2016

Am I Able to Predict How I Will Do? Examining Calibration in an Undergraduate Biology Course

Trisha A. Turner
University of Kentucky, trishadouin@gmail.com
Digital Object Identifier: http://dx.doi.org/10.13023/ETD.2016.138

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Trisha A. Turner, Student

Dr. Ellen Usher, Major Professor

Dr. Kenneth Tyler, Director of Graduate Studies

Am I Able To Predict How I Will Do?
Examining Calibration In An Undergraduate Biology Course

THESIS

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in the College of
Education at the University of Kentucky

By

Trisha Anne Turner

Lexington, Kentucky

Director: Dr. Ellen Usher, Associate Professor of Educational Psychology

Lexington, Kentucky

2016

ABSTRACT OF THESIS

AM I ABLE TO PREDICT HOW I WILL DO?
EXAMINING CALIBRATION IN AN UNDERGRADUATE BIOLOGY COURSE

Students who are self-regulated are more likely to succeed academically, whereas students who have deficiencies in their learning have been recognized as having a lack of metacognitive awareness (Valdez, 2013; Zimmerman, 2002). If students are metacognitively unaware in large introductory courses, they may have difficulty knowing when to self-regulate and modify their learning (Lin & Zabrucky, 1998; Stone, 2000). One manner in which researchers have assessed students' metacognitive awareness is by asking students to estimate how they think they will do on tasks compared to their actual performance, known as calibration. The purpose of this study was to examine students' calibration and study habits. Participants were undergraduates (N = 384) in an introductory biology course at a southeastern U.S. university. Students completed four surveys that assessed their exam score expectations and the study habits they used prior to each exam. Results showed that students' estimates are most discrepant from their actual performance early in the semester and become more accurate at the end of the semester. A closer look at students' study habits revealed that the inaccuracy of students' exam judgments showed little connection to the study strategies that students used. Findings from this study are important for biology instructors.

Keywords: Calibration, Biology, Self-Regulation, Metacognition

Trisha Turner

Student's Signature

4/29/2016

Date

AM I ABLE TO PREDICT HOW I WILL DO?

EXAMINING CALIBRATION IN AN UNDERGRADUATE BIOLOGY COURSE

By

Trisha A. Turner

_______Ellen Usher____________________

Director of Thesis

________Kenneth Tyler_________________

Director of Graduate Studies

__________April 21st 2016______________

Date

Table of Contents

List of Tables iv
List of Figures v
Chapter 1: Introduction 1
Zimmerman's Theory of Self-Regulation 1
Figure 1. Zimmerman's Cycle of Self-Regulated Learning 6
Chapter 2: Literature Review 7
Calibration and Self-Regulation in Science Courses 9
Statement of the Problem 12
Purpose of the Study 12
Chapter 3: Method 15
Participants 15
Procedure 15
Expected exam scores. 15
Actual (earned) exam scores 16
Discrepancy scores. 16
Previous biology experience 17
Proportion of days spent studying. 17
Study tools. 17
Analyses 18
Chapter 4: Results 22
Chapter 5: Discussion 39
Limitations and Future Directions 45
Appendix A - IRB Consent Form 49
Appendix B - Exam One Survey 51
Appendix C- Exam Two Survey 52
Appendix D - Exam Three Survey 54
Appendix E - Final Exam Survey 55
References 56
Vita 60

LIST OF TABLES

Table 1. Description of Study Participants 21
Table 2. Descriptive Statistics for All Exam Discrepancy Variables 28
Table 3. Fit Statistics for Latent Class Growth Analysis Models..29
Table 4. Analysis of Variance for Exam 1 Differences by High School Biology Course.............30
Table 5. Correlations Between Absolute Value Exam Discrepancy Scores and the Proportion of Days Studied 31
Table 6. Study Tools Used For Each Exam..32
Table 7. Correlations Between Raw Exam Discrepancy Scores and Study Tools Used...............33

LIST OF FIGURES

Figure 1. Zimmerman's Cycle of Self-Regulated Learning..6
Figure 2. Frequency of Overestimators and Underestimators...34
Figure 3. Latent Class Growth Analysis Class One 35
Figure 4. Latent Class Growth Analysis Class Two 36
Figure 5. Latent Class Growth Analysis Class Three 37
Figure 6. Mean Exam 1 Grades by Previous Biology Class 38

Chapter 1: Introduction

Being able to pass introductory courses is one of the first steps towards completing an undergraduate degree. Many introductory courses are known as "gatekeeper" courses. These gatekeeper courses are often large and difficult for new undergraduate students. However, success in these courses is necessary for progress towards one's degree. Unsuccessful performance in these courses may force students to change their major or require them to repeat the course. Successful completion of these courses requires that students are skilled in their study habits and have awareness for how well their study habits work in a specific context.

When students lack awareness of how well their study habits work, they may overestimate or underestimate their future or past performance (Bembenutty, 2009). Students who make incorrect estimates of their performance are often referred to as being "poorly calibrated." The goal of this thesis is to examine undergraduate students' self-regulation in an introductory biology class. In this thesis, self-regulation is being examined by asking students about their metacognitive awareness for their academic achievement and by examining students' self-regulatory study habits in the course.

Zimmerman's Theory of Self-Regulation

Self-regulated learning is a process by which students develop goals, select strategies, and monitor their performance in order to complete specific tasks (Zimmerman, 2008). Students who are self-regulated in their approach to studying monitor their achievement and study tactics and modify them when necessary (Stone, 2000; Winne & Jamieson-Noel, 2002). Effective self-regulation requires metacognitive awareness or an ability to think about one's own thinking and learning (McCabe, 2011).

Metacognition is crucial to the learning process. Successful learners must plan, organize, self-instruct, motivate themselves, self-monitor, and self-evaluate (Zimmerman, 2002). Doing so enables learners to implement and modify strategies as necessary to ensure mastery of content.

This cycle of self-regulated learning has been described by Zimmerman (2002). Zimmerman provided a theoretical framework for self-regulatory processes that encompassed three key stages (see Figure 1). These three stages are the forethought stage, the performance stage, and the self-reflection stage. He proposed that these three stages should be viewed in terms of a cyclical process. The cyclical nature of the self-regulation process allows learners' actions at each individual stage to influence the other stages. For example, learners' actions during the self-reflection stage may alter the way they approach the forethought stage (Zimmerman, 2002).

The forethought stage includes two major processes that occur before learning: task analysis and self-motivation beliefs. Task analysis refers to the self-regulatory processes of planning and goal setting. In a college biology course, planning and setting goals for studying are important skills. Self-motivation beliefs deal with learner beliefs towards the learning process. These beliefs may include self-efficacy, outcome expectations, learning goal orientation, and intrinsic value/interest (Panadero & Alonso-Tapia, 2014; Zimmerman, 2002).

The performance stage processes occur during learning. This stage includes the major processes of self-control and self-observation. Self-control processes are often strategies that were selected in the prior stage. Self-control strategies include self-instruction, focusing of attention, task strategies, and imagery. Self-control strategies can be especially important in undergraduate learning. Zimmerman (2002) defined self-control as "the deployment of specific methods or strategies that were selected during the forethought phase" (p. 68). Finding a quiet place to study for college exams, using word association to learn biological concepts, and forming images to remember biological concepts are all examples of self-control. Self-observation is a process in which learners observe their own progress. A learner may record the time it takes to complete a task or time spent studying. Being aware of the time that it takes to study for courses can be a helpful strategy for college students. Learners may also self-experiment by studying with or without a partner to determine the better study method. This example of self-experimentation is known as self-monitoring (Zimmerman, 2002).

The self-reflection stage occurs after the performance has occurred. The two major self-reflection stage processes are self-reaction and self-judgment. The self-reaction processes are those that encompass the adaptive and defensive attitudes as well as self-satisfaction feelings related to the learning event (Panadero & Alonso-Tapia, 2014). Learners' self-reaction can influence their self-efficacy and effort during the forethought stage of the next learning event. The self-judgment process is when the learner reflects on the learning event through either self-evaluation or a causal attribution process (Zimmerman, 2002). For example, a student might evaluate her performance in comparison to her prior performance or against an external standard of performance.

Metacognition plays a central role in each of the three stages previously listed. Metacognition is defined as "the awareness of and knowledge of one's own thinking" (Zimmerman, 2002, p. 65). For example, a student must be aware of her own knowledge and study habits before she can recognize that she may need to implement different self-regulatory strategies. Being able to modify and monitor learning is an important strategy for undergraduate students. Without the ability to monitor progress and modify study habits as necessary, learning may become increasingly challenging. Metacognitive awareness is therefore an important factor when considering the process of self-regulation (Schraw, Crippen, & Hartley, 2006).

One manner in which researchers have assessed students' metacognitive awareness is by asking students to estimate how well they think they will do on important tasks. Researchers then calculate the difference between students' estimations and their actual performance. The discrepancy between expected and actual performance can be viewed as an indication of a student's *calibration* (Lin & Zabrucky, 1998; Stone, 2000). The ability to accurately evaluate one's understanding of information is a critical step in the learning process. Calibration is widely used to examine the accuracy of individuals' self-assessed comprehension (Lin & Zabrucky, 1998). According to Brown, Roediger, and McDaniel (2014), "Testing is not only a powerful learning strategy, it is a potent reality check on the accuracy of your own judgment of what you know how to do" (p. 72). In other words, testing and receiving feedback can prompt metacognitive awareness in the learner.

Bembenutty (2009) found that calibration is an essential part of the metacognitive process and has the power to influence task completion and academic success. Calibration fits within both the forethought and the self-reflection stages of Zimmerman's (2002) model. Students evaluate their performance in the self-reflection stage, and this may subsequently affect the outcome expectations and goals that students set for themselves in the forethought stage. In the following sections I will review findings from research that has examined the role of calibration in college courses, laboratory settings, and within science courses.

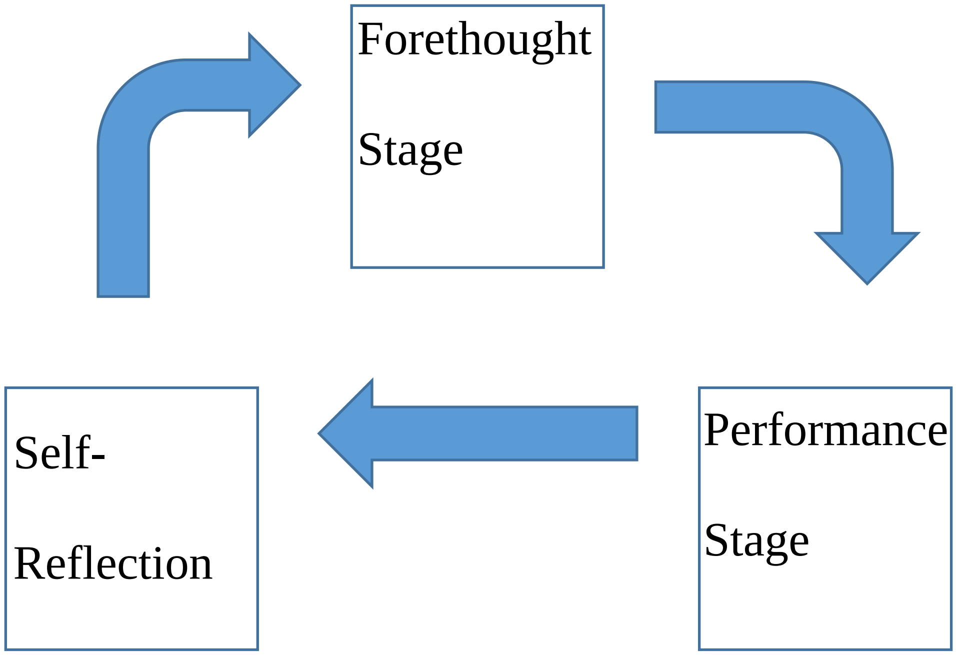

Figure 1. Zimmerman's Cycle of Self-Regulated Learning

Chapter 2: Literature Review

In this section, I will review findings related to calibration, how it has been measured, and how it has been investigated in connection with students' learning and studying. Calibration has been measured in many different ways. Some researchers have attempted to measure this construct using specialized software, such as eye tracking and study habits software. For example, Winne and Jamison-Noel (2002) measured calibration with tracing software. The participants in their study were undergraduates ($N = 69$) at a western Canadian university. Students were given multiple passages to read. While reading, they were also instructed to use tracing software that kept track of all of their study tactics. After the students read the passage and used the software, they estimated their use of study tactics. This process was then immediately followed by a short exam. Students' calibration scores were calculated by subtracting their estimated achievement from their actual achievement. Overall, students were found to overestimate their total achievement. Results revealed that students overestimated their use of their study tactics.

Working from the understanding that calibration is a part of the self-regulation process, Hadwin and Webster (2013) examined the self-regulation processes as described by Zimmerman (2000, 2008). The authors hypothesized that self-regulated learners engage in a cyclical process of self-regulation that involves monitoring their progress towards their goals. When learners' self-evaluations of their progress towards a goal are unaligned with their desired standards, it signals the need to regulate learning by adjusting strategies, goals, or plans (Hadwin & Webster, 2013; Winne & Jamieson-Noel, 2002). To examine their hypotheses, Hadwin and Webster measured calibration in first year undergraduate students. Calibration was calculated using students' confidence judgments and evaluations of the previous week's goal. Confidence judgments were comprised of students' judgments of whether they could attain their weekly goal. Hadwin and Webster found that judgments of confidence were better calibrated with self-evaluations of learners' current goals rather than their past goals. This means that students' confidence judgments were more aligned with the goals that they were currently setting rather than the goals that they had previously set. Students did not become better calibrated over the nine weeks. Their confidence judgments were not aligned with their evaluation of the previous week's goal. However, students did slightly decrease how overconfident they were in their ability to attain their goal. Lastly, Hadwin and Webster found that students who were performing better at their university were more calibrated students.

Calibration has been measured in other ways as well. Some researchers have tried to measure calibration by having students rate their levels of confidence on individual items. This process involves a researcher administering questions one at a time to participants, having participants answer each question, and then immediately asking participants how confident they are in their response for each item. Dinsmore and Parkinson (2013) examined calibration by having students rate their levels of confidence on individual achievement items. The authors examined a sample of undergraduate students (N = 72) who were either inaccurate calibrators or accurate calibrators (defined according to the low or high proximity between confidence ratings and performance) to determine what factors they may consider when making judgments about their performance. They found that students took into account many different factors (e.g., prior knowledge, guessing) when making their judgments.

Lastly, some researchers have asked students to report their estimated grade on an assignment or quiz after they have already completed it (i.e., a post-assessment). For example, Valdez (2013) measured calibration, or as he referred to it "absolute accuracy" (p. 143). Valdez asked undergraduate communication disorder majors ($N = 24$) to rate each question in terms of how confident they were in their answer on each of their six quizzes. Students' confidence ratings were compared to their actual performance on each of the quizzes. Valdez found that students who were better calibrated (their confidence ratings and actual performance scores were closely aligned) performed better on their each of their quizzes and on their final exam.

Why do students tend to make overconfident judgments about their performance? According to Garavalia and Gredler (2002), students who are poor calibrators may deny any valid negative feedback that they receive and will therefore fail to take the corrective actions that are required. Garavalia and Gredler operationalized calibration as expected course grades (measured on a pretest) compared to earned course exam grades. They used a sample of undergraduate students ($N = 69$) in a health science course. The authors found that learners who were more accurate in their calibration earned significantly higher course grades than students who were poorly calibrated.

Calibration and Self-Regulation in Science Courses

Why would science classes be of specific interest when studying self-regulation? VanderStoep, Pintrich, and Fagerlin (1996) found that natural science courses may promote motivation and self-regulated learning differently than do other disciplines (e.g., the frequent use of multiple choice exams in natural science courses, the high task value that is assigned to courses like these, and the hierarchical learning that occurs in natural science courses). The authors therefore suggested that it is important to examine self-regulation in a domain-specific setting.

Examining self-regulation in a domain-specific setting, Lopez, Nandagopal, Shavelson, Szu, and Penn (2013) investigated self-regulatory strategies that undergraduate chemistry students (N = 89) most often utilize in science classes. They found that students most often utilize reviewing-type study strategies when studying for exams. They also found that many common study strategies were unrelated to students' final course grades. Conversely, the authors reported that students rarely engaged in metacognitive strategies (e.g., planning study schedules, quizzing yourself, and making notes from resources), even though these strategies are related to self-regulation and awareness.

Zimmerman and Paulson (1995) suggested that self-monitoring, a metacognitive strategy, leads to more accurate judgments of performance. Self-monitoring involves tracking of one's own time spent studying, recording specific study tools used, and noting earned test grades. Schraw, Crippen, and Hartley (2006) suggested that working in groups, using practice exams, and self-checking are important study strategies that increase the metacognitive awareness in science students.

How do self-regulation and calibration interact specifically in science courses? Zusho, Pintrich, and Coppola (2003) found that, over time, students' confidence in doing well in a chemistry class decreased. Students were less likely to believe that chemistry was important or useful, and there was a significant decrease in motivation as students progressed through the semester. The results suggested that there was a pronounced decline in motivation for the low achieving students. High achieving students reported an increase in self-efficacy toward the end of the semester. Students' use of rehearsal and elaboration strategies decreased over time whereas use of self-regulatory and organizational strategies increased.

Only a handful of studies have measured calibration in science courses. Specifically, few have measured calibration by examining students' expected exam scores before or after important exams. This way of examining calibration may offer an improvement over examining calibration on an individual question-by-question level, because it allows for a real world judgment of performance similar to what students do when preparing for exams. When students prepare for exams, they may set goals for their overall grade rather than how they will perform on a specific question. An exam-related rating may be a more important to measure calibration in this context. For example, Jensen and Moore (2008) measured calibration by having students complete surveys on which they predicted their final grades in a biology course immediately following each exam. Students were asked to report their expected grades at multiple time points during the semester to determine if calibration changed across the semester. Students' grades were converted to a 1.0 (low) - 4.0 (high) scale. Results showed that almost all students anticipated earning a grade of 3.6 or higher at the beginning of the course. Whereas their average estimations decreased at the end of the semester, lower performing students (those who actually earned a C or below in the course) still overestimated their final grade. Higher performing students (those who earned an A or a B) underestimated their final grades.

Past research has examined students' self-regulation in science, but few studies have examined how exam score calibration in the context of natural science courses might be related to self-regulatory strategies (VanderStoep et al., 1996). Moreover limited research has focused on metacognition and self-regulatory strategies in the context of large introductory biology courses. This thesis study fills an important gap in the literature by addressing students' pre-exam expectations in large introductory science courses (Jensen & Moore, 2008).

Statement of the Problem

According to Zimmerman (2002), "Self-regulation is important because a major function of education is the development of lifelong learning skills" (p. 66). Students who are self-regulated are more likely to succeed academically and likely to view their future with a sense of optimism, whereas students who have deficiencies in their learning have been recognized as having a lack of metacognitive awareness (Valdez, 2013; Zimmerman, 2002). Students who are poorly-calibrated are metacognitively unaware of what they know and may be at risk academically (Lin & Zabrucky, 1998; Stone, 2000). If students enrolled in large science courses are metacognitively unaware, they may have difficulty selecting appropriate study strategies, modifying their study habits, or meeting the learning objectives in the course.

Purpose of the Study

The purpose of this study is to examine students' calibration and study habits, both key aspects of students' self-regulated learning, in a large introductory biology course. This study will also address how students' self-regulatory engagement (i.e., calibration and study habits) changes across the course of one semester. Students often have to be metacognitively aware of their learning before they can self-regulate their learning. If students are not metacognitively aware then they are unlikely to be able to accurately predict their grades or to successfully engage in self-regulatory strategies and modify their learning (Lin & Zabrucky, 1998; Stone, 2000).

The research questions guiding this investigation are:

1. How do biology students' expected exam scores compare to their actual exam scores throughout one semester?
2. How do students' discrepancy scores (i.e., the difference between their estimated and actual performance) change across the semester?

Hypothesis 1 and 2: Based on the previous findings in the literature that students overestimate their exam scores during the beginning of the semester and decrease their overestimations at the end of the semester (Jensen & Moore, 2008), I expect to find that students' expected and actual exam scores will be most discrepant on the first two exams but will less so across the semester.

3. How do students' scores on Exam 1 vary as a function of students' high school biology background?

Hypothesis 3: I anticipate that grades earned on Exam 1 will depend on the biology background students had in high school. If students took advanced placement or honors biology in high school they will perform better on Exam 1 compared to students who took regular biology. According to Dinsmore and Parkinson (2013), students take into account prior knowledge when they make calibration judgments. Additionally, Zimmerman's (2002) model of self-regulation suggests that students who are more proficient engage in more effective self-regulation and self-observation processes compared to those students who are novices.

4. What is the relationship between the number of days studied (proportionately) prior to each exam and students' discrepancy scores?

Hypothesis 4: For the relationship between number of days studied and students' discrepancy scores, I do not have a specific hypothesis. The number of days studied may increase students' discrepancy scores because this may lead students to become familiar with the material and have a sense of overconfidence about how much they know. On the other hand, the number of days studied may decrease students' discrepancy scores because the more someone studies the material, the more they will become aware of what they do and do not know.

5. What is the relationship between the type of study tools used prior to each exam and students' exam-specific discrepancy scores?

Hypothesis 5: I expect that students who use study tools that require active participation (i.e., practice exam), will be more accurate in their discrepancies; in other words, they will have a more accurate estimate of how they will perform. I expect that study tools that rely on surface learning (i.e., reviewing the PowerPoint, reading notes) will be related to scores' that are more discrepant. Brown et al. (2014) suggested that by engaging with material students develop a deeper understanding of their knowledge.

Chapter 3: Method

Participants

Data were collected during the spring 2015 semester from 384 undergraduate biology students who were enrolled in two sections of an introductory biology course at a southeastern university in the U.S. The primary instructors for each section were different. Students were primarily White (72%), in their first year of college (40%), and mostly women (60%) (see Table 1 for demographics). This project was proposed and approved by the Institutional Review Board in the fall of 2014. Students were required to provide written consent to be included in the study (see Appendix A).

Procedure

Over the course of one semester, students were invited to participate in seven surveys. The first and final surveys were administered electronically. The second survey took place during the first week of class and was administered on paper immediately before the first quiz. The four remaining surveys, also administered on paper, were distributed prior to each of the four exams in the course. Data from brief paper surveys, administered prior to the four exams, were included in this study. Each of these paper surveys was completed immediately prior to the exams in the course. Exam proctors distributed the survey to the students. Students were given time to complete the survey prior to beginning their exam. Each of the four pre-exam surveys can be found in Appendices B, C, D, and E, respectively.

Expected exam scores. Students were asked to provide their expected exam score, on a percentage scale from 0-100, before the start of each exam. The question read as follows: “Please write the numeric score (one number from 0-100) that you expect to get on this exam.”

Actual (earned) exam scores. Students’ exams were collected and graded via scantron. Students’ grades on each exam were provided by the biology instructors at the end of the semester. Possible scores on Exam 1, Exam 2, and Exam 3 ranged from 0 to 100. Possible scores on the Final Exam ranged from 0 to 150. Final Exam scores were divided by 100 and converted to percentages so they could match students’ expected Final Exam score.

Discrepancy scores. Two types of discrepancy scores were calculated, *raw discrepancies* and *absolute value discrepancies*. When examining students’ averages and using the raw discrepancy score (that considers direction), averages may make the miscalibation of students appear smaller than when they are an absolute value so both were included in this study.

Raw discrepancy scores were computed by subtracting each student’s expected exam score from her corresponding actual exam score. Mean raw discrepancy scores were calculated for each exam. A mean raw discrepancy score with a positive sign indicates that a student expected a higher grade compared to her actual exam grade, meaning that the student overestimated. A mean raw discrepancy score with a negative sign indicates that the student had a lower expectation compared to her actual earned exam score, meaning that the student underestimated.

Absolute value discrepancy scores were created by turning the four raw discrepancy scores into absolute value scores. The raw discrepancy score was calculated by subtracting the expected exam score from the actual exam score. The corresponding positive and negative signs were then removed to allow the numbers to be an absolute value. Including an absolute value discrepancy score is important for interpretation in this study. Using this value allowed me to examine how "off" or discrepant students were from being perfectly calibrated rather than looking at the direction of their miscalibration. Mean absolute values discrepancy scores were calculated for each exam. A higher mean absolute value indicates that, on average, students' were more "off," or more discrepant, from their actual score. A lower mean absolute value indicates that, on average, students were closer in their estimations to their actual earned score.

Previous biology experience. Previous biology experience was measured on the Exam 3 survey. Students were asked, "What kind of high school biology did you take?" with a list of responses as follows: *Advanced Placement*, *Honors Biology, Regular Biology,* and *Other*.

Proportion of days spent studying. On each of the four paper surveys administered, students were asked to draw an "X" on a calendar to indicate every day prior to their exam that they studied (see Appendix B, C, D and E). Each calendar was modified to represent the possible study days between each exam. Each time a student indicated that he studied on one day, he received a "1." The total number of days studied prior to each exam was then calculated. Because the number of days between exams varied, I created a proportion variable by calculating the proportion of days studied prior to each exam (ratio of the sum of days studied/total number of possible study days).

Study tools. Each of the four pre-exam surveys asked students to indicate which study tools they may have used prior to each exam by selecting them from a given list. This list of study tools was created by consulting with the biology professors about all available materials that students were given. Additional study tools were added to the list after a pilot data collection. All pilot data were examined, and study tools that were most often listed by students were added to the list of study tools. Possible study tools were listed as follows: PowerPoint presentations, textbook, deep learning/effective study strategies handout, reviewing the unit learning objectives, clicker questions, Echo 360 recordings, studying in a group, in-class notes, practice exam, unit study recommendations, online resources, and other.

Analyses

To compare students' expected exam scores and their actual exam scores, I ran descriptive statistics to examine the minimum value, the maximum value, the means, and standard deviations of each discrepancy score. I also calculated the absolute value discrepancy scores for each exam and examined these same statistics.

Not only did I want to compare students' expected and actual exam scores, I wanted to identify how students' discrepancy scores changed across the semester. To do this I used the raw discrepancy variable that was created for Research Question 1. I then conducted a repeated measures analysis of variance (ANOVA) to compare the raw discrepancy score means for each exam to one another (Boekaerts & Rozendaal, 2010). With this approach, time acts as the independent variable. This allows for raw discrepancies to be compared across time where measures, specifically exam estimations, have been repeated.

I also used latent class growth analysis to examine how students' raw discrepancy scores changed across the semester. Previous findings by Jensen and Moore (2008) found that students' calibration changed across a semester. To examine how different students' calibration might change across a semester, I used latent class analysis to group students according to how they changed across a semester. Students were grouped into latent classes according to best fit as determined by fit statistics for each model. Using the Bayesian Information Criterion (BIC), Adjusted BIC, Akaike Information Criterion (AIC), Log-likelihood, entropy, and relative size of each latent class, I compared how many latent classes best fit the data. AIC measures how well a specific model fits the data in comparison to other models. BIC and adjusted BIC also provide comparative model fit data, but this measure is affected by sample size (Danner & Toland, 2013). Entropy values near one indicate high certainty in classification (Ram & Grimm, 2009). The groups that are formed can then be examined to see how they may change across the semester. This technique has been used in previous research to identify groups over the course of semester (Rice, Ray, Davis, DeBlaere, & Ashby, 2015).

Although examining how students' discrepancy scores change across the semester was an aim of this study, I also wanted to examine what factors may be related to the exam grades that students earned. Dinsmore and Parkinson (2013) found that students use their prior knowledge when studying for and taking exams. I examined the relationship between students' Exam 1 scores and their self-reported biology background (a categorical variable) using an analysis of variance. Only Exam 1 was used in this analysis because I hypothesized that previous biology background would only have a relationship with the first exam. This relationship was hypothesized because students' previous biology courses would likely be the only reference point that they have in formal biology instruction. To run this ANOVA, I used Exam 1 scores as the dependent variable and students' groups of varying biology background as the independent variables. Students were classified into three different biology backgrounds: honors biology students, AP biology students, and regular biology students.

Next, I investigated whether the number of days that students studied prior to each exam was related to how discrepant students were in their exam estimations. The resulting correlations would reveal if the proportion of days studied was related to students' level of miscalibration. To examine this relationship, I conducted a Pearson's correlation between the proportion of days studied prior to each exam and the students' absolute value discrepancy score on the same exam.

Lastly, I sought to examine the relationship between the tools that students report using to study and their raw discrepancy scores. To examine this relationship, I conducted Spearman's correlations between each study tool used and the students' raw value discrepancy scores on each exam. A Spearman's correlation was used due to the categorical nature of this data. Twelve study tools were used in this analysis. The correlations conducted were between the study tools used for each specific exam and each exam raw discrepancy variable. The resulting correlation coefficients would indicate if individual study tools were related to a student's calibration.

Table 1.

Description of Study Participants

Demographics	***N* = 384**
University Class Designation	
Freshmen	176
Sophomore	89
Junior	31
Senior	23
Missing	65
Gender	
Women	238
Men	122
Missing	24
Race/Ethnicity	
African American	16
Asian / Pacific Islander	25
Hispanic / Latino	11
White	290
Middle Eastern	5
Other	9
Missing	28
First Generation Status	
One of my parents graduated college	85
Both of my parents graduated college	140
Neither parent completed a four year degree	77
Missing	82
Prior Biology Experience	
Regular biology	143
AP biology	89
Honors biology	102
Other	10
Missing	40

Chapter 4: Results

To examine how students' actual exam grades compare to their expected exam scores, I ran descriptive statistics. I also used descriptive statistics to report students' actual exam scores on average, expected exam scores on average, and the frequency of students that were overestimators and underestimators on each exam over the semester (see Figure 2).

First, I examined how students performed on average for each of the exams. Exam 3 had the lowest average score of all four exams (*M* = 70.44, *SD* = 18.46). The Final Exam had the highest average score (*M* = 82.87, *SD* = 11.00). I also examined the average estimated scores for each exam (see Table 2). Students on average estimated that they would do the best on Exam 1 (*M* = 82.89, *SD* = 7.87). For Exam 3, estimations were the lowest (*M* = 80.24, *SD* = 9.17).

I next examined how accurately calibrated students were for each of the four exams. To do this, I examined each of the absolute value discrepancy scores (see Table 2). The largest mean absolute value discrepancy was 13.89 (*SD* = 11.96) points for Exam 3. This means that on average, students were 13.89 points away from being perfectly accurate (i.e., zero discrepancy) in the grade they expected compared to what they received. The smallest mean discrepancy was for the Final Exam, at 8.25 (*SD* = 7.87) points. The mean absolute value discrepancy for Exam 1 was 11.45 (*SD* = 10.96) points and for Exam 2 was 11.36 (*SD* = 9.46) points.

Not only did I want to examine how far off students were from zero discrepancy, I also wanted to examine the average direction of their inaccuracy. I therefore used descriptive statistics to examine each of the raw discrepancy scores (see Table 2). For

Exam 1, the mean raw discrepancy was + 6.44 points (*SD* = 14.49). For Exam 2, the mean raw discrepancy was + 5.58 points (*SD* = 13.70). The largest mean discrepancy was for Exam 3, at + 9.53 points (*SD* = 15.67). In other words, students on average, overestimated their scores on Exam 3 by 9.53 points. The smallest mean raw discrepancy score was for the Final Exam, at -1.39 points (*SD* = 11.32). This means that students, on average, underestimated their Final Exam scores by 1.39 points.

Using raw discrepancy scores, I examined the frequency of students who overestimated or underestimated their scores on each exam. To do this, students raw discrepancy scores were examined. Students who had a raw discrepancy score of +1 or higher were grouped into the overestimation category for a specific exam. Any student who had a raw discrepancy score of -1 or lower was grouped into the underestimation category. Students were put into categories for each of their exam discrepancy scores. Exams 1, 2, and 3 had more students who were classified as overestimators than who were classified as underestimators. The Final Exam had more students who were classified as underestimators than who were classified as overestimators (see Figure 2.)

I next examined whether students' tendency to over or underestimate their scores differed significantly from one exam to the next. I therefore conducted a repeated measures ANOVA to compare the mean raw discrepancy scores at each exam time to one another. There was a statistically significant effect of time on exam discrepancies, $F(3, 618) = 48.575$, $p < .001$, $d = 0.41$. I next examined pairwise means with post-hoc *t* tests using the Bonferroni correction, and found significant differences between all exam discrepancies ($M_{diff} = 9.947$, $-0.766 \leq d \leq 0.277$), except between those for Exam 1 and Exam 2 ($p = .99$, $d = -0.05$).

I next used latent class growth analysis to examine how students' discrepancy scores changed across the semester. For students to be included in this analysis they had to have completed at least two of the paper surveys that were administered. Students who had not completed at least two surveys were excluded from this analysis. Students were grouped into latent classes according to best fit as determined by fit statistics in Mplus. To determine the appropriate number of classes for the data I examined the BIC, AIC, and Log-likelihood for all models. I also examined the entropy. An entropy value when near one indicates high certainty in classification (Ram & Grimm, 2009). By examining these statistics, I determined that the appropriate number of classes to fit the data was three; all of the classes were linear. For the three-class model the fit statistics were, BIC (10126.13), AIC (10078.76), and Log-likelihood (-5027.38). Fit statistics for all models can be found in Table 3. The entropy for this model was (.761), indicating that the model was moderately high in certainty that this is the correct number of classes to best fit the data. The first class consisted of 22 people, the second of 213, and the third of 148. The first group, which I labeled as "consistent overestimators," began the semester by overestimating their scores and continued this pattern throughout the semester (see Figure 3). The second group, referred to as "accurate calibrators," began the semester hardly overestimating and ended the semester in the same way (see Figure 4). The third group, "improving calibration," started the semester by overestimating their scores, but by the end of the semester was better calibrated (see Figure 5).

After examining how students' discrepancy scores change across the semester, I next examined how students' initial exam scores might be related to their previous biology courses. I examined how each group of students performed on Exam 1. Those students who were previously enrolled in honors biology earned, on average, 78.58 points on Exam 1. Similarly, students who were previously enrolled in AP biology earned, on average, 77.91 points on the first exam. Students who reported previously being in a regular biology class earned a 73.19 on average for Exam 1 (see Figure 6).

To examine Exam 1 grade differences based on previous biology courses, I ran a one-way ANOVA. Results indicated a significant difference between Exam 1 grades based on students' previous biology course, F (3, 340) = 3.95, $p < .01$. Using the Bonferroni post-hoc t test, I found a significant difference between those students who took honors biology and regular biology in high school (see Table 4). Those who took honors biology (M = 78.58, SD = 14.92) performed higher on Exam 1 compared to those who took regular biology (M = 73.16, SD = 15.73, p = .040, d = -0.352. I found no differences between those who took AP biology (M = 77.91, SD =14.95) and those who took either honors biology, p = 0.99, d = 0.45, or regular biology, p = .135, d = -0.308.

To further investigate factors that might be related to students' metacognitive awareness, I next examined the relationship between number of days studied and students' discrepancy scores. I did not have a specific hypothesis for this research question. One possible outcome could be that students who spend more time with material may be more aware of their knowledge and performance. I first examined descriptive statistics to find out how many days students reported studying for each exam. Students spent the most time studying for Exam 2 at 25% of the days available. This percentage represents average number of days students indicated that they studied divided by total number of days available between Exam 1 and Exam 2. Students studying on 17% of the days available before Exam 1 and the Final Exam. Students spent the least proportion of days studying for Exam 3, with an average of 11%.

Next, I conducted a Pearson's correlation between proportion of days studied prior to each exam and students' absolute value discrepancy scores on each respective exam. Only proportion of days studied for the second exam and proportion of days studied before the final exam were significant and positively related to students' absolute value discrepancy scores (see Table 5). The proportion of days that students studied for Exam 1 and Exam 3 were not related to students' absolute value discrepancy scores on those exams.

However, some interesting patterns emerged when these correlations were examined across exams. The proportion of days studied for Exam 2 was positively related to students' absolute value discrepancy scores on Exams 1, 2 and 3. In other words, students who studied more for Exam 2 tended to overestimate their grades on Exam 1, 2, and 3. Finally, the proportion of days studied for the Final Exam was positively related to students' discrepancy scores on Exams 1 and 3 but not related to discrepancy scores on Exam 2 or the Final Exam. In other words, students who were more discrepant in their estimations on the first and third exam reported studying more for their Final Exam.

To examine the possibility that discrepancy scores inform study skills and vice versa, the tools students reported using to study and their relationship to discrepancy scores was investigated. I began by examining the frequency of study tools used prior to each exam. The study tool that was used most frequently for each exam was the practice exam. Ninety-one percent of students reported using the practice exam to study for Exam 1. Other study tools that were frequently used prior to each exam were reviewing in class notes (80% of students reported using this approach prior to Exam 1) and reviewing clicker questions (68% of students reported using this approach prior to Exam 1; see Table 6).

To examine the relationship between study tools and students' discrepancy scores, I conducted a Spearman's correlation between study tools used and each of the students' four raw discrepancy scores. No study tools were found to be correlated with all four exam raw discrepancy scores. Reviewing the course PowerPoints was found to be positively correlated with the raw discrepancy scores on Exam 1 ($r = .183$, $p < .01$), Exam 3 ($r = .130$, $p < .05$), and the Final Exam ($r = .127$, $p < .05$) (see Table 7). Use of PowerPoint lectures as study tools had a positive relationship with exam raw discrepancy scores, meaning that students who used PowerPoint to study overestimated their exam scores. Use of the practice exam was found to be negatively correlated with raw discrepancy scores on Exam 1 ($r = -.168$, $p < .01$), meaning that students who used the practice exam to study underestimated how well they would do on their exam.

Table 2.

Descriptive Statistics for All Exam Discrepancy Variables

Variable Name	***N***	***M***	***SD***	**Minimum**	**Maximum**
Exam 1 actual scores	384	75.97	15.45	32.00	100.00
Exam 2 actual scores	383	74.90	16.73	32.00	105.00
Exam 3 actual scores	381	70.44	18.46	23.00	103.00
Final exam actual scores	373	82.87	11.00	30.00	100.00
Exam 1 expected scores	328	82.89	7.87	55.00	100.00
Exam 2 expected scores	346	80.78	8.91	45.00	100.00
Exam 3 expected scores	317	80.24	9.17	40.00	100.00
Final Exam expected scores	315	82.04	9.80	50.00	109.00
Raw discrepancy Exam 1	312	6.44	14.49	-56.00	68.00
Raw discrepancy Exam 2	334	5.58	13.70	-19.00	59.00
Raw discrepancy Exam 3	317	9.53	15.67	-48.00	57.00
Raw discrepancy Final Exam	314	-1.39	11.32	-30.00	78.33
Absolute value discrepancy Exam 1	312	11.45	10.96	0.00	68.00
Absolute value discrepancy Exam 2	334	11.36	9.46	0.00	59.00
Absolute value discrepancy Exam 3	317	13.89	11.96	0.00	57.00
Absolute value discrepancy Exam 4	314	8.25	7.87	0.33	78.33

Table 3.

Fit Statistics for Latent Class Growth Analysis Models

N of Classes	Type	BIC (Adjusted)	BIC	AIC	Log-likelihood	$n_{\text{Class 1}}$	$n_{\text{Class 2}}$	$n_{\text{Class 3}}$	Entropy
2	Quadratic	10197.88	10232.79	10189.36	-5083.681	75	308		0.778
2	Linear	10192.34	10220.89	10185.36	-5083.681	75	308		0.778
3	Quadratic	10096.38	10143.97	10084.76	-5027.378	22	213	148	0.761
3	Linear	10088.06	10126.13	10078.76	-5027.378	22	213	148	0.761

Table 4.

Analysis of Variance for Exam 1 Differences by High School Biology Course

Variable	*SS*	*DF*	*MS*	*F*	*p*
Between Groups	2792.85	3	930.95	3.95	.009
Within Groups	800063.29	340	235.48		

Table 5.

Correlations Between Absolute Value Exam Discrepancy Scores and the Proportion of Days Studied

Variables	1	2	3	4	5	6	7
1. Absolute value discrepancy Exam 1							
2. Absolute value discrepancy Exam 2	.557**						
3. Absolute value discrepancy Exam 3	.616**	.483**					
4. Absolute value discrepancy Final Exam	.362**	.389**	.272**				
5. Proportion of days studied for Exam 1	.056	.025	.016	.053			
6. Proportion of days studied for Exam 2	.144*	.112*	.122*	.004	.581**		
7. Proportion of days studied for Exam 3	.064	.099	.106	.028	.568**	.704**	
8. Proportion of days studied Final Exam	.196**	.093	.182**	-.020	.424**	.594**	.616**

*$p < .05$, **$p < .01$

Table 6.

Study Tools Used For Each Exam

Study Tools	Exam 1	Exam 2	Exam 3	Final Exam
PowerPoint	47.16%	34.06%	30.17%	29.02%
Textbook	62.99%	44.69%	38.79%	30.75%
Deep learning	15.52%	14.71%	16.67%	10.63%
Learning objectives	40.00%	38.21%	27.87%	24.71%
Clicker questions	68.06%	70.84%	80.75%	68.10%
Echo 360 recordings	36.12%	49.05%	37.36%	28.45%
Studying in a group	26.57%	30.25%	27.87%	33.91%
In class notes	88.66%	79.29%	73.56%	68.68%
Practice exam	91.04%	92.92%	94.25%	95.98%
Unit recommendations	33.43%	28.88%	24.43%	20.69%
Online resources	39.40%	43.60%	45.69%	39.37%
Other	5.37%	7.90%	7.47%	6.03%

Note. This table represents the percentage of students who took each survey and reported using a specific tool. The frequency of each study tool being used at each exam was then divided by the sample size at each wave in order to get the percent of students who used each tool at each survey.

Table 7.

Correlations of Raw Exam Discrepancies and Study Tools Used

Study Tools	Exam 1	Exam 2	Exam 3	Final Exam
PowerPoint	.183**	.098	.130*	.127*
Textbook	.069	.154**	.074	.168**
Deep learning	-.027	.016	.064	.098
Learning objectives	.071	.115*	.140*	.140*
Clicker questions	-.105	-.047	.112*	.113*
Echo 360 recordings	-.037	-.034	.113*	-.001
Studying in a group	.014	.091	.006	.122*
In class notes	-.032	.044	.035	.144*
Practice exam	-.168**	-.030	-.078	-.012
Unit recommendations	-.052	-.022	.074	.046
Online resources	.035	-.018	.074	.119*
Other	.001	.089	-.014	.089

*$p < .05$, ** $p < .01$

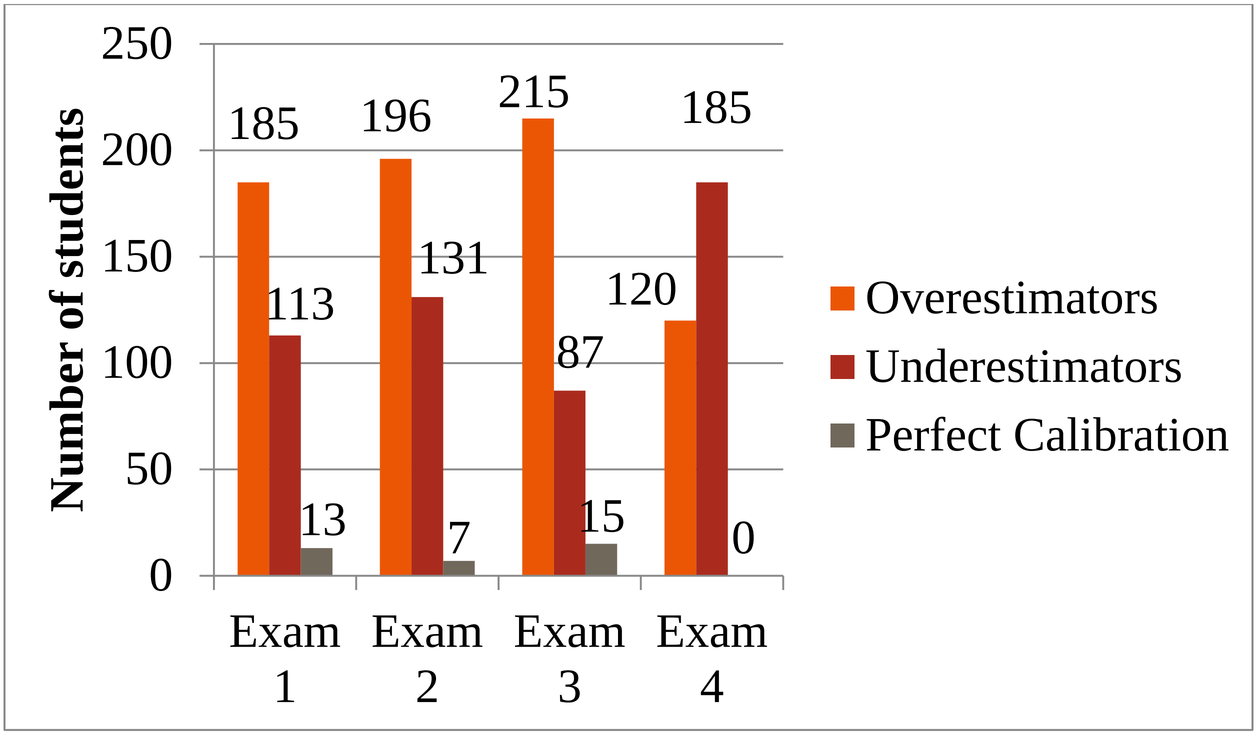

Figure 2. Frequency of Overestimators and Underestimators

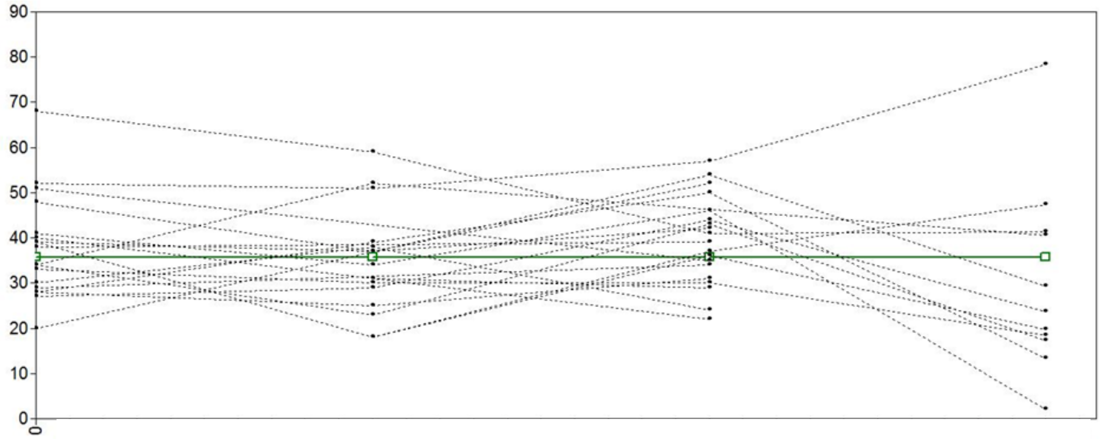

Figure 3. Latent Class Growth Analysis Class One

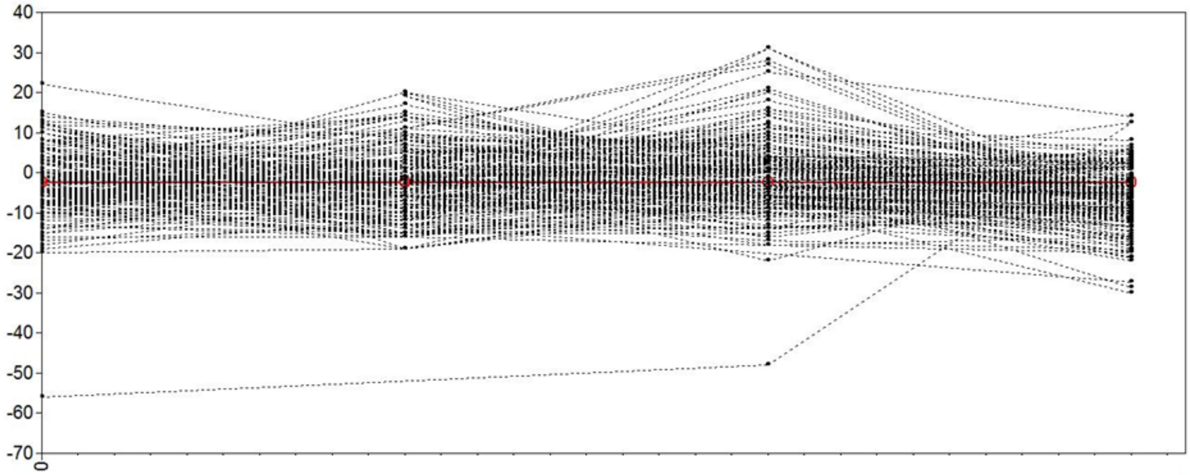

Figure 4. Latent Class Growth Analysis Class Two

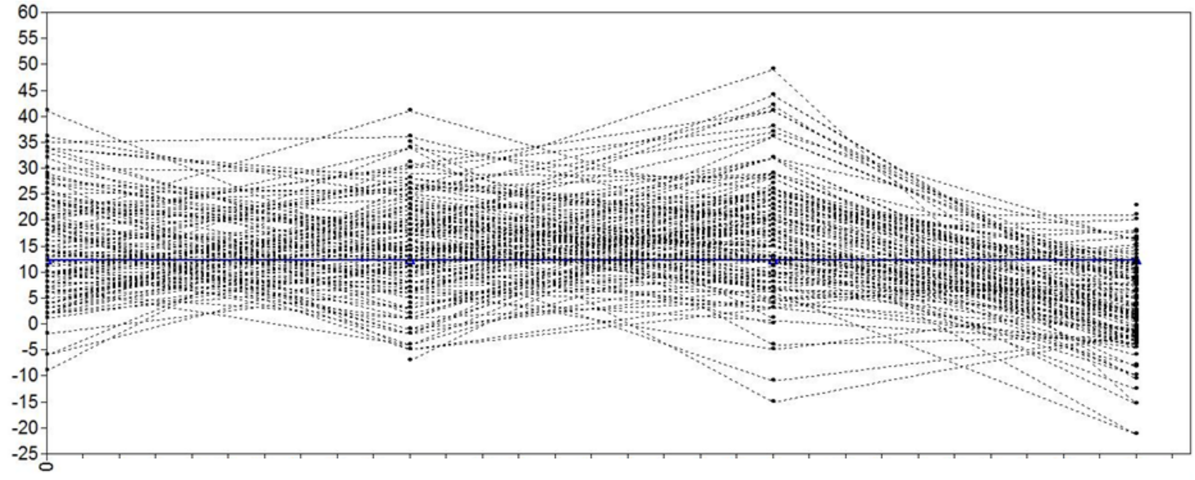

Figure 5. Latent Class Growth Analysis Class Three

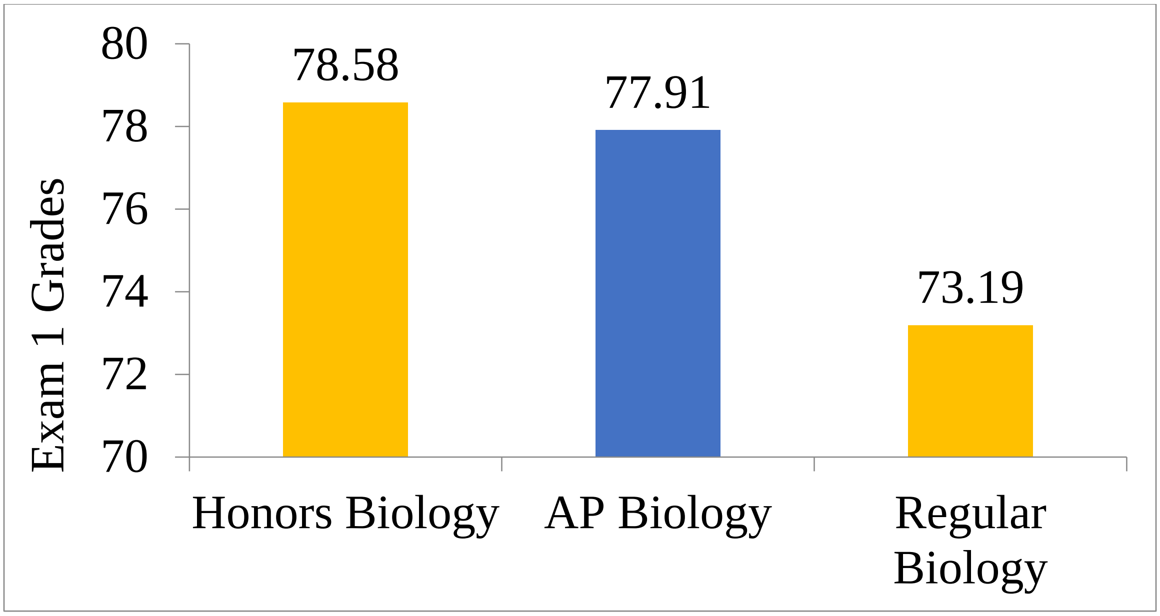

Figure 6. Mean Exam 1 Grades by Previous Biology Class

Chapter 5: Discussion

The purpose of this thesis was to examine how students' study habits and self-regulation interacted with and influenced students' metacognitive awareness and calibration in a large introductory biology course. Students must be aware of their own knowledge if they want to implement different self-regulatory strategies needed to succeed in large undergraduate courses (Zimmerman, 2002). Being an effective self-observer allows students to reflect on their performance, make judgments about what to do in future performances, and be aware of how they may perform (Brown et al., 2014). To examine students' metacognitive awareness, I examined students' calibration on their exams.

When examining how students' expected scores compared to their actual scores, I created an absolute value discrepancy and a raw discrepancy score. The absolute value discrepancy score provides a description of how "off" or inaccurate students were in their exam estimations. If a student estimates that he is going to receive a 90 on his exam, but only received an 84, he would have an absolute discrepancy of 6 points. One of the major findings of this study was that students initially miscalibrate their exam grades, and by the end of the semester, they get better at estimating how they will do. This finding mirrors the research conducted by Valdez (2013) who also found that undergraduate students' absolute discrepancies improved over the course of a semester in a communication disorders class.

Rather than simply examining how inaccurate students may be, I was also able to consider the direction of students' inaccuracy. Considering direction allowed me to see if students on average were underestimating or overestimating. When examining the raw discrepancy score, it appears that students start the semester by overestimating their performance. By the end of the semester, most are underestimating their performance. This may be because students begin to understand the style of the test, what is expected of them, and the nature of the material. This result builds on the findings of Jensen and Moore (2008) that students' discrepancies decreased over the course of a semester.

Both discrepancy scores point to a trend showing that students' estimations seem to be more discrepant in the beginning of the semester. To help raise students' awareness of their inaccurate judgments, instructors could pay more attention early in the course to how students are performing. Instructors should make sure that students are not misunderstanding the material. Students who are encouraged to self-observe and evaluate their understanding of the information are more likely to be aware and be able to effectively self-regulate (Lin & Zabrucky, 1998).

Rather than only examining calibration at each exam individually, I aimed to consider how students' exam discrepancies changed from one exam to another. Latent class growth analysis revealed that students' estimation patterns do not all follow similar trajectories. Students appear to fall into three different groups. One group of students consisted of initial overestimators who became more calibrated. These students were discrepant at the beginning of the course and improved across the semester. The second group of students started the course overestimating their performance and ended the course still overestimating. The last group of students started the semester as "good" calibrators (i.e., were not highly discrepant) and remained well-calibrated across the semester. Using latent class growth analysis helps fills an important gap in the literature. Few other studies have examined calibration trends over time. Not all students improved their calibration across the semester as the full sample means suggest. Using a person-centered method like latent class growth analysis allowed me to examine students on an individual level.

It may be hard for instructors to immediately tell what types of students are in their course. Many students begin the semester by overestimating, and it can be hard to predict who will improve in their calibration and who will not. One way instructors can help detect persistently poor calibration is by checking the students' beginning and midterm calibration. By checking in on how well students are calibrating in the beginning of the semester and in the middle of the semester, instructors will be able to differentiate between those students' who are improving in their judgments and those who are not. Steps can then be taken to help those students who are not improving by the middle of the semester. Zimmerman's (2002) model of self-regulated learning explains that students must be aware before they can self-regulate their strategies. Being aware plays a part in each of the phases in Zimmerman's model. For students to plan out their goals for the next exam or to select appropriate tasks, they must be aware of their current performance.

There is a limitation of using the latent class growth analysis approach with this data. Small sample sizes may have led to an inaccurate number of latent groups. Given the relatively small sample size, it may be valuable to consider other methods of verification, such as conducting interviews with students. Interviewing students could allow for a closer examination of students' calibration patterns and self-regulatory processes. Interviews can be conducted with smaller groups and can provide results about how individuals' calibration changes over a semester. Another limitation was that, in this study, covariates (e.g., gender, year level, study habits, achievement) were not examined by latent group. Examining covariates would provide information about which students comprise each group and about how each group performs. It may also be important to consider how covariates form the groups. It could be that the group to which students belong depends on student characteristics such as gender, year level, specific study habits, or other factors.

Another aim of this study was to examine the factors that may affect students' exam scores and discrepancies. One factor that was of interest was which biology course students had in high school. Dinsmore and Parkinson (2013) suggested that prior knowledge affects students' study habits and exam estimations. I investigated whether Exam 1 scores varied based on which high school biology course students took. I expected that students' Exam 1 grades would be different depending on students' high school biology course. I found a significant difference between scores for students who took honors biology and those who took regular biology in high school. Those who took honors biology performed better on Exam 1 compared to those who were enrolled in a regular biology class. This finding supported my original hypothesis.

Factors such as self-regulatory habits may also play a role in influencing exam grades and exam estimations. By examining how long students studied prior to each exam, I was able to see the relationship between days studied and how well calibrated students were. I did not have a specific hypothesis about the relationship between number of days studied and students' calibration. Overall, the results showed that students' estimations on Exam 1 and the Exam 3 were the only calibration scores in the course that were related to the number of days studied. Results suggest that the relationship between student calibration and the amount of time they studied for exams are reciprocally influential. For example, students who were poorly calibrated on Exam 1 reported studying more prior to Exam 2 and the Final Exam. It could be that the shock of performing lower than one expects leads a student to initiate more intense studying behaviors. Establishing the patterns of these relationships would need further empirical scrutiny and more precise measures of studying time.

One possible reason that the proportion of days studied and overestimations on Exam 3 were related to one another is that, if students who studied a great deal for Exam 2 did well, they may have increased confidence that they would perform well on Exam 3. But it turns out that Exam 3 was the hardest of all the exams. This could also be that students overestimated how they were going to perform on Exam 3 and then modified their study habits for the Final Exam. This could mean that as students progress through the course, they begin to take into account how long they have studied for their exam and relate this to their expected achievement.

The relationships between prior course and performance and between number of days studied and discrepancy scores are supported by Zimmerman’s (2002) model of self-regulation. Students who consider themselves to be more familiar with the content (presumably those who took an honors course) tend to self-regulate more effectively compared to those who are less familiar or less proficient (presumably those who took a regular-level biology course). As the course progresses, students who are less familiar may start to consider themselves more proficient. According to Zimmerman, novices do not engage in the same quality forethought processes (i.e., planning, goal setting, and outcome expectations) as those who are experts. Those who are novices in the subject may instead try to self-regulate after their performance rather than before. This may be why students choose to study more after they perform in unexpected ways on their previous exams. Those who are experts set goals and are more likely to self-regulate before the performance.

Because students are more likely to be novices at the beginning of the class, encouraging students to study more for the first exam can help them engage in the forethought process. It may be important for instructors to emphasize the importance of studying for the first exam. For the first exam it may be more important that students focus on setting specific goals, planning study time, and having more accurate outcome expectations. Engaging in these forethought processes can help students in the performance phase and will allow students to not overestimate their performance.

Not only is how long someone studied an interesting factor to consider when examining calibration, but examining students' particular study tools is of as well. The type of study tool that students choose to use can be important for their calibration. If students use tools that do not help them engage in deep learning, they may feel that they have a sense of understanding with the material even if they do not (Brown et al., 2014). In our this study, the more a student studied using PowerPoint lectures, the more discrepant the student was likely to be when judgment exam performance. Similar to previous findings, this study showed that simply reviewing notes is not a useful strategy (Lopez et al., 2013). In contrast, using the practice exam as a study strategy was found to be inversely correlated with discrepancy scores. This means that for Exam 1, those who used the practice exam were more accurate in their exam estimations.

These relationships may be, in part, because using the PowerPoint and practice exams may evoke different types of learning. Simply reviewing the PowerPoints, reviewing notes, and reading the textbook as forms of studying, does not deeply engage the mind (Brown et al., 2014). Rather, engaging with the material by using practice exams forms a deeper knowledge and understanding of the material (Brown et al., 2014). Study tools that help engage students in the material more deeply can help students become more aware of what they do and do not know. Emphasizing study materials such as self-quizzes, and deep learning strategies that force students to make connections to the material may be one way to help them students from overestimating their scores.

Limitations and Future Directions

This study has several limitations. Students were asked to report their exam estimations prior to their exam. This allowed for a pre-estimation of students' grades. Asking students to estimate their scores prior to their exam allowed me to examine students' expectations based upon how they may have prepared for the exam. However, this method did not give students the benefit of reflecting on their performance to estimate their grade. Using both pre and post estimation scores could have provided a more rounded view of how students' expectations may change according to their performance.

Also worth nothing is that the participants in this study were enrolled in two different sections of the same class. The course was structured the same for both sections and the students received the same exam, but teachers for each section were different. Students may have had different experiences in the course based upon their experience with a specific instructor.

Another limitation of this study is the way that number of days studied was measured. This data source was measured by asking students to indicate how many days they studied on a calendar. This measure did not differentiate between those who skipped the question and those who did not study any days. Anyone who did not have a response to the question was assumed to have not studied at all. Another issue with this measure is that it does not take into account how long a student studied per day. For example, two students who both studied on six days prior to their exam could have spent different amounts of time studying. One student may have studied for 30 minutes each day, while the other student studied for two hours each day. The measure used in this study did not allow for these differences to be examined.

Lastly, a limitation of this study was the way in which calibration was measured in conjunction with study approaches. These relationships were examined using the mean proportion of study days for the full sample. This may limit the information that it can provide. In addition, the choice to use absolute value discrepancy scores did not permit me to investigate the study habits reported by students who overestimated or underestimated their biology performance.

Instead of examining calibration with the absolute value discrepancy score and the different study habits, using latent class growth analysis would be a better method to examine the relationship between discrepancy scores and the various study skills. Using latent class growth analysis could allow for students to be separated into groups of over estimators and under estimators. These groups could be similar to the groups in our latent class growth analysis. If latent class growth analysis was not used, groups could be created by using the average discrepancy for each exam and creating cut off points using the associated standard deviation scores. A larger sample size would be necessary for this analysis.

Another way that groups could be created is by considering the specific study tools used and by number of days studied as covariates. This would allow me to see how different latent classes (e.g., chronic overestimators) spend their study time for each exam. Study habits may look different across groups or may have more of a significant effect on one group and not another. By examining these relationships, researchers could obtain a much clearer picture of what may be happening with students in similar large science classes.

Researchers may want to examine students' study habits in different ways than how they were examined in this thesis. Examining more than just the proportion of days that students studied is important. Considering the distribution of study time is important. In other words, does it matter how many days a student studies or does it matter when prior to an exam that a student studies? Students who study on a weekly basis prior to their exam may perform differently than students who study only at the last minute. Researchers could investigate whether and how distributed studying affects students' calibration compared to studying at the last minute. How often students study and when prior to the exam they study may be related to how aware they are of their knowledge. Examining the number of hours that were studied outside of the classroom may provide a more reliable measure of students' study time.

Aside from examining how students' calibration interacts with students' study habits, researchers may want to examine how students' study habits interact with students' achievement in their course. Do students who use particular strategies perform better in the course? By examining achievement, researchers would be able to determine if specific study tools or if the amount of time students report studying is related to higher achievement on each test and in the overall course. Examination of achievement and study habits may be a practical import for professors.

The implications and results of this thesis can be applied to students in large science courses. Students who enroll in these courses may need different kinds of support to be successful compared to what is needed in other introductory courses. Instructors can be aware of students' course history in order to foster success. They can also check student understanding throughout the course to determine those who may need more specific help. Instructors can also encourage students to use study tools that require deeper learning and suggest effective ways in which students ought to study.

Findings from this paper can inform the broader literature on self-regulation and calibration. Few other studies have examined calibration as a part of the self-regulatory process in large introductory science classes. By examining college students in large classes, I was able to see how students' calibration and self-regulatory habits interacted across a semester. Using latent class growth analysis I was able to examine how students may change in different groups across the semester rather than as a whole.

Appendix A - IRB Consent Form

Consent to Participate in a Research Study on Study Habits in Biology

Greetings Biology Student,

Mrs. Trisha Turner (Master's student) and Dr. Ellen Usher (associate professor) of the Department of Educational, School, and Counseling Psychology, are inviting you to take part in a research study about your study habits and beliefs related to learning biology. You are being invited to take part in this research study because you are enrolled in a biology class. If you volunteer to take part in this study, you will be one of about 980 people this semester to do so. If you are under the age of 18, you should not participate in the study.

Your part in the study will be to allow us to use information from online surveys and brief, in-class questionnaires that will be related to your biology class. These surveys ask you about your study habits and beliefs over the course of the semester. Your participation also involves allowing the researchers access to your grades in the course. The approximate total time commitment will be one hour.

Your responses will be used to help researchers understand how students learn biology. There are no known risks or discomforts as a result of your participation in this research study nor will you benefit directly from taking part in this study. By consenting to participate, you will allow us to access your responses. Some participants may be contacted for an interview if they take part in this research study.

We do everything we can to protect your privacy and confidentiality. We will not tell anybody outside of the research team that you were in this study or what information we collected about you in particular. All personally identifying information will be removed from grades and questionnaire data and a numeric ID known only to the researchers will be used to identify you. Your responses will be combined with those of other students in the course when findings are reported. However, we may be required to show information which identifies you to people who need to be sure we have done the research correctly; these would be people from such organizations as the University of Kentucky. We will make every effort to keep private all research records that identify you to the extent allowed by law. Please be aware, while we make every effort to safeguard your data once received from the online survey/data gathering company, given the nature of online surveys, as with anything involving the Internet, we can never guarantee the confidentiality of the data while still on the survey/data gathering company's servers, or while en route to either them or us. It is also possible the raw data collected for research purposes may be used for marketing or reporting purposes by the survey/data gathering company after the research is concluded, depending on the company's Terms of Service and Privacy policies.
You do not have to be in this study. You may choose not to share your data with us at any time. You will not be punished in any way if you decide not to be in the study or to stop taking part in the study. **If you decide not to take part or to stop taking part in this study, it will not affect your grade in any way.**

If you have any questions or concerns about this study or if any problems arise, please contact Trisha Turner at the University of Kentucky at trisha.douin@uky.edu or Dr. Ellen Usher at ellen.usher@uky.edu.

If you have any questions about your rights as a volunteer in this research, contact the staff in the Office of Research Integrity at the University of Kentucky between the business hours of 8am and 5pm EST, Mon-Fri. at 859-257-9428 or toll free at 1-866-400-9428. We will give you a signed copy of this consent form to take with you.

****

I have read this form and have been allowed to ask any questions I might have.

Participant's Printed Name: ____ UK Student ID: ____________________

Participant's signature: ____ Date: __________________

Course number and instructor's name: ____

Name of person obtaining consent: _____________________________________ Date: ______________

Appendix B - Exam One Survey

BIO 148 - February 9, 2015 **Please write your 8-digit UK ID # (without the "9")** ______________________

Please answer honestly. You will receive in-class activity points for completing this. Individual responses will not be seen by your instructor.

1. Please write the numeric score (one number from 0-100) that you expect to get on this exam: ________
2. Which of the following resources did you use while studying? (check all that apply)

______ PowerPoint presentations	______ Studying in a group
______ Textbook	______ In-class notes
______ Deep learning/effective study strategies handout	______ Practice exam
______ Reviewing the unit Learning Objectives	______ Unit study recommendations
______ Clicker questions	______ Online resources
______ Echo 360 recordings	______ Other: ____________________

3. On the calendar to the right, put a large "X" on any day prior to Exam 1 that you studied biology <u>outside of class time</u> (reviewed, took the practice exam, reviewed unit recommendations, etc.). →

January-February 2015						
M	T	W	T	F	S	S
First day		14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	Feb 1
2	3	4	5	6	7	8
9	← Today (Exam 1)					

4. I would change the way I prepared for this exam (circle one).

Definitely False	Mostly False	A little bit False	A little bit True	Mostly True	Definitely True

5. I have a good idea of how I am going to do on this exam (circle one).

Definitely False	Mostly False	A little bit False	A little bit True	Mostly True	Definitely True

6. I studied the most important information that will be presented on this exam (circle one).

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree

Appendix C- Exam Two Survey

BIO 148 - March 9, 2015 **Please write your 8-digit UK ID # (without the "9")** ________________________

Please answer honestly. You will receive in-class activity points for completing this. Individual responses will not be seen by your instructor.

1. Please write the numeric score (one number from 0-100) that you expect to get on this exam: _________
2. Which of the following resources did you use while studying? (check all that apply)

_____ PowerPoint presentations
_____ Textbook
_____ Deep learning/effective study strategies handout
_____ Reviewing the unit Learning Objectives
_____ Clicker questions
_____ Echo 360 recordings
_____ Studying in a group
_____ In-class notes
_____ Practice exam
_____ Unit study recommendations
_____ Online resources
_____ Other: ____________________

3. On the calendar to the right, put a large "X" on any day prior to Exam 2 that you studied biology <u>outside of class time</u> (reviewed, took the practice exam, reviewed unit recommendations, etc.).

February-March 2015						
M	T	W	T	F	S	S
Exam 1	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	Mar 1
2	3	4	5	6	7	8
9	← Today (Exam 2)					

4. I procrastinate when studying for Biology (circle one).

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree

5. I would change the way I prepared for this exam (circle one).

Definitely False	Mostly False	A little bit False	A little bit True	Mostly True	Definitely True

6. I have trouble starting my Biology work (circle one).

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree

7. I have a good idea of how I am going to do on this exam (circle one).

Definitely False	Mostly False	A little bit False	A little bit True	Mostly True	Definitely True

8. I wait until the last minute to do my Biology work (circle one).

Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree

9. Which of the following best applies to you (for this class):
   a. I have not sought any help this semester.

b. I have sought help from a peer.
c. I have sought help from the instructor.
d. I have sought help from a university resource (i.e., The Study).
e. Some combination of B, C, or D.

10. Which of the following best applies to you:
    a. I have offered to help a peer study for this class.
    b. I have not offered to help a peer study for this class.

Appendix D - Exam Three Survey

BIO 148 - April 13, 2015 **Please write your 8-digit UK ID # (without the "9")** ______________________

Please answer honestly. You will receive in-class activity points for completing this. Individual responses will not be seen by your instructor.

1. Please write the numeric score (one number from 0-100) that you expect to get on this exam: _________

2. Which of the following resources did you use while studying? (check all that apply)

_____ PowerPoint presentations
_____ Textbook
_____ Deep learning/effective study strategies handout
_____ Reviewing the unit Learning Objectives
_____ Clicker questions
_____ Echo 360 recordings
_____ Studying in a group
_____ In-class notes
_____ Practice exam
_____ Unit study recommendations
_____ Online resources
_____ Other: ____________________

3. On the calendar to the right, put a large "X" on any day prior to Exam 3 that you studied biology <u>outside of class time</u> (reviewed, took the practice exam, reviewed unit recommendations, etc.).

March-April 2015						
M	T	W	T	F	S	S
Exam 2	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	← Today (Exam 3)					

4. Have you taken UK 101 or UK 201 (circle one)?

Yes – UK 101 Yes – UK 201 No Unsure

5. What kind of high school biology class did you take (circle one)?

Advanced Placement Honors Biology Regular Biology Other: _________

6. If you took Advanced Placement biology, what was your score on the AP exam (circle one)?

1 (lowest) 2 3 4 5 (highest) Not sure Did not take AP exam

7. Which of the following best applies to you (for this biology class):
   f. I have not sought any help this semester.
   g. I have sought help from a peer.
   h. I have sought help from the instructor.
   i. I have sought help from a university resource (i.e., The Study).
   j. Some combination of B, C, and/or D.

8. Which of the following best applies to you:
   c. I have offered to help a peer study for this class.
   d. I have not offered to help a peer study for this class.

9. What was the most helpful strategy you used to study for this exam?

____

Appendix E - Final Exam Survey

BIO 148 - May 6, 2015 **Please write your 8-digit UK ID # (without the "9")**

____

Please answer honestly. You will receive in-class activity points for completing this. Individual responses will not be seen by your instructor.

1. Please write the numeric score (one number from 0-100) that you expect to get on this exam: __________

2. Which of the following resources did you use while studying? (check all that apply)

______ PowerPoint presentations
______ Textbook
______ Deep learning/effective study strategies handout
______ Reviewing the unit Learning Objectives
______ Clicker questions
______ Echo 360 recordings
______ Studying in a group
______ In-class notes
______ Practice exam
______ Unit study recommendations
______ Online resources
______ Other: ________________________

3. On the calendar to the right, put a large "X" on any day between Exam 3 and the final exam that you studied biology <u>outside of class time</u> (reviewed, took the practice exam, reviewed unit recommendations, etc.).

April-May 2015						
M	T	W	T	F	S	S
Exam 3	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	May 1	2	3
4	5	6	← Today (Final Exam)			

4. Which of the following best applies to you as you prepared for this biology final:
   k. I have not sought any help this semester.
   l. I sought help from a peer.
   m. I sought help from the instructor.
   n. I sought help from a university resource (i.e., The Study).
   o. Some combination of B, C, and/or D.

5. Which of the following best applies to you:
   e. I have offered to help a peer study for this class.
   f. I have not offered to help a peer study for this class.

6. Which of the following best describes you?
   a. Neither of my parents have completed a four-year college degree.
   b. One of my parents has completed a four-year college degree.
   c. Both of my parents have completed a four-year college degree.
   d. Unsure

7. Rank each of your biology exams in order according to <u>how much time you spent studying for it</u>.
   (1 = *spent <u>most</u> time studying* ………… to ………… 4 = *spent the <u>least</u> time studying*)

_____ Exam 1

_____ Exam 2

_____ Exam 3

_____ Final Exam

8. Please write the numeric score (one number from 0-100) that you expect to get overall in this course:

References

Bembenutty, H. (2009). Three essential components of college teaching: Achievement calibration, self-efficacy, and self-regulation. *College Student Journal, 43*(2), 562-570.

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Cambridge, MA: The Belknap Press of Harvard University Press.

Boekaerts, M., & Rozendall, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction, 20,* 372-382.

Danner, F. W., & Toland, M. D. (2012). The interactive role of socioeconomic status, race/ethnicity, and birth weight on trajectories of body mass index growth in children and adolescents. *The Journal of Early Adolescence, 33*(3), 293-314.

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgements made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24,* 4-14.

Garavalia, L. S., & Gredler, M. E. (2002). An exploratory study of academic goal setting, achievement calibration and self-regulated learning. *Journal of Instructional Psychology, 29*(4), 221-230.

Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction, 24,* 37-47.

Jensen, P. A., & Moore, R. (2008). Students' behaviors, grades & perceptions in an introductory biology course. *The American Biology Teacher, 70*(8), 483-487.

Lin, L. M., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*, 345-391.

Lopez, E. J., Nandagopal, K., Shavelson, R. J., Szu, E., & Penn, J. (2013). Self-regulated learning study strategies and academic performance in undergraduate organic chemistry: An investigation examining ethnically diverse students. Journal of Research in Science Teaching, 50(6), 660-676.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory and Cognition, 39*, 462-476.

Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anales de psicologia, 30*(2), 450-462.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, 33, 565–576.

Rice, K. G., Ray, M. E., Davis, D. E., DeBlaere, & Ashby, J. S. (2015). Perfectionism and longitudinal patterns of stress for STEM majors: Implications for academic performance. Journal of Counseling Psychology, 63(3), 247-365. DOI: 10.1037/cou0000097

Schraw, Crippen, & Hartley, (2006). Promoting self-regulation in science education: Metacognition as a part of a broader perspective on learning. *Research in Science Education, 36*, 111-139.

Schunk, D. H., & Usher, E. L. (2011). Assessing self-efficacy for self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 282-297). New York, NY: Routledge.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12*(4), 437-475.

Valdez, A. (2013). Student Metacognitive Monitoring: Predicting Test Achievement from Judgment Accuracy. *International Journal of Higher Education, 2*(2), 141- 146.

VanderStoep, S. W., Pintrich, P. R., & Fagerlin, A. (1996). Disciplinary differences in self-regulated learning in college students. *Contemporary Educational Psychology, 21*, 345-362.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology, 27*, 551-572.

Zimmerman, B. J., & Paulson, A. S. (1995). Self-monitoring during collegiate studying: An invaluable tool for academic self-regulation. In Paul R. Pintrich (Eds). *Understanding self-regulated learning* (no. 63, pp. 13-42). San Francisco, CA: Joey-Bass Publishers.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P.R. Pintrich, & M. Zeider (Eds.), *Handbook of Self-Regulation* (pp. 13-39). San Diego: Academic Press.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64-70.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*, 166-183.

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In B. J. Zimmerman & D. H. Schunk (Eds.) *Handbook of self-regulation of learning and performance* (pp.1-29). New York, NY: Routledge.

Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education, 25*(9), 1081-109

Vita

Trisha A. Turner
Maiden name: Douin

EDUCATION

Expected: 2016	**Masters of Science, Educational Psychology** University of Kentucky (Lexington, KY) Thesis: *Am I able to predict how I will do? Examining calibration in an undergraduate biology course.* Advisor: Dr. Ellen L. Usher GPA: 3.57
2014	**Bachelor of Arts, Psychology** University of Kentucky (Lexington, KY) GPA: 3.52
2012	**Associate of Arts** Somerset Community College (Somerset, KY)
2010	**High School Diploma** Southwestern High School (Somerset, KY)

RESEARCH INTERESTS

My research interests include self-regulation, metacognitive awareness, self-efficacy, and student motivation. The focus of my research has been with students' self-regulation and metacognitive awareness regarding their academic achievement and performance. I am also interested in students' self-efficacy and their motivational beliefs.

PEER-REVIEWED RESEARCH PRESENTATIONS

Pritchard, M. J., **Turner, T. A.,** Usher, E. L., & Jones, F. L. (2016, August). *Being aware of your thinking: Examining mindfulness and metacognition in undergraduate biology students*. Poster to be presented at the annual meeting of the American Psychological Association, Denver, CO.

Turner, T. A., Jones, F. L., & Usher, E. L. (2016, March). *Do I know what I know? Metacognition and self-regulation in undergraduate biology*. Paper to be presented at the 2016 Spring Research Conference. Lexington, Kentucky.

Jones, F. L., **Turner, T. A.,** Page, L. B., & Usher, E. L. (2016, March). *Sleepy? Self-regulated? Successful? An analysis of the relationships among sleep deprivation,* *self-regulation and student performance in undergraduate biology students.* Poster to be presented at the 2016 Spring Research Conference. Lexington, Kentucky.

Osterhage, J., **Turner, T. A.,** & Usher, E. L. (2015, April). *Active study guidance: Learning through practice.* Poster presented at the meeting of the Biology Leadership Community. Austin, Texas.

Turner, T. A., Adams, C. G., Rose, M. A., Butz, A. R., & Usher, E. L. (2015, April*).* *Sources of college-going self-efficacy among rural Appalachian students*. Poster presented at the annual meeting of the American Educational Research Association. Chicago, Illinois.

Turner, T. A., Cheatham, N. N., Jones, F. L., Waiters, B. L., Pritchard, M. J., Covington, S. T., Kiran, D., Wilson, A., & Usher, E. L. (2015, March). *Self-regulation and calibration in undergraduate biology students*. Paper presented at the 2015 Spring Research Conference. Louisville, Kentucky.

Douin, T. A., Meiners, N. C., Rose, M. A., Eddy, Z. S., Adams, C. G., Goodnight, A.C., Butz, A. R., & Usher, E. L. (2014, March). *Sources of college-going self-efficacy in rural Appalachia.* Paper presented at the Spring Research Conference. Cincinnati, Ohio.

Rose, M. A., Adams, C. G., **Douin, T. A**., Meiners, N. C., Eddy, Z. S., Butz, A. R., & Usher, E. L. (2014, April). *Sources of college-going self-efficacy in rural Appalachia*. Paper presented at the National Conference of Undergraduate Research. Lexington, Kentucky.

RESEARCH EXPERIENCE

Fall 2014- Spring 2016	**Graduate Research Assistant,** P20 Motivation and Learning Lab, University of Kentucky. Supervisor: Dr. Ellen Usher, Director
Fall 2014 – Spring 2016	**Graduate Assistant,** Robinson Scholars Program, University of Kentucky. Supervisor: Jeff Spradling, Director
Fall 2015 – Spring 2016	**Research Assistant**, Applied Psychometrics Strategies Lab University of Kentucky. Supervisor: Dr. Michael Toland, Director
Fall 2013-	**Research Lab Assistant**, P20 Motivation and Learning Lab,

Spring 2014	University of Kentucky. Supervisor: Dr. Ellen Usher, Director

AWARDS

Fall 2013	Dean's List Recipient (University of Kentucky)
Fall 2012	Dean's List Recipient (University of Kentucky)
Spring 2012	Dean's List Recipient (Somerset Community College)

PREVIOUS WORK EXPERIENCE

May 2011- August 2014	**Sales Associate**, Shoe Show Inc. Supervisor: Linda Paluzi
June 2010- December 2012	**Photo Editor**, Photography By Angela Vaught Supervisor: Angela Vaught
Jan 2010- May 2010	**Teaching Assistant**, Southwestern High School Supervisor: Monica DeCarlo

PROFESSIONAL DEVELOPMENT

Affiliations

Phi Theta Kappa (Fall 2010 – Spring 2012)

Division 15 of APA (June 2014 – December 2015)

AERA Division C (June 2014 – December 2015)

Motivation SIG (June 2014 – December 2015)

AERA SSRL SIG (June 2014 – December 2015)