(3-4), 300-313.

19. Halmo, S. M., Sensibaugh, C. A., Bhatia, K. S., Howell, A., Ferryanto, E. P., Choe, B., Kehoe, K., Watson, M., & Lemons, P. P. (2018). Student difficulties during structure–function problem solving. Biochemistry and Molecular Biology Education, 46(5), 453–463. https://doi.org/10.1002/bmb.21166

20. Haynie, W. J. (1994). Effects of Multiple-Choice and Short-Answer Tests on Delayed Retention Learning. Journal of Technology Education, 6(1), 32–44. https://doi.org/10.21061/jte.v6i1.a.3

21. Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the Test...or Testing to Teach: Exams Requiring Higher Order Thinking Skills Encourage Greater Conceptual Understanding. Educational Psychology Review, 26(2), 307–329. https://doi.org/10.1007/s10648-013-9248-9

22. Krifka, M. (2011). Be brief and vague! And how Bidirectional Optimality Theory allows for verbosity and precision. Sounds and Systems. https://doi.org/10.1515/9783110894653.439

23. Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from avionics and medicine. *Development of professional expertise:*

Toward measurement of expert performance and design of optimal learning environments, 61-83.

24. Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

25. Lazar, S. (2000). Metacognitive aspects of test-taking and intelligence. Psychologische beitrage

26. Lissitz, R. W., Hou, X., & Slater, S. C. (2012). The Contribution of Constructed Response Items to Large Scale Assessment. Journal of Applied Testing Technology, 13(3), 1–52.

27. Mcdonnell, L., & Mullally, M. (2016). Teaching Students How to Check Their Work While Solving Problems in Genetics.

28. Nedjat-Haiem, Matthew. "Helping Students Write More Effectively; An Investigation of Student Test-Taking Strategies." *EScholarship, University of California*, 25 June 2019, escholarship.org/uc/item/2bb7k11s.

29. Nemoto, T., & Beglar, D. (2014). Likert-scale questionnaires. In *JALT 2013 Conference Proceedings* (pp. 1-8).

30. Ongena, Y., & Dijkstra, W. (2006). Methods of behavior coding of survey interviews. Journal of Official Statistics, 22(3), 419–451.

31. Palincsar, A.S., and Brown, A. L. (1984). "Reciprocal Teaching of Comprehension-fostering and Monitoring Activities." Cognition and Instruction, 1, 117-175.

32. Scardamalia, M., and Bereiter, C., and Steinbach, R. (1984). "Teachability of Reflective Process in Written Composition." Cognitive Science, 8, 173-190.

33. Someren, Maarten W. van. *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. Academic Press, 1994.

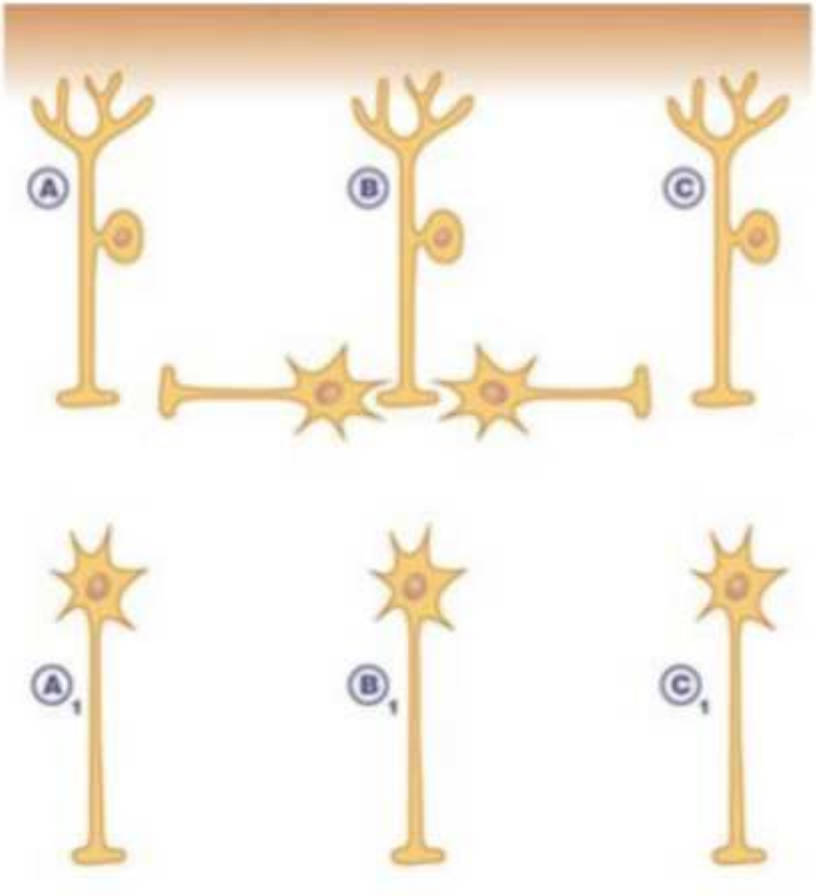
34. Veenman, M. V. J., & Beisuizen, J. J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. Learning and Instruction. https://doi.org/10.1016/j.learninstruc.2004.09.004

APPENDIX

The first worksheet below is an example of an answer key for the modeling worksheet given during discussion only for the treatment section. The students would first have to answer the question on their own, without having access to the modeled responses. Then they would be given the modeling worksheet and asked to underline any extraneous information (**shown underlined**) and give each response a score. They could compare their explanation that they wrote earlier to these examples. Afterward, they would discuss in small groups (3-4) about what a "good" concise answer is and explain why they underlined certain sentences. The Instructional Assistants facilitating the meeting would go over the answer key and explain what concepts were relevant and what constituted full credit and why the red sentences were extraneous (***shown italicized***).

The second worksheet is the answer key for summative assessment given at the end of the quarter, during the Week 10 discussion section. The worksheet was given to both the control and treatment sections. Students were tasked to identify what extraneous information was and to underline it (shown in red) individually. The worksheets were then collected and correctly identified statements (**i.e. the sentence was extraneous, and they underlined it**) and the incorrectly identified statements (**i.e. the sentence was not extraneous, and they underlined it**) was quantified. This was used to see if there was an improvement in the treatment section to identify and recognize what extraneous information is.

Question 6 (5 marks). Use the diagram below to answer questions a) and



Question: Imagine a stimulus is delivered in the middle of the receptive field of the sensory neuron in the middle (cell B), but that the stimulus is outside of the receptive field for the two surrounding sensory cells (cells A and C). That is: we can expect cell "B" to release neurotransmitter, but NOT cells A and C. What will happen to the rate of action potential firing in afferent neurons **A1** and **C1**? Explain your answer. (3 marks)

For each of the 5 student responses below, please **give a grade out of 3** (you can give part Marks if you like). For each answer, **underline any extraneous information** you find (that is: information that is provided but does not add anything to the students' response).

Answers:

1) Cells A1 and C1 will have no action potentials because cells A and C did not release any neurotransmitter onto them, so nothing is causing them to have action potentials.
Action potentials are generated when an incoming graded potential has caused the membrane potential to pass the threshold.

[incomplete and some extraneous. Would be complete if said why cells A and C aren't releasing NT] I'd give this about a 1.5 or 2 out of 3.]

2) Cells A1 and C1 will have no change in their APs produced. Since A and C were not activated by any stimulus, they will not release any neurotransmitter onto cells A1 and C1. Without neurotransmitter release, there is no graded potential that will generate an action potential in these cells.

[inaccurate information - they've forgotten about the interneurons!] I'd give this about a 0.5 out of 3. At least they know about neurotransmitters causing post-synaptic events.

3) Release of NT from cell B will stimulate interneurons, which inhibit the release of NT from Cells A and C. These cells (A and C) have an intrinsic firing pattern at rest, so when their presynaptic termini are inhibited, their 'baseline' NT release is stopped. The baseline firing properties are likely a result of several factors: they might have fewer leak K+ channels (which would both depolarize the cell toward Ena, and increase Rm), and they might have an AP threshold that is lower (eg: -60 mV instead of -50 mV).

[the second half is all extraneous. The first part is great, BUT: they don't actually answer the question! I'd give this a 2 or a 2.5 out of 3.]

4) AP firing in A1 and C1 will decrease b/c NT release from A1 and C1 will be decreased by the interneuron. The interneuron is activated by cell B. Cells A and C fire APs (low level) when at rest, so release small amts of NT onto A1 and C1. Less NT from A and C = fewer APs for A1 and C1.

[full marks; nice and concise]

5) The APs of cells A1, B1, and C1 should be subthreshold at rest since APs of A, B, or C are all subthreshold at rest and cannot stimulate A1, B1, and C1. However, the large stimulus of B should inhibit A's stimulation of A1 and C's stimulation of C1 because the intensity of B is so great. IF the stimulation of the skin had spread to A and C, then they would have increased their action potential firing, possibly overcoming the inhibition that occurred as a result of the interneurons.

[*this is pretty bad. They don't actually answer the question; the first sentence doesn't make sense to me; they do mention inhibition of A and C, so they should probably get 0.5 or even 1, but that is generous.]*

Summative Assessment

Please read the question below and think of your own answer. Then, please read the 3 answers that follow. For each answer, please think of a grade out of 3 that you would give it, and please **underline** any/all information that you think is extraneous. For this purpose, extraneous information is any information that did not contribute anything of value to the response. That is: if you removed it, the score on that question would not change, even if the statement is factually accurate (or inaccurate!).

Question: Alcohol inhibits the release of ADH (vasopressin) by inhibiting calcium channels. Compared to someone who has not ingested alcohol, what would the osmolarity of the filtrate be like at the end of the collecting duct, and what would this do to water reabsorption? Explain, make sure to mention aquaporins. (3 marks)

Answers:

1. Without ADH, it will be impossible for the body to reabsorb any water. ADH is released from the posterior pituitary in response to reduced blood pressure. When ADH binds to ADH receptors, a cAMP cascade is initiated which leads to the insertion of aquaporins into the cells of the collecting duct. These aquaporins allow water to move from the filtrate to the interstitial fluid, where it is collected by the vasa recta. If this person drank alcohol and had less ADH, it would not be possible to reabsorb any water. They would become very dehydrated and their blood pressure would probably go way down.

[this answer provides some extraneous information and doesn't actually answer the question completely. The question asks about the osmolarity in the filtrate, which is not mentioned. They also offer some factually inaccurate information: they say that reabsorption will not

occur, but that is misleading: there will still be water reabsorbed at the levels of the PCT and the loop of Henle. However, they are correct that reabsorption of water at the collecting duct will go down. Since they omit reference to the filtrate osmolarity and are wrong about overall reabsorption, I'd give this a 1 or possibly a 1.5 out of 3.]

2. When ADH levels are suppressed, the amount of reabsorption of water occurring in the collecting duct will decrease because there won't be as much / any aquaporins due to the cAMP pathway not being activated by ADH. It would be harder for water to move from the filtrate to the interstitial fluid. As a result, the osmolarity of the filtrate will be much lower than someone who isn't drinking alcohol.

[I don't think this is missing anything. They've included the reason for the ADH-mediated effects (aquaporins), and there is virtually no extraneous info. I'd give it a 3/3]

3. When we decrease the amount of ADH, it is going to be trouble. The ADH activates ADH receptors in the cells of the collecting duct, which does the following: Gs -> cAMP -> PKA -> aquaporins getting inserted into the membrane of the cells of the collecting duct. This allows water to move between the interstitial fluid and the filtrate. Water is only moving passively because there are no active mechanisms to move water. So water is going to move to wherever the osmotic gradient is higher. In this case (no ADH), because we can't have aquaporins, the osmolarity of the filtrate is going to be way higher than the interstitial fluid. So water wants to be in the filtrate. Since less water leaves the filtrate for the interstitial fluid, the amount of water reabsorbed is going to be less. Also, it is important to note that the ADH is released from the posterior pituitary.

[Some factual errors about the osmolarity of the filtrate should lose a point. They are correct about the amount of water being reabsorbed, but for totally wrong reasons. I'd give this a 1 / 3

max. Also: lots of extraneous info here.]