Fall 2020

Exploring Student Motivation in Quantitative Biology

Alexander Richard Kulacki
University of New Hampshire, Durham

EXPLORING STUDENT MOTIVATION TOWARDS QUANTITATIVE BIOLOGY

BY

ALEXANDER R. KULACKI

Bachelor of Arts, Carleton College, 2016

THESIS

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Master of Science

in

Biological Sciences: Integrative and Organismal Biology

September, 2020

This thesis was examined and approved in partial fulfillment of the requirements for the degree of M.S. in Integrative and Organismal Biology by:

Thesis Director, Dr. Melissa L. Aikens, Assistant Professor of Biology Education (Integrative and Organismal Biology)

Dr. Carrie L. Hall, Assistant Professor of Ecology and Biology Education (Integrative and Organismal Biology)

Dr. Daniel R. Howard, Assistant Professor of Ecology (Integrative and Organismal Biology)

On 07/30/2020

Original approval signatures are on file with the University of New Hampshire Graduate School.

DEDICATION

This thesis is dedicated to my family and friends. To my parents, who have seen me at my best and worst, and who have moved heaven and earth to support me on my journey through life: thank you for being my unwavering beacons in a cold dark sea. To my friends from childhood, college, and currently at UNH: thank you for gifting me the most timeless moments of my life, the most precious gift of our conversations, and the warm hearth of your companionship. Despite being scattered around this great, big, crazy world of ours, wherever you are and whatever you are doing, I'm glad you're here, and there's no place I'd rather be than on this rock with you.

ACKNOWLEDGEMENTS

I would like to thank Dr. Melissa Aikens for sending out that fateful posting on the Ecolog Listserv nearly three years ago that set me off on this journey across the country, for her patience and understanding as my guide through the 'Wild West of DBER', and for showing me how to really get my hands dirty in research. Thank you also to Dr. Carrie Hall and Dr. Daniel Howard, for their guidance and support on my committee as well as letting me survey their courses for my projects. Thank you to Dr. Patty Jarema in DBS and the fine people at CU Boulder who graciously allowed me to survey their students as well, and to the Schuchardt Lab at the U of M in Minneapolis for their collaboration. Thank you to Dr. Jessica Bolker, whose invaluable wit, candor, and unwavering support these past few years I will never forget. Thank you especially to Ciara McCarter, '20 and Jackie Klombers, '21, my fantastic undergraduate research assistants, for all their drive, creativity, and engagement, whose brightness I hope will shine even greater in the next chapters of their lives. Lastly, but certainly not least, thank you to one Jordan Dana Bader, one of the most wonderful, inspiring, and powerful people I know, for her unconditional love and companionship, for being like family when my own is halfway across the world, and for being a light to me when all other lights go out.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The field of biology is becoming increasingly reliant on quantitative tools, methods, and techniques, driving a need for incoming biologists to have robust quantitative skills. However, efforts to incorporate more quantitative skills at the undergraduate level are hampered by low student engagement with math in biology. Students' motivation towards quantitative biology can provide insight into how best to increase their engagement and thus performance with these topics. This thesis examines students' motivation towards math in biology through two key constructs: 1) students' self-efficacy, through the theoretical lens of Social Cognitive Theory; and 2) students' task-values, through the theoretical lens of Expectancy-Value Theory.

In Chapter 1, I explore how students' self-efficacy towards quantitative biology problems is impacted by their experiences when working together in small groups to tackle mathematical problems in a biological context. In two sections of an introductory biology class, I surveyed students about their self-efficacy before and after completing two separate group work assignments about evaluating Hardy-Weinberg Equilibrium and modeling population growth, as well as asked them to report through short responses their experiences during those assignments which increased or decreased their confidence towards these kinds of problems. I qualitatively coded students' short responses and found that students draw from a breadth of experiences to evaluate their self-efficacy. In particular, students reported many mastery experiences which increased their self-efficacy, through opportunities to practice solving these problems, confirming their success with them, or even being able to teach and guide their peers through the problems. Students also valued how group work fostered an availability of help and support from

their peers which built their self-efficacy, through discussion, collaboration, and being able to simultaneously receive and seek help from their peers. I performed logistic regression to find that students' self-efficacy level before entering each group work assignment predicted their likelihood of reporting mastery experiences or help availability from peers as the source of their increased self-efficacy, with higher self-efficacy students more likely to report mastery experiences and lower self-efficacy students more likely to report the availability of help from their peers.

Meanwhile, I found that while most students did not report any experiences which decreased their self-efficacy, those who did described a wide range of specific experiences. Most commonly, a lack of mastery decreased self-efficacy, ranging from simply not understanding the problem or making mistakes on the problem, to being unable to complete the assignments due to a lack of time or their group rushing ahead of them, to groups not even checking their answers or progress. Some students also described a lack of availability of help from their peers or instructors, with some groups failing to communicate openly or fully collaborate to group members simply being unable to help them with no one else around for support. Students also described a handful of experiences where they compared themselves unfavorably to their peers, feeling like they were falling behind or otherwise lacking in skill, as well as a general sense of anxiety from working in groups. I performed a logistic regression to find that students' self-efficacy level before entering each group work assignment also predicted their likelihood of reporting a lack of mastery which decreased their self-efficacy, with lower self-efficacy peers more likely to describe a lack of mastery than their higher self-efficacy peers.

In Chapter 2, I explore how an alternative, multidimensional model of task-values compares to a more traditional model of students' task-values towards statistics, and how these task-values relate to their statistical understanding. I surveyed life-sciences students at two institutions about their task-values towards statistics and measured their performance on an assessment of their understanding of biological variation in an experimental design context. I performed confirmatory factor analyses to find that students' task-values towards statistics are better represented using a multi-dimensional model which differentiates the four canonical task-values—intrinsic value, attainment value, utility value, and cost—into multiple task-value 'facets', each capturing a specific aspect of each task-value, such as 'utility for school' or 'emotional cost'. After excluding attainment value due to its poor fit, my model of task-value facets includes: 1) intrinsic value, with no facets; utility value with five facets ('utility for school', 'utility for daily life', 'social utility', 'utility for career/job', 'utility for future life'); cost with three facets ('effort required', 'emotional cost', 'opportunity cost'). Using multiple linear regression, I found that students' utility value for statistics for school and emotional cost of statistics predicted their performance on the statistical assessment; students with higher utility value for statistics for school performed better than their peers with lower utility value for statistics for school, and students with lower emotional cost of statistics performed better than their peers with higher emotional cost of statistics.

My findings show how exploring students' motivation towards quantitative biology can be a helpful lens for better understanding how students engage with math in biology. I reveal a mechanism by which in-class experiences can impact students' confidence, highlighting a need for more focused work into how these specific experiences arise and how they relate to and interact with each other to shape students' self-efficacy beliefs. Understanding this mechanism may reveal more effective and positive ways to increase students' engagement with quantitative biology and reinforce their quantitative skills. Furthermore, I show how a more focused model or characterization of students' task-values can predict their performance, providing a useful tool for educators and instructors to develop lessons or interventions to bolster their students' values to increase their performance. Future work into students' values about statistics should center around exploring this multi-dimensional model of task-values in a variety of circumstances with students of different backgrounds and experiences to broaden our understanding of how these values relate to their performance and understanding.

INTRODUCTION

The field of biology is becoming increasingly reliant on sophisticated quantitative tools, methods, and techniques (NRC, 2003). These tools are required to understand and solve mounting problems in the environmental, agricultural, energy, and public health sectors (NRC, 2009). However, despite this strong impetus, biology education struggles to train undergraduates to meet these challenges, leading to numerous national calls to reform undergraduate biology curricula to better integrate quantitative skills (NRC, 2003; NRC, 2009; AAAS, 2011).

Recent reforms to directly integrate quantitative skills into introductory biology curricula have demonstrated promise through a variety of approaches. Much work has been done to varying degrees of success, such as through quantitative literacy interventions (Speth et al., 2010), online (Thompson et al., 2010) and in-class modules (Hoffman et al., 2016), as well as topic and skill related projects (Wightman & Hark, 2012; Metz, 2008). Other attempts approach the problem of integration from the other side of the equation by incorporating more biological topics and contexts such as systems modeling (Chiel et al., 2010) into mathematics courses and modules (Duffus et al., 2010; Rheinlander & Wallace, 2011). Dedicated 'math for life sciences' courses have also been developed to bridge the gap between biology and mathematical disciplines, such as calculus (Usher et al., 2010; Thompson et al., 2013b), and statistics (Watkins, 2010), with some approaches even developing completely integrated biology-math courses from the ground up in both general contexts (Depelteau et al., 2010) and specific applications like biological modeling (Hoskinson, 2010). The long-term success of these courses is still unclear, however (Marsteller et al., 2010).

Instructors face many challenges towards greater adoption and success of these approaches (Bialek & Botstein, 2004) but a significant hurdle to further integrating quantitative skills in biology classrooms is students' motivation, particularly their attitudes towards math (Colon-Berlingeri & Burrowes, 2011; Thompson et al., 2013a). Biology students often hold more negative attitudes towards math than students in other natural sciences (Wachsmuth et al., 2017), with some students potentially avoiding more math-intensive courses or instructors because they perceive them to be more difficult (Colon-Berlingeri & Burrowes, 2011; Hood et al., 2012). Addressing this slump in students' motivation is critical to further improving quantitative biology education because motivation and in particular students' attitudes can profoundly impact their willingness to engage with the curriculum (Poladian, 2013; Rheinlander & Wallace, 2011) as well as influence their perceptions towards biology as a career (Glynn et al., 2007; Matthews et al., 2013).

Social Cognitive Theory and Self-Efficacy

Social Cognitive Theory is a framework which describes how individuals' behavior is shaped by their environment as well as the behaviors of others (Bandura, 1986), and has been studied and applied in a variety of contexts including public health and education. A key component of this framework is the beliefs people hold about their capabilities and how these beliefs influence their behavior and choices, which Bandura calls self-efficacy (Bandura, 1986; 1997). In an academic context, self-efficacy represents a students' beliefs about their ability to succeed at a given task (Bandura, 1997; Bong, 2001; Bong & Skaalvik, 2003). Self-efficacy can strongly influence a student's academic success, predicting students' motivation, engagement, and thus performance and achievement on academic tasks (Klassen & Usher, 2010; Lee et al., 2014). Furthermore, self-efficacy beliefs can have an amplifying effect on motivation. Students who are highly self-efficacious and motivated can become even more so upon achieving greater success, while students who suffer repeated setbacks and low motivation frequently reinforce negative self-efficacy beliefs (Pajares, 2003). Self-efficacy is specific to different domains or subjects, with students shaping and leveraging their self-efficacy beliefs depending on what kinds of tasks they face (Bong & Skaalvik, 2003; Usher & Pajares, 2008). In particular, self-efficacy has been shown to factor into students' motivation and achievement across a variety of subjects and fields (Woolcock et al., 2016; Hutchison et al., 2006; Ainscough et al., 2016), including sub-domains like statistics (Finney & Schraw, 2003) within larger fields like mathematics (Pajares & Kranzler, 1995). Self-efficacy can also strongly impact identity with respect to science (Trujillo & Tanner, 2014), especially when students have strong or direct ties to an established mentor (Joshi et al., 2019), as well as feed into a student's ability to academically self-regulate (Lee et al., 2014). Consequently, self-efficacy beliefs can have strong impacts on career aspirations both in young children and adolescents (Bandura et al., 2001) as well as in undergraduate students (Jones et al., 2010).

Bandura (1997) describes four primary sources of self-efficacy: mastery experiences, vicarious experiences, social persuasions, and physiological states, each of which can build or harm students' self-efficacy beliefs. The most common and most impactful source of self-efficacy is the mastery experience, when students experience either success on a task through their own effort, or fail at a task despite that effort (Bandura. 1997; Usher & Pajares, 2008). Students do not have to strictly experience success or failure, however; mastery experiences can also arise as a judgement of success or a judgement of failure based on their perceptions of the outcome of the experience (Pajares et al., 2007; Usher & Pajares, 2008). Mastery experiences are especially powerful when students overcome significant challenges or obstacles, especially if they perceive the task to be difficult for them or for others (Usher & Pajares, 2009). Additionally, mastery experiences can have long-lasting effects on a student's self-efficacy (Usher & Pajares, 2008), and students often draw from prior mastery experiences when evaluating their current self-efficacy (Butz & Usher, 2015).

Oftentimes, many of the tasks students face may not have immediately observable or absolute measures of proficiency. In these circumstances, students draw from vicarious experiences (also called 'social comparisons'; Bong & Skaalvik, 2003; Butz & Usher, 2015) to gauge their abilities and skill in comparison to others. They are similar to mastery experiences in that they are a judgement of one's success but couched in relation to another person rather than one's own effort solely. Students compare themselves to a variety of others, most commonly their peers, friends, and classmates (Usher & Pajares, 2008; 2009), but also sometimes to the adults in their lives like their family or teachers (Butz & Usher, 2015). These comparisons often rely on a form of social modeling, where students draw heavily on the success, struggle, or failures of particular individuals, and the degree to which they relate to or identify with the model can affect how strongly their self-efficacy beliefs are shaped by those experiences (Schunk, 1987; Schunk & Pajares, 2002).

Social persuasions reflect the direct feedback from others about their abilities, skills, or performance on a task. This feedback can consist of encouragement or compliments directed towards the student, which a student can draw from to build their self-efficacy, or the feedback can be negative or denigrating to a student, hurting their self-efficacy (Bandura, 1997). Consequently, social persuasions are highly sensitive to the context of the situation and task, making them a more transient or fleeting influence on students' self-efficacy (Usher & Pajares, 2008). Additionally, social persuasions may be misinterpreted by students if they believe they are not genuine, and positive social persuasions may have the ability to undermine students' self-efficacy or reinforce their existing negative perceptions if a student believes the persuasions are disingenuous or placatory (Bandura, 1997; Usher & Pajares, 2008; Butz & Usher, 2015).

Lastly, physiological states (also called 'emotional' or 'affective states'; Bandura, 1997) reflect a student's emotions and feelings towards a task, or the feelings and emotions they experience while performing or after performing the task. Students can interpret their physiological state towards a task as an indicator for their expected success or failure (Usher & Pajares, 2009). Typically, these emotions include anxiety over potential failure, stress from performing the task, or dread towards engaging with the task, but ameliorating these emotions and increasing students' emotional well-being can reduce the negative impact of these stressors or even increase self-efficacy by re-framing students' anxieties (Bandura, 1997; Usher & Pajares 2008; 2009).

While these four sources are well-studied across a variety of subjects and domains, as Bandura (1997) and others (Schunk & Pajares, 2005; Usher & Pajares, 2008) have long suggested that the unique personal conditions and experiences which generate self-efficacy beliefs can manifest in more than just mastery experiences, vicarious experiences, social persuasions, and physiological states. These four sources have largely been studied using quantitative measures but may fail to capture some of the underlying complexity of the sources of self-efficacy (Usher & Pajares, 2008), leading to calls to study these sources through a qualitative lens. Recent examples of such qualitative approaches have indeed revealed the possibility for additional sources. Butz and Usher (2015), in a study focusing on primary- and secondary-school students and experiences which contributed to their self-efficacy towards math and reading, found that students not only described experiences reflective of the four primary sources of self-efficacy but also events or circumstances which seem to fall outside the confines of those four sources. They described how being able to get support or help from their peers, teachers, or other adults influenced their self-efficacy beliefs, through the 'availability of help from peers/teachers/adults/etc.,' or how the 'teaching style' or pedagogical approach of their instructor helped shape their beliefs, or even how the classroom and 'learning environment' itself was structured could influence their self-efficacy. This window into additional sources of self-efficacy beyond the four traditionally-described sources reveals an important avenue for further inquiry into how students develop their self-efficacy beliefs.

Students also frequently draw from multiple sources to build their self-efficacy. While mastery experiences are typically considered the most influential, many students will rely on a variety of vicarious experiences and social persuasions to build their beliefs (Usher & Pajares, 2008; Bandura, 1997). Early undergraduate engineering students' self-efficacy beliefs were predominantly shaped by vicarious experiences and social persuasions when working collaboratively on projects or building support groups to help each other survive the major (Hutchison et al., 2006; Hutchison-Green et al., 2008). Furthermore, self-efficacy sources may have different effects in conjunction or interaction with each other: self-efficacy beliefs may be additive, where the more information is available to form self-efficacy beliefs, the greater the belief is enhanced; multiplicative, where sources interact to shape self-efficacy beliefs; or configurative, where sources may have varying influence depending on the presence or absence of other source information (Bandura, 1997).

The sources of self-efficacy can also affect students differently depending on a variety of characteristics. In, the major sources of self-efficacy have been shown to vary along gender lines. For example, primary- and secondary-school boys experience generally higher exposures to mastery and vicarious experiences compared to social persuasions and physiological states for girls (Usher & Pajares, 2006; Butz & Usher, 2015). In undergraduate students, vicarious experiences more strongly predicted the probability of passing the course in women, while mastery experiences more strongly predicted the probability of passing for men (Sawtelle et al., 2012). Self-efficacy sources can also differ by grade and experience level, with older or more advanced students reporting less exposure and response to sources of self-efficacy than their younger peers (Pajares et al., 2007; Butz & Usher, 2015). Therefore, understanding this complexity of self-efficacy beliefs and their sources in biology specifically could therefore provide insight into how undergraduate biology students experience these sources and build their self-efficacy, increase engagement, and achieve more in their biology courses (Gogol et al., 2017). This insight may also lead to instructional development to facilitate and encourage the development of self-efficacy beliefs to further bolster students' engagement with quantitative material.

Expectancy-Value Theory and Task-Values

Expectancy-Value Theory (EVT) is a theory of motivation with widespread applications in education research, and has been studied largely in psychology (Wigfield & Eccles, 2020). In an educational context, EVT argues that a student's engagement and thus performance on a given learning task is influenced by the combination of their expectancies of success on the task as well as the set of their personal values towards that task, collectively called task-values (Eccles et al., 1983; Wigfield & Eccles, 2000; Eccles & Wigfield, 2002). The expectancy component has often been related to Social Cognitive theory and characterized as a student's self-efficacy (Eccles & Wigfield, 2002; Pajares, 1996). A student has four main task-value constructs: intrinsic value, attainment value, utility value, and cost. Intrinsic value is the enjoyment a student experiences from a given task, and reflects the degree to which they consider the task to be an end to itself (Ryan & Deci, 2016). Intrinsic value can also be characterized as a student's individual interest in the task (Wigfield & Cambria, 2010). Attainment value represents the importance of performing well on the task to a student's identity, or the extent to which the task allows them to express or confirm important aspects of their sense of self (Wigfield & Eccles, 2020). Utility value is the perceived usefulness of the task to a student's goals, both near and far, and in some way reflects a student's 'extrinsic motivations' (Ryan & Deci, 2016) towards a task, as compared to intrinsic value and motivation; the task is no longer an end to itself but merely a means to an end. Lastly, cost reflects a student's perception of the negative effects, penalties, or burden on themselves that they would incur through engaging with a task.

Task-values are extensively studied in primary and secondary-educational contexts (Wigfield & Cambria, 2010; Wigfield & Eccles, 2020). They develop early in childhood and change over time depending on academic domain (Jacobs et al., 2002), significantly impacting academic achievement (Lee et al., 2014; Simpkins et al., 2006). While less work has been done at the post-secondary level, there is evidence to suggest that task-values play a role at this level as well. In life-sciences students, students' task-values towards math in biology related to their characteristics and their likelihood of taking further quantitative biology courses (Andrews et al., 2017). Task-values may also impact performance and achievement at this level as well, both in general (Bong, 2001; Jones et al. 2010) and within domains like chemistry (Zusho et al., 2003) and mathematics (Elliott et al., 2001). Recent work has delved more deeply into the specific connection between students' task-values and their performance, but the precise nature of this relationship is not clear. Through targeted interventions, some studies have found a direct relationship between a single targeted task-value and performance, such as with utility value (Hulleman et al., 2010; Harackiewicz & Hulleman, 2010; Durik & Harackiewicz, 2007) or intrinsic value (Durik et al., 2015). Others found that task-values can relate to performance, but typically in conjunction with or moderated by other motivational constructs like self-concept or self-efficacy (Steinmayr et al., 2019; Bong, 2001; Guo et al., 2016). Notably, success expectancies in math moderate the effects of utility value on students' performance on an assessment (Durik et al., 2015). This interaction between expectancies and task-values may also be present across all task-values, as suggested by Trautwein et al. (2012) and Nagengast et al. (2011), painting a complex picture of the landscape of how this component of students' attitudes directly influence their performance.

Part of this complexity may revolve around the typical characterization of students' task-values. Traditionally, task-values have been explored as monolithic constructs, either as simply a 'task-values' component of a larger overall relationship (Steinmayr et al., 2019; Wigfield & Eccles, 2020), or often individually as part of a specific intervention (e.g., Hulleman et al., 2010; Harackiewicz & Hulleman, 2010; Durik et al., 2015), but this singular focus on individual task-values often fails to capture the underlying complexity encompassed by that task-value.

In particular, utility value may describe the usefulness of a task towards many different short- and long-term goals. Students are able to differentiate between different 'domains' of these goals, such as their daily lives, their academics, or their careers (Peetsma & van der Veen, 2011). Therefore, studies which examine only the 'utility' of a task as a single construct may not capture precisely what it means for students to find that task 'useful' to them. For example, students often see statistics courses as an important means to an end for their schooling (Evans, 2007) while also failing to recognize the relevance of statistics to their careers in biology (Evans, 2007; Hagen et al., 2013), representing a dissonance between various facets of their overall utility value. Additionally, students of different cultural backgrounds express their utility for math in different terms, with some students focusing more on the utility of math to immediate or proximal goals and others focusing instead on future or distal goals (Shechter et al., 2011). Previous studies which included utility value as part of their investigations sometimes included items which pertained to some of these different flavors or domains of utility while still considering them as part of a single utility-value construct (Conley, 2012; Luttrell et al., 2010; Trautwein et al., 2012). Conley (2012) examined math task-values in middle-school students and used utility value items from different domains: "Being good at math will be important when I get a job or go to college" refers to a student's math utility for their career or future education whereas "Math will be useful for me later in life" refers to a student's general utility for future life. Luttrell et al. (2010), in developing the Mathematics Value Inventory designed to measure each of the canonical task-values, included utility items like "I do not need math in my everyday life", referring to how students' value the utility of math on a short-term, day-to-day basis. However, these studies did not explicitly distinguish between life-domains like academics or career as separate utility value facets and treated utility value only as a singular construct, highlighting a need to systematically explore how students evaluate their utility towards different aspects of their lives.

Cost may also have distinctive underlying dimensions, and recently has been the subject of much investigation. The original definition of cost by Eccles et al. (1983) described three components to the overall task-value: the amount of time and energy spent on the task and lost for other activities (opportunity cost), the anticipated negative emotions to performing a task (emotional/psychological cost), and the effort required to succeed on the task (effort required).

Similar to utility value, prior studies included items in their cost measures which captured some (Luttrell et al., 2010; Trautwein et al., 2012; Conley, 2012) or all of these dimensions (Chiang et al., 2011), but again typically characterized cost as a singular task-value construct. Recently, more work has been done to systematically distinguish between the dimensions of cost, demonstrating evidence that students do in fact identify those dimensions (Perez et al., 2014; Perez et al., 2019; Flake et al., 2015). Perez et al. (2014) identified the three components of cost as originally described. Furthermore, they found different relationships between the three cost components and achievement, with effort required having the strongest effect, suggesting that students evaluate the various cost components differently. Flake et al. (2015) extended these results and found not only the three original components of cost, but evidence for a fourth component: "outside effort cost", representing the cost incurred by students for tasks and activities outside of the task of interest which impose an additional burden towards completing said task. Additionally, they found evidence for a relationship between effort required and expectancies, as well as a relationship between emotional cost and performance, suggesting that the specific dimensionality within cost must be considered when exploring students' motivation.

Recent research has explored a model of task-values which differentiate them into more granular, specific dimensions, showing that students indeed distinguish between various aspects of their attainment value, utility value, and cost towards a task (Gaspard et al., 2015) and that these dimensions can vary significantly between students of different backgrounds, academic success, and gender (Gaspard et al., 2017). These studies suggest that examining multiple dimensions is a ripe avenue for developing a more complete understanding of the development of students' task-values and how they impact their engagement.

Project Goals

The goal of this thesis is to better understand how biology students' motivation towards math in biology influences their engagement and performance with quantitative biology. While there are many components to understanding motivation, this thesis will focus specifically on the constructs of self-efficacy and task-values as they are described by Social Cognitive Theory and Expectancy-Value Theory. These constructs are well-studied across educational research contexts, but their significance with respect to the specific challenges in biology education is less apparent.

In Chapter 1, I explore how students' math self-efficacy beliefs can be shaped through an instructional strategy such as group work when working on quantitative biology tasks and the relationship between students' self-efficacy and what experiences they draw from to shape their beliefs. I survey introductory biology students both before and after they work together in groups to complete two different in-class quantitative biology students, asking them to report their self-efficacy towards the problems and to describe any experiences during the group work which increased or decreased their self-efficacy. Through qualitative coding of their responses, I seek to better characterize the distinct group work experiences which contribute to self-efficacy and understand how they relate to the sources of self-efficacy. I also aim to relate students' self-efficacy towards the problems prior to working in their groups with the sources of self-efficacy they report, to better understand how students of different self-efficacy levels benefit or are hurt by their experiences during group work. I hypothesize that students with higher incoming self-efficacy may rely more on mastery experiences to shape their self-efficacy beliefs, while students of lower self-efficacy may rely more on more social sources of self-efficacy to shape their self-efficacy beliefs. Ultimately, better understanding of how group work experiences influence students' self-efficacy towards quantitative biology will provide insight into how best to design or implement this instructional strategy to increase students' engagement with math in biology.

In Chapter 2, I investigate students' task-values towards statistics using a more multi-dimensional model to better understand the relationship between students' task-values and their understanding of statistical concepts as measured by their performance. I survey introductory statistics students at two different institutions about their specific task-values towards statistics as well as measure their understanding of statistical concepts as measured by their performance on an assessment. Using confirmatory factor analyses, I seek to validate a multi-dimensional model of task-values in comparison to a more traditional model of the four main task-values, which I believe will be a better representation of students' task-value beliefs than as traditionally described. I also relate students' task-values towards statistics to their performance on the assessment, hypothesizing that students with higher utility values towards statistics will perform favorably to their peers with lower utility values. Additionally, I believe that students who perceive lower cost to engaging with statistics will perform better on the assessment than their peers with higher perceived cost towards statistics. Characterizing students' task-values on a more granular level may provide a more specific understanding of how students value an important aspect of quantitative biology such as statistics, and how these values impact their performance and understanding of statistical concepts. Understanding students' motivation through students' task-values and self-efficacy can provide biology educators and instructors insight into how best to incorporate and integrate quantitative biology into modern curricula and improve students' skills for using math in biology.

CHAPTER 1 — WHAT HAPPENS WHEN STUDENTS WORK TOGETHER? THE IMPACT OF GROUP WORK ON SELF-EFFICACY TOWARDS QUANTITATIVE BIOLOGY

Introduction

Despite numerous calls emphasizing the importance of quantitative skills in biology education (NRC, 2003; AAAS, 2011), educators struggle to incorporate quantitative biology into the undergraduate classroom. A variety of approaches to directly integrate quantitative skills into introductory biology curricula exists—to varying degrees of success—such as folding quantitative topic modules and projects into biology courses (Thompson et al., 2010; Hoffman et al., 2016; Metz, 2008), biological topic models into math courses (Chiel et al., Duffus et al., 2010; Rheinlander & Wallace, 2011), and even developing fully-integrated 'math for life-sciences' courses from the ground up (Usher et al.; 2010, Watkins, 2010, Depelteau et al., 2010). A significant challenge towards greater adoption and success of these approaches is students' motivation and engagement with math in biology (Colon-Berlingeri and Burrowes, 2011; Thompson et al., 2013a), with many biology students expressing a strong negative perception of math (Wachsmuth et al., 2017), leading to their poor engagement (Poladian, 2013; Rheinlander and Wallace, 2011).

One strategy to address these issues is through active learning strategies such as group work. Much work has been focused around the effectiveness of group work and its impact on student performance and engagement (Hodges, 2018). The structured and interactive nature of group work can increase student performance on high-risk assessments (Haak et al., 2011). Working in groups provides students with a variety of social and cognitive benefits, such as opportunities to learn from peers and leverage their unique knowledge and background, as well as practice building a consensus with their peers, which can also increase engagement (Nokes-Malach et al., 2015). Group work is not without its limitations, however. Students frequently report displeasure at having to work in groups despite their increase in performance, citing issues of unequal participation of group members and a perception of group activities as 'busy work' (Chang & Brickman, 2018). Performance is also sensitive to how a group was formed, the characteristics of the students in the group, and the resulting group dynamics when working together (Donovan et al., 2018; Chang & Brickman, 2018). These dynamics can create stressful interactions between overconfident or overbearing students with their peers, stifling discussion and harming group cohesion, and ultimately impacting their performance (Theobald et al., 2017), but their underlying complexity makes the mechanism by which they impact students' engagement less clear.

Theoretical Framework

Social Cognitive Theory, proposed by Bandura (1986) to describe how individuals and their behavior are shaped by their environment and their peers, may provide a useful framework for examining students' engagement through group work. Specifically, Bandura identifies the importance of self-efficacy, which represents a students' beliefs and judgements about their ability to succeed at a given task (Bandura, 1997; Bong, 2001; Bong & Skaalvik, 2003). Self-efficacy can strongly influence a student's academic success in numerous ways in a wide range of fields (Woolcock et al 2016; Hutchison et al., 2006; Ainscough et al., 2016), and can predict students' level of engagement with a task and their academic achievement (Klassen & Usher, 2010; Lee et al., 2014; Pajares & Kranzler, 1995; Britner & Pajares, 2006). Furthermore, self-efficacy beliefs can have a compounding effect on motivation: highly motivated students can become even more so upon achieving greater and greater success, while struggling students frequently reinforce negative self-efficacy beliefs with every subsequent failure (Pajares, 2003).

Students derive their self-efficacy beliefs from a handful of sources: mastery experiences, vicarious experiences, social persuasions, and physiological states (Bandura, 1997). Mastery experiences are when students personally experience success or failure at a task through their own work, generating new self-efficacy beliefs after reflecting upon the outcome of their efforts (Bandura, 1997). Mastery experiences do not necessarily have to explicitly represent success or failure, only that a student judges themselves as gaining or losing mastery (Bandura, 1997; Usher & Pajares, 2008). Vicarious experiences (also known as 'social comparisons'; Bong and Skaalvik, 2003; Butz and Usher, 2015) are when students compare their own level of success with that of their peers, generating self-efficacy beliefs upon determining their relative level of success to their peers (Bandura, 1997). Comparing one's own success to that of one's peers through a vicarious experience can depend strongly on how similar one is to the subject of the comparison; if a student believes that their peer is similar to them and observes them succeed on a task, they may feel more strongly that they can also achieve the task than if they aspire to their peer or look down on them (Usher & Pajares, 2008). Social persuasions occur when students receive direct feedback from their peers or instructors about their performance on a task, the evaluation of which can result in self-efficacy beliefs (Bandura, 1997). They are also sensitive to the tone of the feedback and context in which the feedback was given (Usher & Pajares, 2008); disingenuous or unwarranted feedback may have a stronger negative effect than positive. Physiological or affective states represent a student's emotions and feelings towards a task, such as anxiety over potential failure or satisfaction of performing the task. These feelings can contribute or amplify existing self-efficacy beliefs (Bandura, 1997).

While students predominantly express the significance of mastery experiences in contributing to their self-efficacy both positively and negatively, students also weigh other sources when forming self-efficacy beliefs. Many students also draw from vicarious experiences and social persuasions to build their beliefs (Usher and Pajares, 2008; Bandura, 1997). For instance, in engineering undergraduate students, working in teams and collaborating with each other produced opportunities for vicarious experiences and social persuasions, which significantly increased self-efficacy beliefs across the board compared to standalone mastery experiences (Hutchison et al., 2006; Hutchison-Green et al., 2008). When asking primary and secondary students faced with reading and math tasks, students reported specific aspects of vicarious experiences, such as distinguishing comparisons between their peers and the adults in their lives, as sources of self-efficacy (Butz & Usher, 2015). Additionally, this study also found evidence for sources of self-efficacy outside the four traditionally described by Bandura (1997) and self-efficacy theory. Students identified the nature of guidance and help provided by their peers, teachers, and other adults, such as a teacher's instructional style or how available they were to the student, as salient sources of self-efficacy (Butz & Usher, 2015). This complexity in how students characterize and experience the various sources of self-efficacy highlights the need to better understand the process by which these experiences arise and how they influence students' self-efficacy beliefs.

Influencing Self-Efficacy through Group Work

At first glance, the interactive nature of group work lends itself to generating experiences which build self-efficacy (Usher & Pajares, 2008). Group work can impact students' self-efficacy and engagement by fostering a sense of collaboration rather than competition with their peers (Springer et al., 1999). When working in small groups, social work students experienced an increase in self-efficacy compared to working independently (Öntaş & Tekindal, 2015). Working together in groups builds a sense of collective efficacy which in turn strongly relates to an individual student's self-efficacy beliefs and performance (Lent et al., 2006). Specifically, groups with higher self-efficacy beliefs tend to not only reinforce those beliefs but also increase the use of high-level cognitive skills during class discussions, as well as academic performance (Wang & Lin, 2007). Undergraduate engineering students also frequently reported group work as a significant contributor to their self-efficacy beliefs and persistence in their academic career (Hutchison et al., 2006; 2008).

However, it is unclear how students experience the sources of self-efficacy through group work and how different sources may arise depending on specific group work interactions. For example, students working on a problem in class in their groups can engender mastery judgements by providing them opportunities to succeed or fail at the task. When working or discussing collaboratively, students may compare their problem-solving methods, strategies, or abilities and those of their peers, allowing them to leverage vicarious experiences to develop their self-efficacy beliefs. The nature of these discussions may also result in different outcomes, such as whether a conversation promotes encouraging and constructive social persuasions or creates a confrontational or judgmental environment where students may experience negative social persuasions. Working in groups may also alleviate or amplify a students' physiological states or emotions towards a task, either by creating an environment where they feel supported or generating stress when the group fails to work together effectively. Consequently, characterizing what students actually experience when working together in groups specifically with respect to their self-efficacy and what sources it stems from may generate insight into how to better design group work activities to increase their engagement and understanding of difficult-to-wrangle topics like those in quantitative biology.

Research Goals

Our study had two primary goals. First, we sought to better understand how self-efficacy beliefs arise and are influenced through the specific experiences students have when working together in groups. We explored what experiences introductory biology students had during quantitative biology group work which positively and negatively impacted their self-efficacy with respect to two specific mathematical tasks: calculating Hardy-Weinberg Equilibrium and modeling population growth. We hypothesize that group work generates experiences which influence students' self-efficacy beliefs, and we expect to find a preponderance of mastery experiences given their significance in developing self-efficacy beliefs (Bandura, 1997; Usher & Pajares, 2008), but given the social nature and complex emergent dynamics of group work, we also anticipate a variety of experiences which draw from the more social sources of self-efficacy, such as availability of help or social persuasions (Butz & Usher, 2015). Second, we sought to understand how students of differing self-efficacy levels experience and report the sources of self-efficacy through group work. We hypothesize that, after controlling for students' gender as well as the specific topic of the group work assignment, students' self-efficacy level prior to the group work relates to their likelihood of describing a mastery experience, the availability of help from their peers, or a lack of mastery. We predict that students with higher self-efficacy may be more inclined to report mastery experiences than their lower self-efficacy peers when evaluating what experiences increased or decreased their self-efficacy, because their higher self-efficacy towards the mathematical problems may be the result of prior mastery experiences and therefore are more likely to draw similarities between their present and past mastery. We also predict that lower self-efficacy students might rely more on more help availability, a social source of self-efficacy, to influence their beliefs than their higher self-efficacy peers.

Methods

Participants and Setting

We surveyed 337 undergraduate students at a large public research university in the Northeast. These students were drawn from two sections of an introductory biology course held in the Fall of 2019, each section taught by a different instructor. This course is one of two required introductory biology courses for many life-science majors at this institution, although non-life-science students may take this course to fulfill a general education requirement. The course was structured as a large lecture—roughly 150 to 200 students per section—in which students were assigned to groups of three to five students, meeting several times per week. Groups were assigned according to students' seating preferences, such as preferring to be near the front of the room or off to one side. These groups sat together during class sessions and remained together throughout the semester. Groups would frequently work together during a class session to complete a collaborative, low-stakes assignment based on the lecture content of the prior class session. These assignments consisted of several multi-part questions about the lecture topic within an authentic biological context. The curriculum covers a variety of topics in evolution, ecology, and biodiversity, such as the mechanisms and principles of evolution, biological speciation, ecological competition, and included specific quantitative biology topics such as evaluating Hardy-Weinberg Equilibrium and modeling population growth. Participants' demographic information is summarized in Table 1.1. This study was approved by the Institutional Review Board at the University of New Hampshire, IRB # 7005.

Data Sources and Sampling Method

Our study centers around two in-class group-work assignments: one where students evaluated Hardy-Weinberg Equilibrium (HWE), and another where students modeled population growth (PG). In one section, the HWE assignment centered around a modified case study about the conservation of Timber Rattlesnakes, *Crotalus horridus*, in New England (Drott & Sarvary, 2016), while the assignment in the other section centered around scenarios about two physical traits in humans, hair texture and the unibrow phenotype. For PG, the assignment for both sections was the same, centering around the population growth of two invasive species. In both course sections, the HWE assignment occurred early on in the semester (week 3 of 15 in one section, week 4 of 15 in the other), while the PG assignment occurred in the latter half of the semester (week 12 of 15 in both sections).

On the day of each assignment, prior to starting the group work, we provided students with a pre-survey, consisting of a sample problem similar to the questions on the actual assignment, and asked students to simply consider the problem but not actually solve it (Appendix A). For HWE, this example consisted of a table of genotype frequencies in a population of a single trait with two alleles, asking students to “calculate the predicted number of individuals of each genotype under the conditions of Hardy-Weinberg Equilibrium.” For PG, the example consisted of a short scenario describing the initial population of a fishing stock, the initial observed births and deaths of the population, the carrying capacity of the population, and a period of time over which to model the growth of the population, asking students to “calculate the population size in the year 2022.” We then asked students to report their confidence in their ability to solve the sample problem using a five-point scale, ranging from “1 - Not at all confident” to “5 - Completely confident.” After we collected this pre-survey, students worked together in their groups over the course of the class session to complete the assignment. The instructors of each section as well as a number of undergraduate and graduate teaching assistants would circulate throughout the room to supervise students and answer any questions which arose during the group work. Following the end of the class session, we administered a post-survey using the online service Qualtrics to students, delivered via the course's online learning management system. This post-survey consisted of the exact same sample problem we provided students on the pre-survey, once again asking them to only consider the problem and report their confidence in their ability to solve the problem using the same five-point scale (Appendix A). This combination of pre- and post-surveys provided us with students' self-efficacy towards HWE and PG both before and after working with their group. Additionally, this post-survey asked students to "describe any experiences and/or interactions which *increased* your confidence in your ability to [solve the sample problem]," as well as "describe any experiences and/or interactions which *decreased* your confidence in your ability to [solve the sample problem]." These short responses provided us with a qualitative description of students' group work experiences which may have increased or decreased their self-efficacy for both HWE and PG. Students' self-efficacy scores for each assignment, pre- and post-, were then paired with their respective short responses for analysis. Students received course credit for completing the post-survey online equivalent to one homework assignment, while participation in the research study was optional.

Data Analyses

Qualitative Coding of Short Responses

To explore what experiences students had during group work which affected their self-efficacy, we qualitatively coded students' responses. We drew heavily from Social Cognitive

Theory and the sources of self-efficacy described by Bandura (1997), Usher & Pajares (2008) and expanded upon by Butz & Usher (2015) to inform our coding choices and to provide a theoretical basis for this goal of our study. We relied on both *deductive* (theory-based, codes established *a priori*) and *inductive* (codes emergent from the data) strategies to code students' short responses into salient experiences which influenced their self-efficacy. To capture these experiences, we conducted *process coding*, which involves using gerunds ('-ing' words) to describe events, occurrences, or ongoing action in a situation of interest (Saldaña, 2015). Process coding was especially useful for our study because we were interested in encapsulating specific moments during a student's overall group work experience which may have impacted their self-efficacy (Saldaña, 2015).

To accomplish this, students' short responses describing their group work experiences were extracted from the survey results and de-identified. Students who did not complete the post-survey for an assignment (and therefore did not provide any short responses) and students who declined to give consent to the use of their responses in our study were excluded from our analyses for each topic. This resulted in a sample of 311 students out of the 337 surveyed. We compiled students' responses to each question into four documents, one for each set of question topic and the direction of the experience: 230 students reported experiences which increased confidence in HWE, 218 students reported experiences which decreased confidence in HWE, 231 students reported experiences which increased confidence in PG, and 230 students reported experiences which decreased confidence in. We started our exploration by reviewing students' responses in each set to get an overall sense of what experiences students described. Three members of the research team parsed through all student responses to develop a preliminary codebook of students' group work experiences, both those which increased their self-efficacy and those which decreased their self-efficacy. Throughout this preliminary coding, the team relied on a process of regular introspection and reflection of students responses and how each researcher elected to code each response, called 'writing analytic memos', which helped us organize thoughts, identify patterns and notable responses, and develop a deeper understanding of the responses as a whole to ensure consistency in the codebook (Saldaña, 2015). This preliminary codebook included *a priori* codes which were intended to capture the defining experience of each of the sources of self-efficacy, based on theory (Bandura, 1997) and findings of other self-efficacy studies (Butz & Usher, 2015). Each code was then categorized by the source of self-efficacy to which it pertained.

For experiences which increased self-efficacy, two members of the research team independently process-coded two 'training rounds' of 40 responses each, drawn from both HWE and PG and from both sections of the course. Following each training round, we discussed the codes we assigned to each response and resolved any disputes or disagreements. We iteratively re-examined previously coded responses whenever we added new codes to the codebook or definitions of *a priori* codes changed, based on our discussions and the analytic memos we wrote and reflected upon. Following these two training rounds and the solidification of the codebook, the research team independently coded an additional third and fourth round of 40 responses each, following a similar pattern to the training rounds, while also assessing inter-rater reliability (IRR) for each of our developed codes. Per the recommendations of Xu and Lorber (2014), we established Holley and Guilford's G-index (Holley and Guilford, 1964) as our metric for IRR, because of its general robustness with skewed responses, which we expected our data to exhibit. We established thresholds for achieving IRR based on the recommendations of Hruschka et al. (2004): *index of agreement* (G) > 0.80 for most (e.g., > 90%) of codes. We calculated G-indices for the third and fourth rounds of coding and achieved our threshold for IRR across all of our codes, with a minimum *G*-index of 0.75, which occurred on only one code in the third round of coding. Following this, one member of the research team coded all of the remaining experiences which increased self-efficacy independently. They then shared a random selection of 30% of these remaining coded responses with the other member of the research team to confirm our IRR again. We also met our threshold for IRR for this last set of codes, with a minimum *G*-index of 0.93.

For experiences which decreased self-efficacy, we followed a similar procedure to experiences which increased self-efficacy. Two members of the research team independently process-coded two training rounds of 40 responses each, compiled identically as before. Once again, we iteratively re-coded previous items as new codes were added to the codebook or as definitions changed. These two rounds required more revision and discussion than with the increased responses, so we elected to conduct a third training round of 40 responses to further solidify our codebook. Following these training rounds, the research team conducted a fourth and fifth round of coding, again 40 responses each, and calculated *G*-indices for all our codes. While we achieved our threshold for IRR across all of our codes, with a minimum *G*-index of 0.7 on one code in the fourth round of coding, the research team decided to both independently code all the remaining decrease experiences rather than simply a random subset as with increase experiences. We based this decision on our discussions during the fourth and fifth coding rounds, as we developed two additional codes in the fifth round of coding. For these remaining experiences, we also met our threshold for IRR, with a minimum *G*-index of 0.98.

Generalized Linear Mixed Models of Students' Self-Efficacy and Reported Sources of Self-Efficacy

To investigate whether students of varying math self-efficacy report different sources of self-efficacy during the group work, we examined several generalized linear mixed models (GLMM). Given the wide variety of experiences students reported and the range of sources of self-efficacy, we elected to model only the most commonly reported sources across both sets of responses to address this research goal. For experiences which increased self-efficacy, we modeled the relationship between students' incoming self-efficacy towards the quantitative problems prior to the group work assignment and whether or not a student reported a mastery experience, as well as the relationship between incoming self-efficacy and whether or not a student reported help availability from their peers. For experiences which decreased self-efficacy, we modeled the relationship between incoming self-efficacy and whether or not a student reported a lack of mastery.

For all our models, we included as predictors: 1) the question type—HWE or PG—about which the student was describing in their response, as we wanted to control for any differences in sources of self-efficacy reported due to the type of problem students were facing; 2) gender, as we wanted to control for any differences in reported self-efficacy sources between genders, as gender differences in self-efficacy have been previously found (Usher & Pajares, 2008; Butz & Usher, 2015). Our explanatory variable in each model was students' incoming self-efficacy as measured by their pre-survey self-efficacy score, and our response variable was students' reported self-efficacy source, expressed as a binary outcome of whether or not a student reported that source in their response. Because we sampled the same students from two different points in the semester, because students were assigned into groups which meant that some students shared similar experiences, and because we sampled from multiple sections, we initially included these factors as random effects. However, due to lack of convergence, we instead included class section and assigned group as fixed effects instead (Theobald et al., 2018). Model selection and quantitative analyses were conducted using the standard R v. 4.0.2 packages for generalized linear modeling (R Core Team, 2020), 'lme4' (Bates et al, 2015) to test each model and evaluate odds ratios, and 'effects' (Fox & Weisberg, 2019) for plotting our results. The final models for each self-efficacy source are detailed below, with included random effects marked in parentheses.

Mastery Experiences - Increased Self-Efficacy

Reporting of Mastery Experiences ~ Self-Efficacy Level + Question Type + Gender + Section + Group + (Student)

Help Availability (Peers) - Increased Self-Efficacy

Reporting of Help Availability from Peers ~ Self-Efficacy Level + Question Type + Gender + Section + Group + (Student)

Lack of Mastery - Decreased Self-Efficacy

Reporting of a Lack of Mastery ~ Self-Efficacy Level + Question Type + Gender + Section + Group + (Student)

Results and Discussion

We assigned a total of 1036 process codes to our 983 recorded responses: 541 codes to experiences which increased self-efficacy and 495 codes to experiences which decreased self-efficacy. We identified ten distinct process codes for experiences which increased self-efficacy (Table 1.2) and twenty distinct process codes for experiences which decreased self-efficacy

(Table 1.3). We additionally included two 'neutral' codes which are shared between increase and decrease experiences: "no impact", for when students expressed that no experiences occurred during group work which impacted their self-efficacy, and "non-answer", for when students provided an incomplete, unintelligible, or irrelevant response. We identified and categorized our codes into seven sources of self-efficacy. These sources were derived from the four original descriptions by Bandura (1997) as well as three additional sources as identified by Butz and Usher (2015): help availability from peers, help availability from teachers, and learning environment. In this section, we discuss the overall patterns in students' experiences which increased or decreased their self-efficacy, going through each source of self-efficacy within both sets of experiences. Within relevant sections, we also detail the results of our GLMMs exploring how students' self-efficacy level relates to their likelihood of reporting a specific source, e.g., mastery experiences.

Experiences which Increased Self-Efficacy

As we expected, a considerable proportion of experiences which increased self-efficacy reported by our students were mastery experiences (Figure 1.1). However, the most common experiences reported by students actually reflected the availability of help from their peers. A few students described the availability of help from their teachers/instructors. Students reported no vicarious experiences, social persuasions, or physiological states increased their self-efficacy. Encouragingly, only a small proportion of students reported no experiences increased their self-efficacy.

Mastery Experiences

We identified three salient mastery experiences (Figure 1.1) which increased students' self-efficacy: 1) 'accomplishing it', which represents the act of succeeding at a part of all of the group work, as well as a general feeling of accomplishment simply by working on the problem and practicing it; 2) 'confirming their answers', which represents confidence in their answers or progress through checking their work and results with their group members; and 3) 'teaching / guiding others', which represents a feeling of mastery because they were able to help explain or teach a part of the group work to their peers. While mastery experiences can represent moments where students experience success and can verify it directly, as with 'accomplishing it' or 'confirm their answers', mastery experiences also encompass moments where students judge themselves as successful based on their sense of progress or accomplishment, as is the case with 'teaching / guiding others' and feeling confident in their own ability because they are skilled enough to guide their peers.

The most commonly reported mastery experience was 'accomplishing it' (Figure 1.1). In general, students noted the benefit of simply being able to practice solving problems about Hardy-Weinberg Equilibrium and modeling population growth. Many students indicated that prior to starting the group work, they were apprehensive about their understanding of the topics and were not sure whether they could succeed at the assignment, but once they had worked through the problems in their groups their confidence in their ability to do so increased. While this particular aspect is not necessarily unique to a group work setting, as simply providing students with individual practice may also benefit them, there are still some ways in which group work is especially impactful. Some students specifically indicated that after their group had worked through some problems together initially, they were then able to complete later problems on the assignment on their own. This 'easing into the water' interaction in groups appears to allow students to overcome an initial doubt or lack of confidence with their peers and lets them build their own mastery, which is crucial for building salient self-efficacy beliefs (Bandura, 1997; Usher & Pajares, 2008). They also valued being able to work on specific parts of the assignment collaboratively rather than needing to navigate the entire assignment by themselves. These experiences highlight how the collaborative environment of group work can provide benefits on top of merely providing students with independent practice, by helping mitigate some of the cognitive load of working through problems independently (Nokes-Malach et al., 2015).

The second most commonly reported mastery experience was ‘confirming their answers’ (Figure 1.1). Students valued being able to compare their answers with their peers to get a sense of their success and progress, often reporting that working on a problem on their own, then checking their answers with their peers and getting the same result boosted their confidence in their own answers. Some students specifically mentioned that going into the assignments, they believed they understood how to evaluate Hardy-Weinberg Equilibrium or model population growth and had methods to approach these problems, and upon seeing their group-mates get the same answer as their own, they felt more confident that their method was indeed successful. This self-checking and reinforcing aspect of group-work is notable; in a large-lecture setting with hundreds of students, instructor attention is at a premium even with the help of teaching assistants, so this increased ability for groups to resolve small or simple misconceptions and mistakes is especially helpful to prevent instructor assistance from being spread too thin.

Lastly, some students also reported a considerable increase in confidence in their own ability to solve the problems because they were able to teach or guide their peers about the topic of the assignment (Figure 1.1). They described how being able to explain to someone else how to solve a part of the assignment or how to reason through a problem reflected that they fully understood the material themselves, which increased their confidence in solving the problem on their own. Some even mentioned how systematically working through problems with peers who needed help reinforced their own knowledge and understanding, even if they were not fully confident going into the group work. While the focus of most instructors is typically on helping to increase the self-efficacy of less-confident students, these experiences are especially exciting and encouraging, as they showcase not only the best of how students are able to help each other learn and increase their confidence and engagement, but also how the act of teaching or guiding someone else is beneficial even to confident students. This also dovetails with our previous observation about the reinforcing and self-corrective aspects of group work, as students who are able to explain concepts or problems to their peers additionally serve to optimize the attention of the instructors to the groups which most need help.

To investigate our second research goal, we wanted to determine whether students of different incoming self-efficacy levels were more or less likely to report mastery experiences as a source of self-efficacy. We examined a GLMM examining students' pre-assignment math self-efficacy levels and their reporting of an experience which increased their self-efficacy which reflected a mastery experience, controlling for the question topic (HWE or PG), their gender, class section, and assigned group as fixed effects, with student as the only random effect. We checked the assumptions of this model (Theobald et al., 2019) and verified that: 1) the outcome variable is binary (reported a mastery experience, yes or no); 2) our observations are not independent, but we are accounting for this lack of independence using our random effects; and 3) our predictor variable is linearly related to the logit of the outcome variable. We found that students' math self-efficacy significantly increased the log odds of reporting a mastery experience (β: 0.622; standard error 0.116, $p < 0.001$; Table 1.4). For a one-unit increase in self-efficacy levels, the odds of a student reporting a mastery experience are 1.9 times greater than the odds of a student not reporting a mastery experience ($e^{\beta} = e^{0.622} = 1.862$). Therefore, higher self-efficacy levels are related to a greater probability of reporting a mastery experience when making a judgement about a positive group work experience (Figure 1.2). This aligns with our expectations that students with higher incoming self-efficacy would tend to rely more on mastery experiences when evaluating an increase in self-efficacy compared to their lower self-efficacy peers. Butz and Usher (2015) also observed this tendency with their students when asked about self-efficacy sources with respect to math. One reason for this tendency may be that students who have such high self-efficacy going into the assignment have experienced success in math before these quantitative group work assignments, making them more inclined to consider new successful experiences when further evaluating their self-efficacy (Usher & Pajares, 2008). A handful of our students did in fact indicate that they worked with Hardy-Weinberg Equilibrium problems earlier in their academic career and that working on it again and solving the problems correctly increased their self-efficacy further.

Availability of Help from Peers and Teachers

As described by Butz and Usher (2015), the availability of help from peers represents how help or guidance from a student's peers or even merely the potential for such help contributes to the student's self-efficacy beliefs. Similarly, the availability of help from teachers reflects how the assistance or presence of instructors can affect how a student judges their self-efficacy. We identified three distinct experiences for help from peers (Figure 1.1): 1) 'discussing / working together', which represents the benefit of simply being able to talk to group mates, discuss ideas, and work through the problems together; 2) 'being taught / guided', which represents experiences where students received clarification or help from their group mates; and 3) 'asking questions', which represents being able to actively ask questions and seek help from their peers. For help from teachers, we identified one experience: 'consulting with a teacher', which represents experiences where students received help from an instructor or teaching assistant, or were able to ask an instructor for help directly.

The most commonly described experience of help availability from their peers was ‘discussing / working together’ (Figure 1.1). Similarly to ‘accomplishing it’, most students indicated that simply being able to work in a group and talk to their peers was beneficial, as it allowed them to share ideas and thoughts. Some students even expressed that they communicated or discussed answers with people outside of their own group, revealing the possibilities for larger ‘super groups’ to emerge in large shared spaces like a lecture hall. Many students expressed that discussing through a problem helped them feel better about their own answers or methods, even if there was no explicit checking as with ‘confirming their answers’. Some students appreciated that their group members had different opinions and perspectives, which allowed them to refine their own methods or correct misconceptions, increasing confidence in their ability to tackle the problems, which represents a known benefit of working in groups (Nokes-Malach et al., 2015). Students reported that the setting also enabled the group to collectively assist someone who was having difficulty, in some ways acting as a more distributed version of ‘teaching / guiding others’, which further demonstrates the self-guiding and adjusting element of group work. Some students even felt that talking with their group mates was easier and more comfortable than asking an instructor for help, which, while uncommon, is an important consideration for instructors, especially those in a large-lecture setting where students may already feel distant (Gill, 2011).

Being able to both ask questions and receive help from their peers were also important interactions for students (Figure 1.1). In many specific circumstances, not only did students report that they were able to ask their group mates questions about difficulties they were having, they also described instances where other group members stepped in to help them of their own accord, without the student soliciting help directly. Many students reported that even though they may have struggled or even gotten something incorrect, having other group members around them provide clarification or support enabled them to ultimately build their confidence in solving later problems because they now knew the correct approach. These responses are especially interesting as they provide a small glimpse into how experiences which could potentially harm a student' self-efficacy can end up being resolved in the moment through a group interaction and turned into constructive judgements instead. Bandura (1997) argues that moments such as these may not be as impactful as achieving mastery directly through one's own effort, but may still provide an avenue for students to build their self-efficacy beliefs. Several students also reported that receiving help from the instructors and teaching assistants was beneficial in general. Some valued that they were able to receive one-on-one attention from an instructor, or were able to resolve confusion or misunderstandings which deadlocked the student's or their groups' progress. This particular kind of experience highlights the importance of instructor interaction in a group-work setting, as even though groups are able to self-guide and reinforce each other, not all issues can be overcome by the student or group themselves, and having the fallback of reaching out to an instructor is still a significant component of improving students' self-efficacy.

We also wanted to determine whether students of different self-efficacy levels were more or less likely to rely on help availability from peers to evaluate their self-efficacy. We examined a GLMM examining students' post-assignment math self-efficacy levels and their reporting of an experience which increased their self-efficacy which reflected the availability of help from their peers, controlling for the question topic (HWE or PG), gender, class section, and assigned group as fixed effects, with student as the only random effect. We checked the assumptions of this model (Theobald et al., 2019) and verified that: 1) the outcome variable is binary (reported of availability of help from peers, yes or no); 2) our observations are not independent, but we are accounting for this lack of independence using our random effects; and 3) our predictor variable is linearly related to the logit of the outcome variable. We found that students' math self-efficacy significantly decreased the log odds of reporting the availability of help from peers, (β: -0.374; standard error 0.118, $p = 0.002$; Table 1.4). We additionally found that class section was a significant predictor (β: -1.467; standard error 0.448, $p < 0.001$; Table 1.4). For a one-unit increase in self-efficacy level, the odds of a student not reporting the availability of help from peers are 1.5 times greater than the odds of a student reporting the availability of help from peers ($1/e^{\beta} = 1/e^{-0.374} = 1.453$). Therefore, greater self-efficacy levels are related to a lower probability of reporting the availability of help from peers when making a judgement about a positive group work experience (Figure 1.3). This also aligned with our expectations that students of lower initial self-efficacy may rely more on the social sources of self-efficacy like help availability than their higher self-efficacy peers. Unlike some of their higher self-efficacy peers, students with lower self-efficacy going into the assignment may not have had experience with solving Hardy-Weinberg Equilibrium or modeling population growth, meaning they have no prior mastery from which to draw a sense of confidence. These students instead may look to their peers for guidance and seek to 'talk it out' with others rather than needing to work alone, as we found in many of our students' responses.

Vicarious Experiences, Social Persuasions, and Physiological States

For vicarious experiences and social persuasions, we determined our codes *a priori* based on the self-efficacy literature and theory about these sources (Table 1.2). For vicarious experiences, we established the code "comparing themselves positively", which reflects the canonical definition of a vicarious experience where a student compare their abilities to those of their peers and makes a favorable or positive judgement about themselves. For social persuasions, we established the codes ‘getting positive feedback from peers’ and ‘getting positive feedback from teachers’, which similarly reflect the definition of social and verbal persuasions by Bandura (1997) where students receive encouragement or appraisal of their abilities from their group mates or instructors. Lastly, while emotional and physiological states are known to impact self-efficacy beliefs (Bandura, 1997), these emotions are predominantly described as negative influences, such as anxiety, stress, or dread related to a task. Therefore, we did not expect to find any responses which increased self-efficacy through a physiological state and established no *a priori* codes for these experiences.

We did not observe any instances of emotional or physiological states resulting in an increase in self-efficacy, yet we notably did not observe any responses where students compared themselves positively to their peers (vicarious experiences), or where students expressed that they received encouragement or feedback from their group mates or the instructors outside of simply receiving help or clarification on the problems (social persuasions). The lack of vicarious experiences is conspicuous for a number of reasons. As group work is transparent and collaborative by design, given that students are working in close physical proximity on a shared task, this setting would seem to provide ample opportunities for students to compare their abilities with that of their peers. The absence of social persuasions in the form of positive feedback from peers or instructors is less unusual, however and may reveal a possible explanation for the absence of either source in our responses. Social persuasions may be limited in how significant or impactful they are in improving students’ self-efficacy, and more often may undermine students’ self-efficacy beliefs rather than reinforce them (Bandura, 1997; Usher &

Pajares, 2008). Our surveys asked students to simply report any experiences which increased or decreased their confidence in solving the example problem in a short response, meaning that students were free to describe any number of the wide variety of experiences that they did. More often than not, students typically responded by describing only one or two salient experiences; it is distinctly possible that due to the relative significance or importance of other sources of self-efficacy—such as mastery experiences or help availability—in making judgements about one's own experiences, students simply did not end up reporting any vicarious experiences social persuasions even if they did occur. Therefore, we cannot claim that students working in groups do not rely on these two sources to develop their self-efficacy beliefs.

Experiences which Decreased Self-Efficacy

Encouragingly, we found that the most commonly reported experience which decreased self-efficacy (Figure 1.3) was none at all: the majority of students who responded reported that group work had 'no negative impact' on their self-efficacy. Of the responses which did indicate a decrease in self-efficacy, we found that most of the experiences students reported were reflective of a lack of mastery, followed by a lack of availability of help or support from their peers. A small proportion of students also reported vicarious experiences comparing themselves negatively to their group mates, while marginal proportions of students expressed a lack of availability from teachers, social persuasions, or physiological states. Notably, a number of these negative experiences appeared to be 'transient' during the group work, in that students reported that an experience decreased their confidence but later events or experiences during the group work resolved the issue or even increased their confidence in the end. For our analyses, we coded the relevant parts of these responses with our decrease codes while leaving out non-negative experiences.

Lack of Mastery Experiences

In contrast to experiences which increased self-efficacy, we identified a considerable breadth of experiences in which students expressed a lack of mastery over the assignment. Students reported nine distinct mastery experiences (Figure 1.3). Most commonly, students reported that generally 'lacking understanding' of the content or the assignment decreased their confidence in solving the problem. While most of these lapses in understanding were general or non-descript, many students specifically noted that what decreased their confidence were moments when the entirety of their group were lacking understanding and unable to figure out a problem or part of the assignment. Ultimately, we coded such experiences as 'lacking consensus', where students felt a lack of mastery because they were unable to agree or figure out a solution to their misunderstanding. Related to these experiences are when students indicated their group 'failed to confirm their answers', representing how students remained unsure of their answers or methods because their group did not engage in cross-checking. We highlight these experiences because they reveal a possible deficiency in the self-corrective and guiding aspect of group work that we identified in the previous section. If a group is unable to or unwilling to collaborate in this way, students are left without the ability to create their own mastery.

This theme of an 'inability' to create their own mastery appears in other experiences as well (Figure 1.3). Students occasionally reported that, aside from 'failing to accomplish' the task and making mistakes, some groups decided to split up their efforts in order to complete the problems, meaning that some students did not actually work on all parts of the assignment, resulting in a feeling that they did not really engage with the material and get the practice that they wanted. Students also expressed in a variety of different ways that they felt pressured for time when working on the assignment; many students felt that their group was ‘rushing through’ the problems, not allowing them to work at their preferred pace to build their understanding and confidence on their own terms, or even not finishing the assignment by the end of the class session, resulting in a lack of confidence that they can actually complete the problems.

Lastly, there were a small number of students who experienced a decrease in confidence apparently because they were unable to solve a problem on their own or with their group, requiring the assistance of their peers or the instructors (Figure 1.3). This particular experience is interesting because it relates to an important consideration for the development of self-efficacy beliefs, that students may interpret the same set of circumstances and make different judgements about their ability (Bandura 1997; Usher & Pajares, 2008). Where one student may find that being able to ask their peers or the instructor questions reflects an availability of help and leverages that support to feel more confident in themselves, other students appear to judge themselves negatively because they ‘required support’, perhaps feeling that a true mastery over the material means that they must be able to ‘accomplish it themselves’, and being able to do so is a failing on their part.

As with experiences which increased students’ self-efficacy, we wanted to determine whether students of different incoming self-efficacy were more or less likely to report a decrease in confidence from a lack of mastery. We examined a GLMM examining students’ pre-assignment self-efficacy level and their reporting of an experience which increased their self-efficacy which reflected a lack of mastery, controlling for the question topic (HWE or PG), gender, class section, and assigned group as fixed effects, with student as the only random effect. We checked the assumptions of this model (Theobald et al., 2019) and verified that: 1) the outcome variable is binary (reported a lack of mastery, yes or no); 2) our observations are not independent, but we are accounting for this lack of independence using our random effects; and 3) our predictor variable is linearly related to the logit of the outcome variable. We found that students' pre-assignment self-efficacy significantly decreased the log odds of reporting a mastery experience (β: -0.288; standard error 0.110, $p = 0.009$; Table 1.4). For a one-unit increase in self-efficacy level, the odds of a student not reporting a lack of mastery are 1.3 times greater than the odds of a student reporting a lack of mastery ($1/e^{\beta} = 1/e^{-0.288} = 1.334$). Therefore, greater self-efficacy levels are related to a lower probability of reporting a lack of mastery when making a judgement about a negative group work experience (Figure 1.2). This is contrary to our expectations that students with higher incoming self-efficacy will rely more on mastery sources than their lower self-efficacy peers when evaluating their own self-efficacy. This expectation is based on Butz and Usher's (2015) findings, where students of higher self-efficacy tended to evaluate their beliefs through mastery more than their lower self-efficacy peers, but as they only examined how their students' self-efficacy increased and not how they decreased, this tendency may not be present when considering what decreased self-efficacy. Furthermore, given that many lower self-efficacy students may not have had prior experience with HWE or PG, it is possible that the group work experiences they described were their first significant exposure to these problems, marking these moments of a lack of mastery as particularly significant in harming their self-efficacy beliefs. We also found that gender was a significant predictor (β: -0.986; standard error 0.278, $p < 0.001$; Table 1.4). The odds of a female student reporting a lack of mastery are 2.7 times greater than a male student ($1/e^{\beta} = 1/e^{-0.986} = 2.680$). This reflects the gender differences found by Butz and Usher (2015) with their students.

Lack of Availability of Help from Peers and Teachers, Vicarious Experiences, Social Persuasions, and Physiological States

The most common expression by students of a lack of help from their peers was that their group suffered a breakdown of communication or were simply unwilling to work together (Figure 1.3). Sadly, some students indicated that their peers not only neglected to speak up or contribute to the group's efforts, but also that some individuals actively discouraged discussion or withheld their answers from the group. This experience highlights one of the primary pitfalls of group work, when group dynamics break down and inhibit collaboration rather than foster it (Chang & Brickman, 2018; Donovan et al., 2018; Nokes-Malach, 2015).

Notably, in contrast to experiences which increased self-efficacy where we did not find vicarious experiences, social persuasions, or physiological states, students described a small number of experiences which reflected these sources of self-efficacy (Figure 1.3). Related to the sense of time pressure that other groups experienced, some students expressed that they felt like they were 'falling behind' their peers in their progress on the assignment, judging themselves less confident in their abilities because they felt unable to keep up with their more skilled group mates or even feeling like a burden on their group because they were taking longer to work through the problems. Some students also indicated that they felt less confident in their abilities because their peers were effortlessly proceeding through the assignment while they themselves struggled. A handful of students described how combative or overly critical peers made them question their abilities and decreased their self-efficacy. As with the lack of vicarious experiences and social persuasions in our increase responses, the relative influence of the sources of self-efficacy may explain why they appear in the set of decrease responses. These experiences appear to describe significant moments of self-doubt, without an immediate or subsequent resolution as we observed across many of the decrease responses. Therefore, these specific experiences might have stood out more to our students than their positive equivalents. We additionally observed a small number of students express anxiety or stress over working in the group, hinting at the general dislike of group work observed by Chang and Brickman (2018).

Teaching Implications

This study explored how group work may influence students' engagement with quantitative biology, but our results have important implications for understanding the impact of group work more broadly. Crucially, there appear to be several negative experiences which instructors should be mindful of when designing and supervising their group work. A key experience is when groups are unable to come to a consensus or are otherwise confused and unable to figure out a problem even after working together. This sometimes went hand-in-hand with a failure of groups to openly communicate with each other. Instructors can try to mitigate these experiences by reinforcing the importance of talking through problems as a group, or reaching out to neighboring groups if the whole group is struggling, leveraging the potential for 'super groups' to magnify the collaborative and self-guiding benefits of group work we observed. A strategy to enforce these internal discussions may be to assign roles to each student in a group. Group roles such as a group scribe or discussion leader (Bailey et al., 2012) can enhance students' engagement in information sharing (Mesmer-Magnus & DeChurch, 2009) and can ensure collaboration and equal participation (Savadori et al., 2001). Care must be taken, however, to ensure that students understand the specific role they are assigned to and that each role is sufficiently meaningful, as sometimes groups may diminish or ignore their roles, even resulting in experiences similar to our students where they merely worked independently or only on a discrete portion of the assignment, consolidating their work in the end (Chang & Brickman, 2018).

Additionally, students often expressed some feeling of time pressure when working in a group, desiring more time to work with the problems and build their own understanding, or feeling rushed by their group mates. Instructors can take care to design their assignments and problems to be completable well within the time allotted for a class session, break up a group assignment over multiple sessions, or provide additional practice for students outside of class, but this may not always be possible. Instead, to alleviate the feeling of being rushed or falling behind during class, group work assignments can be designed to balance providing sophisticated problems which promote student gains (Kirschner et al., 2011) while also reducing the stakes of the assignment by grading charitably or not at all, instead assessing students on the assignment content through other means, like group quizzing or polling questions throughout class (Hodges, 2018).

Encouragingly, there are also several positive experiences which arise from group work which instructors can foster. Instructors should lean heavily into the collaborative benefits of group work and provide ample opportunities for students to discuss their ideas and results throughout the group work assignment, as students predominantly found this aspect beneficial. Instructors can therefore design their assignments to include frequent checkpoints or opportunities for students to share and confirm their answers. For example, instructors can incorporate problems into a group work assignment which ask students to discuss among themselves and form a consensus before proceeding, or by segmenting a group work assignment for whole-class discussions (Gillies, 2013) or calling on students to lead or support discussions about the group work (Eddy et al., 2015). These structures can help provide students with validation of their efforts and verification of their success and abilities, increasing their self-efficacy. Additionally, several students expressed the advantage of being able to teach or guide their peers in reinforcing their own understanding and self-efficacy. Establishing group roles to help facilitate discussion and directed help (Bailey et al., 2012) may benefit these more confident students as well as their peers who need help. Lastly, and perhaps most simply, students most frequently reported that simply being able to work on these problems and achieve success was highly beneficial in increasing their self-efficacy. Instructors can design group work assignments to solve complex or more involved problems which require students to work together (Scager et al., 2016) and provide meaningful opportunities for students to succeed at a sophisticated task, potentially increasing the significance and endurance of that success and mastery in shaping their self-efficacy (Bandura, 1997; Usher & Pajares, 2008).

Limitations and Future Directions

When we set out to investigate what specific experiences students have during group work which affect their self-efficacy, we elected to survey students using short open responses and a basic assessment of self-efficacy. This approach meant that we could survey a large number of students, allowing us to capture and characterize a broad set of distinct group work experiences, both positive and negative, and how those experiences feed into students' self-efficacy beliefs. This breadth, however, carries with it the significant trade-off of precision.

Short open-response questions like those on our surveys are simple for students to complete but are likely unable to capture the full range of students' experiences of students in great detail. Our questions asked only generally about students' group work experiences which increased or decreased their confidence in their ability to solve the quantitative problems, which represents a huge range of possibilities. Students may not have thought particularly deeply or thoroughly about the experiences which were most meaningful to them; they may simply have reported the first group work experience that they could remember during the survey.

Recent research has shown that students may rely on several different sources of information or experiences to varying degrees when evaluating their self-efficacy beliefs (Chen & Usher, 2013). Our own findings, especially among experiences which decreased self-efficacy, revealed how many experiences appear related to each other, which is reflective of the complex social dynamics and interactions of group work (Nokes-Malach, 2015; Donovan et al., 2018) One distinct experience may contribute to the significance or salience of an entirely different experience on students' self-efficacy beliefs. For some students, perhaps all that was necessary to increase their self-efficacy was to simply work together in groups, leveraging the availability of help from peers, while for another student in their same group, merely talking with their peers may not be enough. Additionally, the differential presence and absence of sources across experiences which increased or decreased self-efficacy may be due to their relative importance in different contexts. In our students, we observed that vicarious experiences appear to play a role in harming their self-efficacy, but not in helping their self-efficacy. This variability in the magnitude or significance of how each source of self-efficacy influences students in different context or valences further underscores the limited depth of analyzing short responses and broader questions about what experiences shaped students' self-efficacy.

The highly-specific nature of experiences which influence students' self-efficacy beliefs may instead be more thoroughly captured through the use of interviews with students about their group-work experiences. Interviews offer researchers an opportunity to explore a rich account of students' individual experiences and self-efficacy judgements. Students could be asked to expand upon particularly salient experiences, revealing and making explicit the possible relationships between that experience and others during the group work. Additionally, while we were able to show a relationship between students' self-efficacy levels and their likelihood to report a given source of self-efficacy, more insight could be garnered through interviews about the specific characteristics of an individual student, such as their academic background, socioeconomic status, experiences relating to their race/ethnicity, and the myriad other biases which could further influence their self-efficacy beliefs, how they are formed, and how they interpret group work experiences to develop those beliefs. In particular, our sample lacks the demographic breadth found in other courses at other institutions (our students were predominantly female, for instance), further highlighting a need to capture the experiences of students from a variety of backgrounds and characteristics.

Conclusion

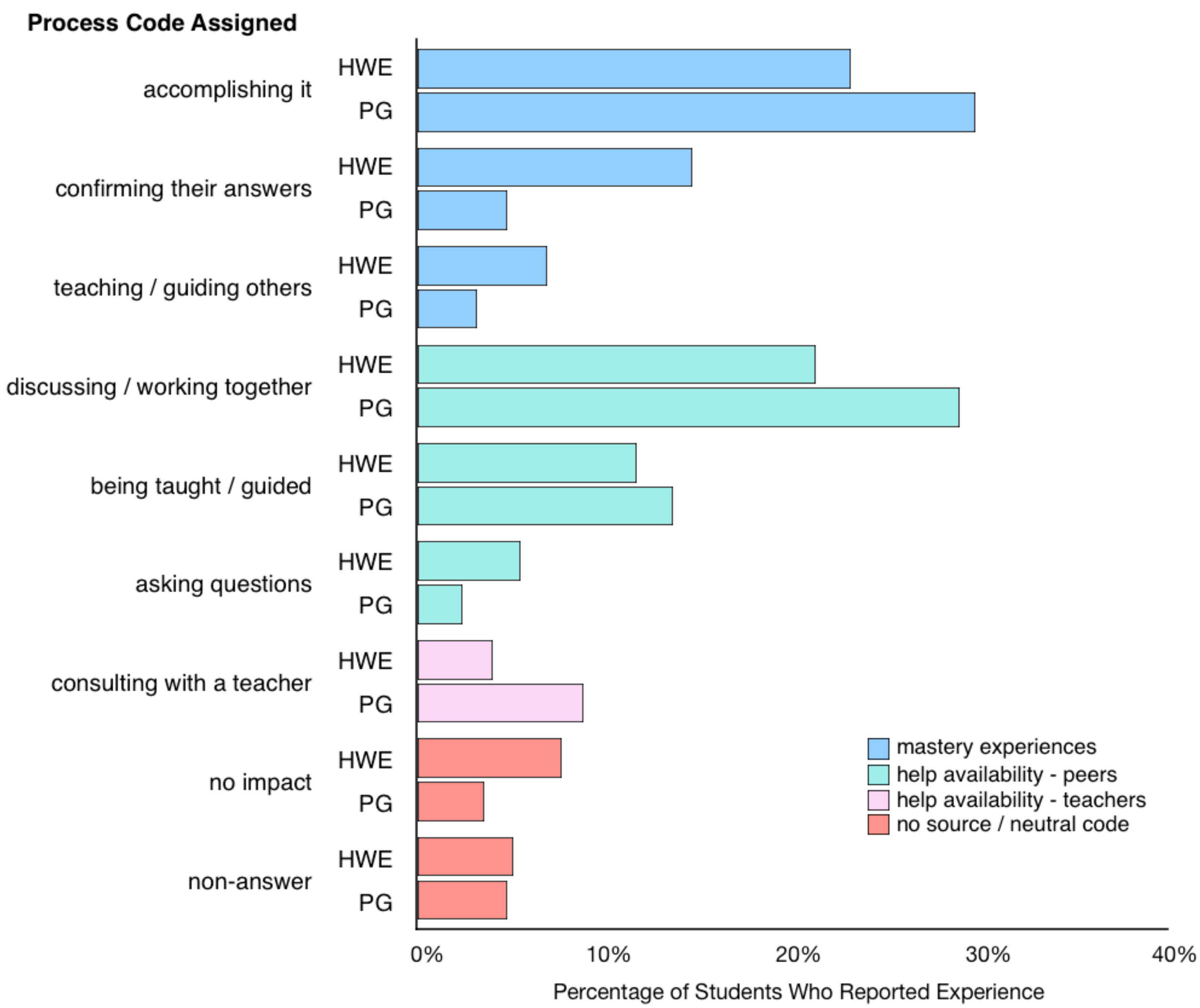
This study found that introductory biology students working in groups to complete quantitative biology tasks like calculating Hardy-Weinberg Equilibrium and modeling population growth draw their math self-efficacy judgements from a wide variety of experiences, providing a window into how group work especially may foster or harm students' self-efficacy. In particular, when their confidence increased, students reported a preponderance of constructive mastery experiences, finding that working on these problems in their groups provided them additional practice, the ability to verify their answers with their peers, and even opportunities to leverage their mastery to help their peers, reinforcing their own abilities. Students also found that the group work frequently allowed them to self-correct and guide each other through the problems, highlighting the collaborative, interactive, and supportive benefits of working in groups by making those experiences accessible to students. These group work benefits have been shown to increase students' engagement (Nokes-Malach et al., 2015), and may also help reshape students' avoidance of math in biology by providing them with opportunities to develop their quantitative skills in a supportive environment. When we examined how the prevalence of these sources related to students' math self-efficacy levels, we found that higher self-efficacy students were more likely to report mastery experiences when evaluating their self-efficacy, and lower self-efficacy students were more likely to report that the availability of help from their peers increased their self-efficacy, highlighting how experiences during group work and the sources of self-efficacy may be differently interpreted by students based on their existing self-efficacy beliefs. Instructors can design group work assignments and tasks to engage students with meaningful challenges and provide them opportunities to demonstrate their mastery, while also building in frequent discussion questions or checkpoints to reinforce and encourage groups to collaborate with each other and ensure that every member is contributing and understanding the problems.

Meanwhile, when considering how their confidence decreased, students reported a wide range of highly specific experiences in which they felt a lack of mastery, such as making mistakes or being rushed for time, a lack of support from their peers due to a breakdown in communication and collaboration, or experiences in which they compared themselves to their peers and were unable to keep up or fully participate with their group. In contrast to our earlier results, we found that lower self-efficacy students were likely to report a lack of mastery compared to their higher self-efficacy peers when judging experiences that decreased their self-efficacy. Instructors should take care to structure the group work to minimize the impact of momentary losses in group cohesion through being active in the classroom during group work and keeping a close eye on groups who appear to be struggling, intervening when necessary while still allowing groups to exercise their ability to self-correct.

While the responses we collected from students provide a window into the variety of experiences which influence their self-efficacy judgements, our data is limited in capturing the complexity of interactions and group dynamics in which these experiences arise. The breadth and specificity of the experiences we observed warrants a more thorough investigation into how they interact with each other, how students interpret these experiences, and how these interpretations and judgements ultimately affect students' self-efficacy beliefs. Our hope is that biology educators and instructors will gain better insight into how the instructional strategy of group work can impact their students' engagement with quantitative biology, and ultimately help them implement this strategy in their own classrooms.

Figures

Figure 1.1: Percentage of Students Who Reported A Given Experience Which Increased Self-Efficacy. Colors correspond to the source of self-efficacy that each code was categorized under (ref. Table 1.2). Only students who had both consented to the research and completed the surveys were included (HWE: n = 273; PG: n = 249). The *a priori* codes 'comparing themselves positively', 'getting positive feedback from peers', and 'getting positive feedback from teachers' did not appear in this set and were excluded from the figure. 'No impact' represents responses where students expressed that no experiences increased their self-efficacy. 'Non-answer' represents unintelligible, incomplete, or irrelevant responses.



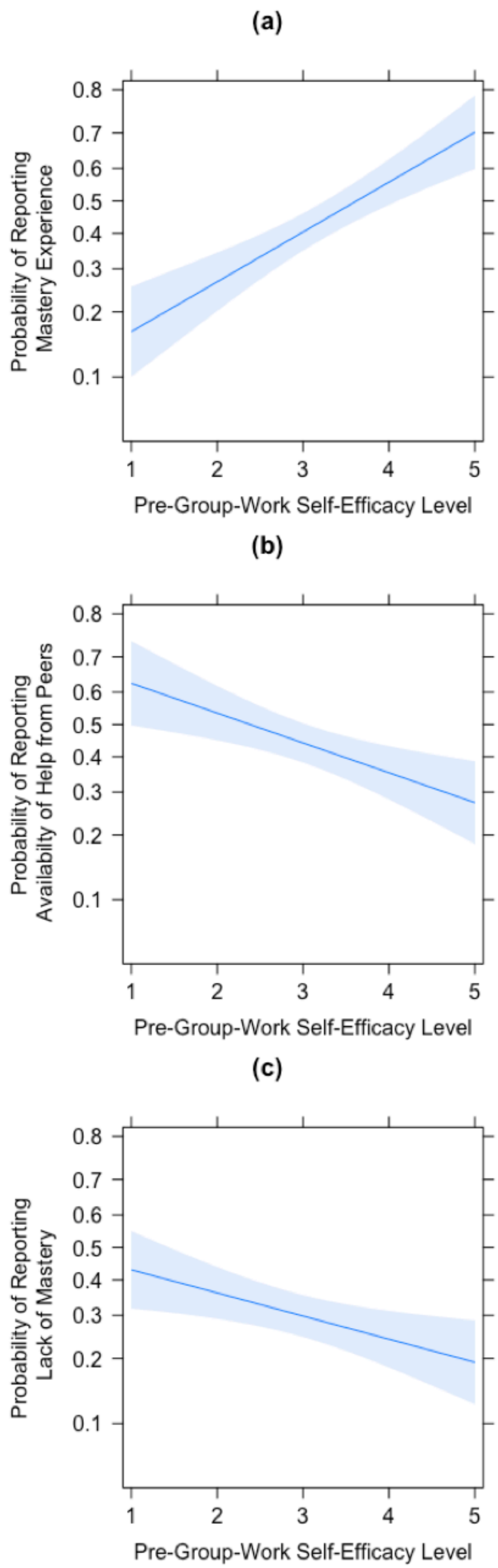
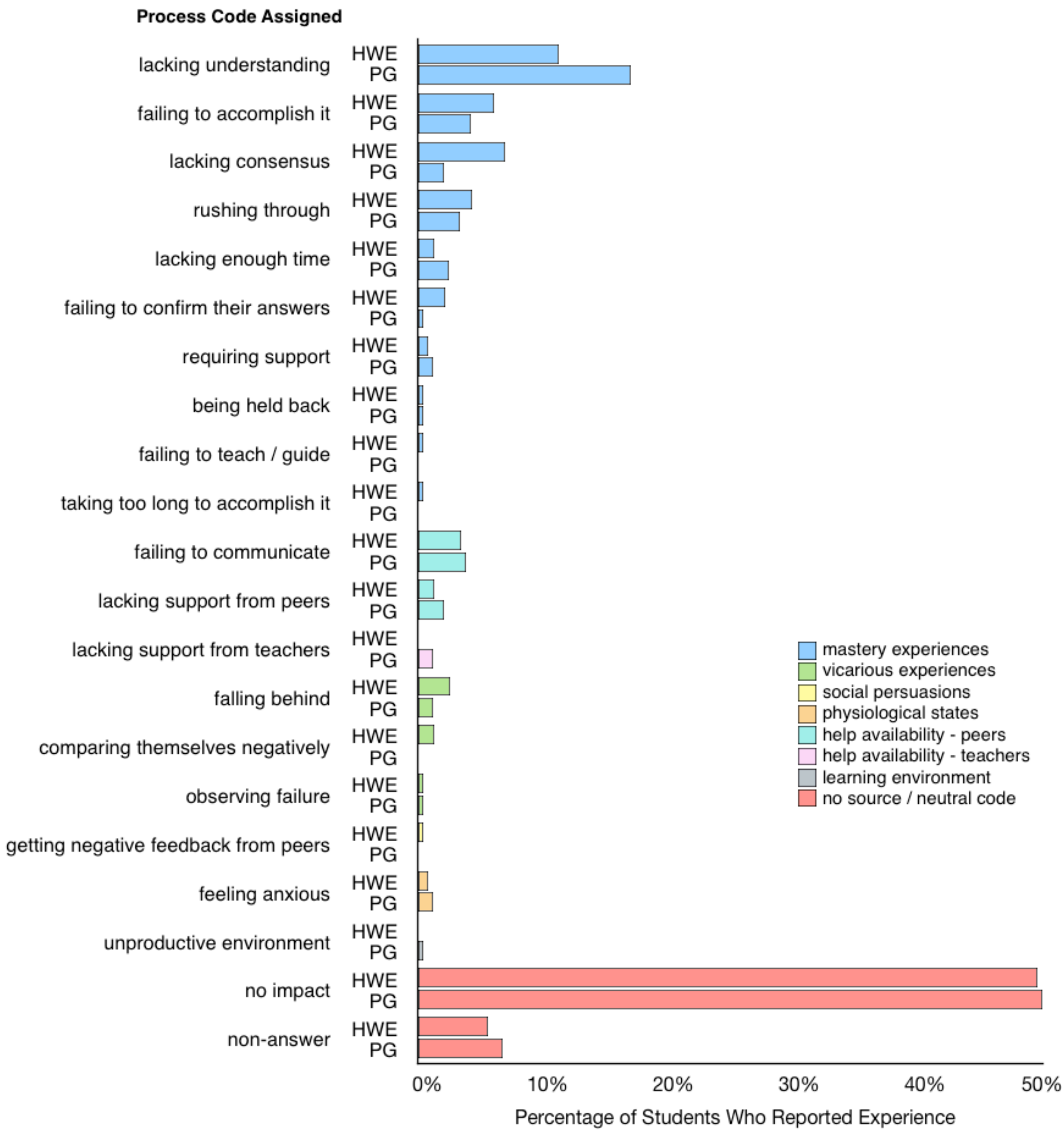


Figure 1.2: Relationship Between Students' Pre-Group-Work Self-Efficacy and their Probability of Reporting a Given Source of Self-Efficacy. All models controlled for students' gender, the type of problem (solving Hardy-Weinberg Equilibrium or modeling population growth), class section, and assigned group, including student as the only random effect. The shaded regions represent the 95% point-wise confidence interval of the estimated effect. **(a)** n = 460 responses. Students with higher self-efficacy prior to starting the group work were more likely to report a mastery experience which increased their self-efficacy than their lower self-efficacy peers (***β*: 0.622; standard error 1.126, *p* < 0.001**). **(b)** n = 460 responses. Students with higher self-efficacy prior to the group work were less likely to report that the availability of help from their peers increased their self-efficacy than their lower-self-efficacy peers (***β*: -0.374; standard error 0.118, *p* = 0.002**). We additionally found a significant effect for class section (***β*: -1.467; standard error 0.448, *p* < 0.001**). **(c)** n = 447 responses. Students with higher self-efficacy prior to the group work were less likely to report that a lack of mastery decreased their self-efficacy than their lower self-efficacy peers (***β*: -0.288; standard error 0.110, *p* = 0.009**). We additionally found a significant effect for gender. The odds that a female student reported a lack of mastery were 2.7 times greater than the odds of a male student to report a lack of mastery (***β*: -0.986; standard error 0.278, *p* < 0.001**).

Figure 1.3: Percentage of Students Who Reported A Given Experience Which Decreased Self-Efficacy. Colors correspond to the source of self-efficacy that each code was categorized under (ref. Table 1.3). Only students who had both consented to the research and completed the surveys were included (HWE: n = 230; PG: n = 235). The *a priori* code 'getting negative feedback from teachers' did not appear in this set and was not included in the figure. 'No impact' represents responses where students expressed that no experiences decreased their self-efficacy. 'Non-answer' represents unintelligible, incomplete, or irrelevant responses.



Tables

Table 1.1: Demographics of Study Participants. n = 311. Some characteristics may not have percentages which total 100%, due to excluding students who did not or preferred not to respond for a given characteristic.

Characteristic	Percentage of Participants
Gender	
Male	33%
Female	66%
Other	1%
Year in School	
First Year	71%
Second Year	14%
Third Year	10%
Fourth Year	2%
Other	1%
Highest Math Course Achieved	
Algebra or Geometry	11%
Trigonometry	7%
Pre-Calculus	46%
Calculus	35%

Table 1.2: Process Codes for Group Work Experiences Which Increased Self-Efficacy. Ten distinct experiences which increased self-efficacy were identified. Table contains the name of the code, the definition of the code, the percentage of the code across all reported experiences which increased self-efficacy (n = 515), and the source of self-efficacy which reflects the code.

Code	Definition	Percentage	Source
accomplishing it	The student achieves success/progress on a task through their own effort, or that they just practice as part of the group work session	45%	mastery experiences
confirming their answers	The student checks their own answers with other members of the group and feels more confident in their own	17%	mastery experiences
teaching / guiding others	The student's SE is impacted by teaching or guiding someone in their group themselves	9%	mastery experiences
discussing / working together	The student describes that the group discussed with each other or worked together to solve the problem	44%	help availability - peers
being taught / guided	The student's SE is impacted by someone else in their group or students around them teaching or guiding them through the problem	22%	help availability - peers
asking questions	The student seeks help by directly asking their group members	8%	help availability - peers
consulting with a teacher	The student's SE is impacted by consulting, asking, or otherwise seeking OR receiving help from an instructor or teaching assistant	12%	help availability - teachers
getting positive feedback from peers	The student's SE is impacted through some sort of encouragement from their group mates	0%	social persuasions
getting positive feedback from teachers	The student's SE is impacted through some sort of encouragement from their instructor or teaching assistants	0%	social persuasions
comparing themselves positively	The student compares their answers, methods, or abilities to those of their peers and judges themselves as better	0%	vicarious experiences

Table 1.3: Process Codes for Group Work Experiences Which Decreased Self-Efficacy. Twenty distinct experiences which decreased self-efficacy were identified. Table contains the name of the code, the definition of the code, the percentage of the code across all reported experiences which increased self-efficacy (n = 468), and the source of self-efficacy which reflects the code.

Code	Definition	Percentage	Source
lacking understanding	The student experiences doubt or confusion about the overall content or questions/assignment which undermines their SE	22%	mastery experiences
failing to accomplish it	The student's SE is impacted by failing at a task and explicitly indicates that they made a mistake / got something wrong, or that they simply didn't do the assignment	8%	mastery experiences
lacking consensus	The student experiences doubt about their answers or their method because other members of their group got different answers, used different methods, or questioned the answers/methods of the student	7%	mastery experiences
rushing through	The student's SE is impacted because they or their group worked at a faster pace than they wanted or could keep up with	5%	mastery experiences
lacking enough time	The student's SE is impacted because they felt pressed for time and/or did not finish the assignment	3%	mastery experiences
failing to confirm their answers	The student feels less confident in their answers because their group did not check their answers or was otherwise unable to check their answers	2%	mastery experiences
requiring support	The student's SE is impacted because they needed to refer to their notes, ask questions, or receive help from others in order to complete a problem or the assignment	2%	mastery experiences
being held back	The student's SE is impacted because they felt slowed down or impeded by their group	1%	mastery experiences
failing to teach / guide	The students' SE is impacted by having difficulty or being unable to explain their answers or methods to their group	<1%	mastery experiences
taking too long to accomplish it	The students' SE is impacted by feeling like they needed more time or effort to solve a problem than they wanted / thought they should take	<1%	mastery experiences
failing to communicate	The student describes that the group neglected to, was unable to communicate openly, or miscommunicated about/during the group work	6%	help availability - peers

lacking support from peers	The student's SE is impacted by being unable to seek help or guidance from their peers because their group members were unable to help them or were not present/unavailable	3%	help availability - peers
lacking support from teachers	The student's SE is impacted by being unable to seek help or guidance from a teacher or that no teachers were present, able, or willing to help them	1%	help availability - teachers
falling behind	The student's SE is impacted because their group members were ahead of them and they judged themselves negatively due to their lack of speed	3%	vicarious experiences
comparing themselves negatively	The student compares their answers, methods, or abilities to those of their peers and judges themselves as worse	1%	vicarious experiences
observing failure	The student's SE is impacted by observing others around them fail at a task or get something wrong, causing them to doubt their own ability on the task	1%	vicarious experiences
getting negative feedback from peers	The student's SE is impacted through some sort of discouragement from their group mates	1%	social persuasions
getting negative feedback from teachers	The student's SE is impacted through some sort of discouragement from their instructor or teaching assistants	0%	social persuasions
feeling anxious	The student's SE is impacted by experiences during class which produce anxiety, stress, or frustration	2%	physiological states
unproductive environment	The student's SE is impacted by the structural and environmental components of the group work assignment	<1%	learning environment

Table 1.4: Logistic Regression Outputs for Students' Self-Efficacy and Reporting of a Source of Self-Efficacy. n = 447 for experiences which increased self-efficacy; n = 460 for experiences which decreased self-efficacy. Regressions were conducted on student responses from both Hardy-Weinberg Equilibrium (HWE) and population growth (PG) group work sessions. Asterisks indicate significant predictors ($p < 0.05$)

Source of Self-Efficacy	Unstandardized Coefficients			Odds ratio
	β	S.E.	p-value	e^{β}
Mastery Experiences - Increase				
Intercept	-2.584	0.508	0.000	
Self-Efficacy Level	0.622	0.116	0.000*	1.862
Gender - Male	0.266	0.241	0.271	1.305
Question Type - PG	-0.098	0.221	0.658	0.907
Class Section - 2	0.003	0.382	0.993	1.003
Assigned Group	0.006	0.007	0.390	1.006
Help Availability from Peers - Increase				
Intercept	0.899	0.499	0.072	
Self-Efficacy Level	-0.374	0.118	0.002*	0.688
Gender - Male	-0.516	0.274	0.060	0.597
Question Type - PG	0.346	0.235	0.141	1.413
Class Section - 2	-1.467	0.448	0.000*	0.231
Assigned Group	0.011	0.008	0.170	1.011
Mastery Experiences - Decrease				
Intercept	-0.058	0.456	0.899	
Self-Efficacy Level	-0.288	0.110	0.009*	0.749
Gender - Male	-0.986	0.278	0.000*	0.373
Question Type - PG	-0.395	0.235	0.093	0.674
Class Section - 2	0.541	0.383	0.157	1.718
Assigned Group	0.007	0.007	0.293	1.007

CHAPTER 2 — INVESTIGATING A MULTI-DIMENSIONAL MODEL OF STUDENTS' VALUES TOWARDS STATISTICS

Introduction

One of the most important quantitative topics in biology education is statistics (NRC, 2003; AAAS, 2011). Many new biological tools and techniques depend on a strong foundation of statistical understanding, and a good statistics education is fundamental to scientific literacy (Ben-Zvi & Garfield, 2004). In order to build such a foundation, students must be able to grasp key concepts and ideas but also connect between them, necessitating grounding of these concepts in authentic biological examples and applications (Ben-Zvi & Garfield, 2004; Colon-Berlingeri & Burrowes, 2011). Authentic examples can be provided through statistical problem-based learning curricula (Karpiak, 2011) or interactive workshops in lab-based settings (Olimpo et al., 2018). Many approaches have attempted to better address this need for statistical education in biology by integrating statistical concepts such as probability (Liu & Zhu, 2016) into biology courses (Metz, 2008; Colon-Berlingeri & Burrowes, 2011), while others have approached integration from the other direction, bringing more biological contexts into statistics courses (Masel et al., 2015). While both approaches show promising results in building the statistical foundation for students in biology, one significant hurdle facing instructors is that students often still hold crucial misconceptions about statistics (Castro Sotos et al., 2007) and scientific principles in general (Gormally et al., 2012). Even after taking a course in introductory statistics, students struggle to evaluate probabilities, interpret visualizations such as box plots, and draw statistical conclusions (Delmas et al., 2007). Additionally, students may also have difficulty interpreting statistical language and symbology, preventing them from building a deeper understanding of fundamental concepts such as sampling distributions (Kim et al., 2016).

A key factor in the development of these misconceptions may lie with students' attitudes towards statistics. Undergraduate students tend to hold ambivalent or negative attitudes towards statistics (Gal & Ginsburg, 1994). In particular, students are often uninterested in statistics (Gal & Ginsburg, 1994) or fail to find it useful (Evans, 2007). Students also frequently exhibit anxiety and low confidence in their ability towards math (Chang & Beilock, 2016) and statistics (McKim, 2014), which can be exacerbated upon students' first contact with statistics in an authentic context (Ruggeri et al., 2011). These negative attitudes can decrease students' engagement and thus performance, inhibit learning of statistical concepts, and create uncertainty in how to apply those concepts to real-world situations (Gal & Ginsburg, 1994; García-Santillán et al., 2013; Garfield & Ben-Zvi, 2007).

Theoretical Framework

One of the most widely-used frameworks to explore students' attitudes is expectancy-value theory (EVT; Eccles et al., 1983), which argues that students' performance on a task and their achievement is impacted by their expectancies of success on the task as well as the values they hold towards the task, or "task-values"(Wigfield & Eccles, 2000). Students have four distinct task-values, hereby referred to as the 'canonical task-values': intrinsic value, attainment value, utility value, and cost (Wigfield & Eccles, 2000). Intrinsic value is how much a student enjoys performing a given task or their interest in the task. Attainment value is how important it is to the student to perform well on the task. Utility value is how useful the task is to the student towards achieving their goals. Cost is characterized by the negative effects the student perceives they will incur as a result of performing the task or in order to succeed at the task (Eccles et al., 1983; Wigfield & Eccles, 2000).

While the four primary task-values have each largely been examined as monolithic, singular constructs, Eccles (1983) and Wigfield (2000; 2002) have long suggested that certain primary task-values may instead encapsulate multiple, more nuanced dimensions of the overall construct. For example, when a student considers how costly a task may be, they may consider specific and distinct aspects of cost, such as the emotional burden of engaging with a task, how much effort they believe they need to expend to engage in the task, or consider how engaging with the task trades off with their other goals, representing an opportunity cost (Eccles et al., 1983; Wigfield & Eccles, 2000; Eccles & Wigfield 2002). Much recent work has focused around evaluating these emerging dimensions of task-values, showing that in particular separating out these dimensions for cost can more effectively capture student motivation and outcomes than lumping them together (Conley, 2012; Perez et al., 2014, Flake et al., 2015). Utility value may also consist of several dimensions, each capturing a different aspect of what it means for a task to be useful for students; for example, students can differentiate between different domains of goals in their lives, such as their academic goals versus those for their career or even their daily lives (Peetsma & van der Veen, 2011).

A promising approach to accounting for this extra dimensionality may be a model which differentiates students' task-values even further than just the four canonical task-values. Gaspard et al. (2015) investigated and established a model dividing the four canonical task-values into eleven task-value facets. These facets represent specific aspects or dimensions within each canonical task-value, such as the various life-domains for utility value, or the specific kinds of cost a student might perceive. Specifically, they performed confirmatory factor analyses of the canonical task-values and found evidence in secondary-school math students for two facets within attainment value, five facets within utility value, and three facets within cost, while intrinsic value had no further dimensionality than what was previously described (Figure 2.1). Furthermore, they found that students' task-values were better represented through differentiating the eleven facets than simply relying on the four canonical task-values. Empirical evidence suggests that these facets may be distinguishable in different domains and subjects than just mathematics (Nagengast et al., 2013), such as English (Trautwein et al., 2012), and various natural sciences like physics, chemistry, and biology (Guo et al., 2017). This highlights the need to examine the validity of such a model in a specific biology educational context. Here, we examine the validity of this multi-dimensional model for measuring task values in the context of statistics in biology at the undergraduate level.

Research Goals

Our study had two primary goals. First, we sought to extend the model of task-value facets established by Gaspard et al. (2015) to a new context (statistics) and population (undergraduate life-sciences students). We asked whether such a model with multiple facet constructs more precisely represents students' task-values than the canonical model of four task-values as single constructs. Given the empirical evidence for task-value facets in other contexts and populations, we hypothesize that undergraduate life-sciences students' task-values towards statistics will be better described using a model of multiple task-value facets rather than a model of the canonical task-values as single constructs. We predict that students' attainment value, utility value, and cost will consist of multiple facets, consistent with the model established by Gaspard et al. (2015). Second, we sought to describe the relationship, if any, between students' task-values towards statistics and their understanding of statistical concepts in an applied context. We asked, after controlling for students' overall academic achievement as measured by GPA and their prior exposure to statistics, how students' intrinsic value, utility value, and cost relate to their performance on a statistical assessment, and we hypothesize that students' intrinsic value and utility value will relate positively to students' performance, while students' cost will relate negatively to students' performance. We predict that students with higher intrinsic value towards statistics will perform better on the assessment than their peers with lower intrinsic value towards statistics. We also predict that students with higher statistics utility for school, daily life, or career will perform better on the assessment than their peers with lower utility value facets towards statistics. Lastly, we predict that students who express lower effort required and emotional cost for statistics will perform better on the assessment than their peers who perceive high cost towards statistics.

Methods

Participants and Setting

We surveyed 366 undergraduate life-sciences students across two public research universities: a large Northeastern university (n = 286), and a large Western university (n = 80). These students were drawn from two introductory-level statistics courses. At their respective institutions, some life-science majors list the course as an explicit requirement, while other life-science majors list it as an option towards an overall 'math' component of the major. At the large Northeastern university, the introductory statistics course is conducted as a single section per semester of roughly 150 students in a large-lecture format, covering topics such as: probability distributions, distributions of sample statistics, regression and correlation, and analysis of variance. Students frequently engaged with statistical examples drawn from actual biological data and endeavored to produce a final project developing and testing a statistical hypothesis using a provided data set. We surveyed this course during the Spring of 2019 and the Fall of 2020. At the large Western university, the introductory statistics course is conducted as two sections of roughly 40 students during the Spring semester, covering similar topics to the course at the Northeastern university. This course had a specific focus on working with data using R, and frequently asked students to write short metacognitive reflections about their learning as a component of the course. We surveyed this course during the Spring of 2020, and note that due to the COVID-19 pandemic, this institution had transitioned to an online instructional format partway through the semester; our survey was administered after this transition. Participants' demographic information is summarized in Table 2.1. This study was approved by the Institutional Review Board at the University of New Hampshire, IRB #8077, and an IRB approval waiver was received at the outside institution.

Measures

We administered two separate surveys to each section of students: an attitude survey, asking students to report their task-values towards statistics, and a knowledge assessment of statistical concepts in an applied biology context. We distributed these surveys to participants through each course's respective online course management system as a class assignment spread out over two consecutive weeks, one survey per week. Students received course credit equivalent to one homework assignment for completing both surveys, awarded in two parts, one for each survey. We used the online service Qualtrics to design and administer the surveys during the latter third of each semester, to ensure that students had sufficiently covered the material asked on the surveys. From the Northeastern university, 232 students responded to the surveys, and 52 students responded from the Western university.

Students' Task-Values Towards Statistics

To quantify students' task-values towards statistics, we adapted a task-value instrument developed by Gaspard et al. (2015). This instrument consists of 37 individual survey items grouped into 11 task-value facets, each of which falls under one of the four canonical task-values (ref. Table 2.2, Appendix B). Each task-value facet on the survey contains between two and six items which relate to the facet. Each item asked students to rate their agreement with a statement about the respective task-value facet using a seven point scale, ranging from 1 - "Strongly Disagree" to 7 - "Strongly Agree". This instrument is functionally identical to that which Gaspard et al. (2015) developed, save for replacing all instances of the word 'math' with 'statistics', keeping all other verbiage and survey structure intact.

Students' Understanding of Statistical Concepts in an Applied Biology Context

To evaluate students' understanding of statistical concepts in an applied context, we used the Biological Variation in Experimental Design and Analysis instrument (BioVEDA, Hicks et al., 2020). This instrument consists of multiple-choice questions relating to sources of variation in biological experiments, how to control variation when designing an experiment, and how such variation impacts the results of statistical inferences based on these experiments. We selected this instrument because of the central significance of variation in statistical analysis, which represents a key statistical concept that students should be able to tackle as a result of their statistics training (Finney & Schraw, 2003; Horton & Hardin, 2015). For our study, we administered an early version of the instrument which had 20 items; however, subsequent validation of the instrument revealed only 16 items (Hicks et al., 2020). Therefore, we examined only those 16 validated items when conducting our analyses.

Data Analyses

Confirmatory Factor Analysis for Task-Value Models

We conducted several Confirmatory Factor Analyses (CFA) aimed at verifying whether or not the model of task-value facets identified by Gaspard et al. (2015) more precisely represents students' values towards statistics over simply looking at the four canonical task-values. All task-value CFAs were conducted using the responses from all 284 of our sampled students. We started by first exploring each task-value and its facets, specifying single-factor models for each canonical task-value, and multi-factor models where each task-value facet was its own factor. Intrinsic value has no hypothesized dimensionality; therefore, we tested a model of only one factor. For attainment value, we specified a single-factor model for the canonical task-value, and a two-factor model for its facets: 'importance of achievement' and personal importance'. For utility value, we specified a single factor-model for the canonical task-value, and a five-factor model for its facets: 'utility for school', 'utility for daily life', 'social utility', 'utility for career/job', and 'utility for future life'. For cost, we specified a single-factor model for the canonical task-value, and a three-factor model for its facets: 'effort required', 'emotional cost', and 'opportunity cost'. Following this, we explored and compared two combined models: a 3-factor model with each canonical task-value as a distinct factor, and a 9-factor model with each task-value facet as a distinct factor (refer to Table 2.2 for all facets and items). The initial exploration of the factor structure for attainment value proved inconclusive; therefore, we elected to exclude attainment value and its facets from these combined models and our future analyses, as the regression models are contingent upon a clear factor structure.

Because of the large number of scale degrees (7) and the ordinal nature of our surveys, we selected the robust maximum-likelihood estimator (MLR) as our estimator of variances in the data (Knekta et al., 2019). To evaluate the fit of our models, we relied on multiple fit indices: 1) the chi-square value from the robust MLR (MLR χ^2); 2) the comparative fit index (CFI); 3) the Tucker-Lewis index (TLI); 4) the root-mean-square error of approximation (RMSEA); and 5) the standardized root-mean-square residual (SRMR). We established thresholds (Table 2.3) for each fit index per the recommendations of Hu & Bentler (1999). All factor analyses were conducted in R v. 4.0.2 (R Core Team, 2020) using the packages 'lavaan' (Rosseel, 2012) and 'psych' (Revelle, 2019). Additionally, when evaluating the factor loadings for each factor and its items, which indicate how much of the variance in the responses for each item are related to the factor versus the error variance unique to each item, we declared factor loadings greater than 0.7 as 'high', per Knekta et al. (2019).

Regression Analyses for Relationship Between Task-Values and Performance on BioVEDA

To investigate the relationship between students' task-values and their performance on BioVEDA, we conducted multiple linear regression, which allowed us to control for a variety of extraneous factors which may affect that relationship (Theobald & Freeman, 2014). We used the model of students' task-value facets we identified from Goal 1, where students' task-values were represented as a single intrinsic construct, multiple utility value facets, and multiple cost facets, to inform the specification of our regression models. We decided to relate each set of task-value facets separately rather than in a single comprehensive model, to better understand the specific relationship between the facets of a given task-value and performance on the statistical assessment. We decided to relate only a subset of the five utility value facets and the three cost facets to students' performance. Some facets, such as students' social utility or utility for future life, had a small number of survey items which we believed were not as relevant to our undergraduate student population as they engage with statistics. Additionally, the items representing students' opportunity cost were similarly indistinct. Thus, we examined only the utility for daily life, utility for school, and utility for career/job facets within utility value, and only the effort required and emotional cost facets within cost.

For all our regression models, we started by including as predictors: 1) students' self-reported GPA, which serves as a reasonable proxy for their academic achievement despite their tendency to misreport it (Wright et al., 2009), which could impact their score on the assessment, and 2) students' prior exposure to statistics, as indicated by whether or not they completed a statistics course prior to taking their current statistics course, which could also impact students' understanding of variation. We also initially included students' institutions/schools as a random effect, but due to the small number of levels within this factor, this model failed to converge on a solution, so we instead included school as an additional fixed effect (Theobald, 2018). Our explanatory variables in each model were students' task-value facets, calculated as a mean across all items for each facet. In the case of intrinsic value which has no facets, we instead simply included the mean across all intrinsic value items as our explanatory variable in that model. Our outcome variable was students' score on the BioVEDA assessment, calculate as a sum score with a maximum of 16, which represents students' understanding of variation in experimental design and analysis. After excluding students who did not report a GPA or did not complete BioVEDA, and including only students from Fall 2019 and Spring 2020 due to minor modifications to our instruments made after Spring 2019, our sample for our regressions was 101 students across both institutions (Northeastern University, n = 73; Western University, n = 28). We conducted all our analyses using the standard R v. 4.0.2 packages for linear regression (R Core Team, 2020), 'lme4' (Bates et al., 2015), and 'effects' (Fox & Weisberg, 2019). Equations for our regression models are described below.

Intrinsic Value

BioVEDA Score ~ Intrinsic Value + GPA + Prior Stats Exp. + School

Utility Value Facets

BioVEDA Score ~ Utility for Daily Life + Utility for School + Utility for Career/Job + GPA + Prior Stats Exp. + School

Cost Facets

BioVEDA Score ~ Effort Required + Emotional Cost + GPA + Prior Stats Exp. + School

Results

Goal 1: Exploring Students' Task-Values Towards Statistics

Single-Factor Model for Interest Value

Examining a single-factor model for intrinsic value revealed a good model fit (Table 2.4). The chi-squared test of model fit was insignificant (MLR $\chi^2 = 0.476$, df = 2, p = 0.788), and both CFI and TLI were firmly above the threshold for good fit (CFI = 1.000, TLI = 1.000). RMSEA and SRMR were similarly indicative of good fit (RMSEA = 0.005; SRMR = 0.005). Our Cronbach's α for this model was 0.94. As there is no dimensionality within intrinsic value (Table 2.2), we maintain that with our students, the items measuring intrinsic value indeed load onto a single construct as previously described.

Single-Factor Model for Attainment Value Versus a Multi-Factor Model for Attainment Facets

The single-factor model for attainment value indicated poor model fit (Table 2.4) by all measures (MLR $\chi^2 = 244$, df = 35, p < 0.000; CFI = 0.810; TLI = 0.756; RMSEA = 0.175; SRMR = 0.089). In comparison, the two-factor model for the attainment facets of importance of achievement and personal importance indicated a better model fit than attainment value as a single construct (Table 2.4), yet still yielded poor model fit across nearly all measures (MLR χ^2 = 177, df = 34, p < 0.000; CFI = 0.874; TLI = 0.833; RMSEA = 0.145; SRMR = 0.076). We examined the factor loadings for both models to better understand their misspecification (Table 2.2). Factor loadings for the single-factor model were mostly within the range of 0.71 - 0.78, with substantially lower loadings for two items in importance of achievement and one in personal importance, indicating that the model does not explain these three items well (Knekta et al., 2019). Factor loadings for the two-factor model were moderately higher (mostly within 0.75 - 0.88), but the same items in importance of achievement and personal importance were still much lower. Given that most items loaded weakly and were poorly explained by the model, and that both the single-factor and two-factor models for attainment value appear to be misspecified, we decided to exclude attainment value and its facets from further analyses.

Single-Factor Model for Utility Value Versus a Multi-Factor Model for Utility Value Facets

The single-factor model for utility value also indicated poor model fit (Table 2.4) by all measures (MLR χ^2 = 662, df = 54, p < 0.000; CFI = 0.616; TLI = 0.531; RMSEA = 0.226; SRMR = 0.125). Factor loadings (Table 2.3) for the single-factor model of utility value were virtually all below 0.6, with only the factor loadings for utility - daily life items above 0.8. In contrast, the five-factor model for the utility value facets ‘utility for school’, utility for daily life’, ‘social utility’, ‘utility for career/job’, and ‘utility for future life’ fared better, indicating a good model fit across most measures (MLR χ^2 = 90, df = 44, p < 0.000; CFI = 0.972; TLI = 0.958; RMSEA = 0.068; SRMR = 0.050). Cronbach’s α for this multi-factor model were 0.82 for utility for school, 0.91 for utility for daily life, 0.84 for social utility, 0.64 for utility for career/job, and 0.87 for utility for future life. Factor loadings for the five-factor model of utility value facets (Table 2.2) were all within the 0.8-0.9 range save for three items, one in utility - school (0.7), social utility (0.67) and utility - career/job (0.525). These specific items showed a moderate correlation to other task-value facets both within utility and different task-values (see Appendix C), which could explain their poorer factor loadings. Despite this, given that the model fit for utility value facets was good compared to a model of only utility value, we argue that our students were indeed differentiating between the five hypothesized facets of utility value.

Single-Factor Model for Cost Versus a Multi-Factor Model for Cost Facets

The single-factor model for cost indicated poor model fit (Table 2.4) (MLR $\chi^2 = 556$, df = 44, $p < 0.000$; CFI = 0.728; TLI = 0.667; RMSEA = 0.266; SRMR = 0.132). Factor loadings (Table 2.2) for the single-factor model of cost were all below 0.6 for items in emotional cost and opportunity cost, but were very high in effort required (0.90 - 0.95). In contrast, the three-factor model for the cost facets ‘effort required’, ‘emotional cost’, and ‘opportunity cost’ indicated a good model fit across most measures (MLR $\chi^2 = 74$, df = 41, $p < 0.000$; CFI = 0.985; TLI = 0.980; RMSEA = 0.065; SRMR = 0.029). Cronbach’s α for this model were 0.97 for effort required, 0.89 for emotional cost, and 0.94 for opportunity cost. Factor loadings for the three-factor model were all consistently high (0.81 - 0.96), indicating that the cost items were well explained by a model which distinguished the three cost facets. Thus, we argue that for our students, cost is better represented through these three facets than as a single construct.

Three-Factor Model of Canonical Task-Values as Single Constructs

After removing attainment value and its facets from our model of canonical task-values, we additionally examined the items for the remaining three canonical task-values as part of a ‘combined model’ with three factors, one for each task-value. The results from this CFA indicated a poor model fit (Table 2.4). The chi-squared test of model fit was significant (MLR $\chi^2 = 1739$, df = 321, $p < 0.000$), and the CFI and TLI both indicated poor fit (CFI = 0.734; TLI =

0.709). Furthermore, both the RMSEA and SRMR also indicated poor fit (RMSEA = 0.140; SRMR = 0.103). Factor loadings (Table 2.3) for this model were mixed: while cost- effort required and some items from intrinsic value were high (> 0.9), most values were below 0.7, especially in the remaining cost facets and the utility value facets, indicating that this model poorly explained these items. Thus, we argue that our students' task-values are not well-described using only the four canonical task-values.

Nine-Factor Model of Task-Value Facets as Distinguishable Constructs

Lastly, we examined the nine remaining task-value facets (after excluding the two facets in attainment), and the results from this CFA indicated a noticeably better fit across all measures MLR $\chi^2 = 488$, df = 288, $p < 0.000$; CFI = 0.973; TLI = 0.967; RMSEA = 0.047; SRMR = 0.048). Factor loadings (Table 2.3) for this model were mostly high (> 0.8) indicating that this model well-describes most of the task-value items, although some items in intrinsic value, utility for school, and emotional cost were moderately lower (0.7-0.79). Only two items in utility for career/job and social utility were considerably low (0.51 and 0.67, respectively). Despite these two items, in light of the good model fit and compared to the results of the three-factor model, we argue that our students' task-values as a whole are better described using these nine task-value facets. Therefore, we decided to use this model of students' task-value facets when relating them to students' statistical understanding for the second goal of this study.

Goal 2: Regression Analyses Relating Students' Task-Values to BioVEDA Scores

Intrinsic Value and Students' Assessment Scores

We performed a multiple linear regression examining students' mean intrinsic value and their scores on the BioVEDA assessment, controlling for GPA, prior statistics course, and school (Table 2.5). We checked the assumptions of linear regression for this model and found: 1) the data contained no outliers (standard residual minimum = -6.02, standard residual maximum = 5.35); 2) low multicollinearity, with VIF values ranging from 1.02 to 1.15 (O'Brien, 2007); 3) the Q-Q plot of standardized residuals indicated approximately normally-distributed errors; 4) little to no heteroscedasticity as indicated by a plot of residuals vs. fitted values. This regression model was not statistically significant ($F(4, 96) = 2.189$, $p = 0.076$), and accounted for less than 5% of the variance in BioVEDA scores (Adj. $R^2 = 0.045$). Thus, we argue that students' intrinsic value does not predict their performance on the BioVEDA assessment.

Utility Value Facets and Students' Assessment Scores

We performed a multiple linear regression examining students' mean utility for daily life, utility for school, and utility for career and their scores on the BioVEDA assessment, controlling for GPA, prior statistics course, and school (Table 2.5). We checked the assumptions of linear regression for this model and found: 1) the data contained no outliers (standard residual minimum = -6.00, standard residual maximum = 6.26); 2) low multicollinearity, with VIF values ranging from 1.09 to 1.95; 3) the Q-Q plot of standardized residuals indicated approximately normally-distributed errors; 4) little to no heteroscedasticity as indicated by a plot of residuals vs. fitted values. This regression model was statistically significant ($F(6, 94) = 2.657$, $p = 0.020$), and accounted for roughly 9% of the variance in BioVEDA scores (Adj. $R^2 = 0.090$). We found two significant predictors: utility for school (β: 0.568; standard error: 0.267; $p = 0.036$) and self-reported GPA (β: 1.359; standard error 0.572; $p = 0.020$). Therefore, we argue that students' utility for school does predict their performance on the BioVEDA assessment: students with higher utility for school perform better on BioVEDA (Figure 2.2a).

Cost Facets Versus Students' Assessment Scores

We performed a multiple linear regression comparing students' mean effort required and emotional cost to their scores on the BioVEDA assessment, controlling for GPA, prior statistics course, and school (Table 2.5). We checked the assumptions of linear regression for this model and found: 1) the data contained no outliers (standard residual minimum = -5.37, standard residual maximum = 4.98); 2) low multicollinearity, with VIF values ranging from 1.12 to 2.43; 3) the Q-Q plot of standardized residuals indicated approximately normally-distributed errors; 4) little to no heteroscedasticity as indicated by a plot of residuals vs. fitted values. This regression model was statistically significant ($F(5, 95) = 4.307$, $p = 0.001$), and accounted for slightly over 14% of the variance in BioVEDA scores (Adj. $R^2 = 0.142$). We found a single significant predictor: emotional cost (β: -0.985; standard error: 0.283; $p < 0.001$). Thus, we argue that students' emotional cost does predict their performance on the BioVEDA assessment: students with lower emotional cost perform better on BioVEDA (Figure 2.2b).

Discussion

Dimensionality Within Students' Task-Values Towards Statistics

For our first research goal, we sought to understand whether our students' task-values could be better represented by distinguishing between specific task-value facets, or by simply using the four canonical task-values. The results from our factor analyses confirmed multi-dimensional models which differentiate between task-value facets. Utility value and cost were both better represented using a task-value facet model than treating them as singular constructs, confirming and extending the findings of Gaspard et al. (2015). With respect to utility value, students differentiate between different 'life-domains' such as school, career, or everyday life

(Peetsma & van der Veen, 2011), and that the various items used by other surveys to investigate utility value in general, such as those of Conley (2012) and Luttrell et al. (2010), are distinguishable to students as separate facets of their overall utility value. One thing to note, however, was the low internal reliability of the facet ‘utility for career/job’, as indicated by a Cronbach’s α of 0.64 and limited items for this facet. Closer inspection of these two items (Table 2.2) suggests that perhaps the item “Good grades in statistics can be of great value to me later on” may have been interpreted by students as relating to their academic goals instead; there is in fact a moderate correlation between ‘utility for school’ and ‘utility for career/job’ (Appendix C). Nevertheless, we argue that using these utility-value facets is a meaningful way to more thoroughly characterize students’ utility value towards statistics.

With respect to cost, students indeed differentiate between the three originally-hypothesized dimensions: ’opportunity cost’, ‘emotional cost’, and ‘effort required [to succeed]’ (Eccles et al., 1983; Wigfield & Eccles, 2000). Our findings corroborate those of Perez et al. (2014), who assessed cost using those three dimensions or ‘sub-factors’ (referring to ‘emotional cost’ as ‘psychological cost’) and were able to separate them as distinct through an exploratory factor analysis. We also corroborate the investigation by Flake et al. (2015), and although they argued for a model of cost which includes an additional fourth dimension—“outside effort cost”— that we did not include in our model, we share the conclusion that cost can be described with more nuance than other previous approaches, such as “task effort cost” (effort required), “emotional cost”, and a “loss of valued alternatives” (opportunity cost) (Flake et al., 2015; Conley, 2012; Trautwein et al., 2012).

However, our students did not appear to distinguish facets within attainment value. Neither a two-factor model of ‘importance of achievement’ and ‘personal importance’ or a single-factor model of attainment value achieved a good fit for our students' responses. This is in contrast to the findings of Gaspard et al. (2015) where they found that students differentiated between the two facets of attainment value. It is possible that the misspecification of these models is due to underlying interactions or correlations. While the factor loadings for both the single-factor model and two-factor model were moderate overall, there were three attainment value items with very low factor loadings (Table 2.2) suggesting that the poor model fit stems largely from these items. Looking more closely, the poorly-loading item from 'attainment - personal importance', "Statistics is not meaningful to me", could have be interpreted as relating to intrinsic value based on the wording. When examining the correlation coefficients of task-value facets within our students, we indeed observed moderate correlations between 'personal importance' and intrinsic value, as well as 'utility for daily life', suggesting that this item may have been captured by intrinsic value or utility value. Additionally, the 'attainment - importance of achievement' items "Performing well in statistics is important to me" and "Good grades in statistics are very important for me" may have been interpreted similarly to items relating to utility facets like 'utility for school' or 'utility for career/job'. We noted moderate to high correlations between 'attainment - importance of achievement' and 'utility for career/job', suggesting that like with 'personal importance', these items were instead captured by other task-value facets.

These observations are similar to those made by Gaspard et al. (2015) with respect to attainment value. While their students did manage to distinguish the two facets of attainment value, they also noted that attainment-value facets often correlated strongly with other task-value facets or even an entire task-value. For example, 'personal importance' correlated moderately with 'social utility' in their students, perhaps reflecting a relationship between their students' math identity and their interpersonal relationships, such as “impressing others with math competencies” (Gaspard et al., 2015). Intrinsic value and ‘personal importance’ were also more highly correlated with each other than the other attainment value facet, ‘importance of achievement’, was with ‘personal importance’. These results highlight a potential risk with exploring students’ task-values through task-value facets, that the correlations between facets may mean that each facet cannot be treated as strictly distinct from others. A multi-dimensional model of task-value facets should therefore be confirmed in other educational contexts before using it to compare to other variables.

Students’ Task-Values and Performance on BioVEDA

For our second research goal, we sought to determine whether students’ task-value facets as identified by our first research goal were predictive of their understanding of statistical concepts. We found that, after controlling for students’ academic achievement via GPA, prior statistics experience, and institution, students’ value of the utility of statistics for school and emotional cost of statistics predicted their understanding of variation in experimental design and analysis as measured by their performance on the BioVEDA assessment. Students with higher utility of statistics for school performed better than their peers with lower utility for school, with a one-unit increase in statistics utility resulting in a 0.568 point increase in BioVEDA scores (Figure 2.2a). Students with lower emotional cost towards statistics performed better than their peers with higher emotional cost, with a one-unit increase in emotional cost resulting in a 0.985 decrease in BioVEDA scores (Figure 2.2.b). Intrinsic value, utility for daily life, utility for career/job, and effort required did not present significant linear relationships to students’ BioVEDA scores.

The relationship between students' statistics utility for school and their performance but their lack of relationships between other task-value facets may be related to the generally negative or apathetic attitudes students hold towards statistics (Gal & Ginsburg, 1994), in particular their interest (intrinsic value) towards the topic and its perceived utility for students' career aspirations (Evans, 2007). Evans (2007) also identifies the importance of students' individual backgrounds and experiences in influencing their task-values, especially their interest. While our surveyed students were all self-reported life-sciences majors, the breadth of the life sciences and diversity of available majors at both institutions may mean that, for an equally broad subject such as statistics, our students' precise interests may misalign with the topics and material discussed and assessed in each respective course, obfuscating or limiting the relationship between those interests and their performance with the course content. In particular, the items for intrinsic value were very general and broad, painting interest in terms of 'fun doing statistics,' or 'simply enjoying dealing with statistical topics,' without capturing specific aspects of what it could mean to be intrinsically motivated by statistics. The breadth of student experiences may also explain why we did not see a relationship between utility for daily life or utility for career/job and performance. Our students have a wide variety of lived experiences and career aspirations, and similarly to their interests, the courses at the surveyed institutions simply may not have emphasized in a way that resonates with our students. Additionally, each utility value facet had only a couple of general items which described the facet; for example, utility for career had only two items, and only one such item explicitly mentioned the terms 'career' and 'job'. Thus, items for these facets may not have captured the utility for statistics for daily life or career that our students actually hold.

The observed relationship between emotional/psychological cost and performance but not between effort required and performance is also striking. Our results reinforce the evidence from Flake et al. (2015) suggesting that emotional cost may be more closely related to performance than the other cost facets. This also reflects the findings of previous studies which found that students often expressed considerable anxiety towards statistics which could unduly impact their engagement and performance (McKim, 2014; Chang & Beilock, 2016), especially with unfamiliar or less-familiar contexts and applications (Ruggeri et al., 2011). Conspicuously, our lack of a relationship between effort required and performance contrasts with previous studies which found relationships between effort and performance and achievement (Perez et al., 2014; Perez et al., 2019). This may be partially because of the specific wording of our survey items describing this facet (Table 2.2). Our items centered heavily around the 'energy' expended by engaging with statistics, while the items characterizing effort required in previous studies asked about not only energy, but time, money, and general 'effort' as well (Perez et al., 2014; Flake et al., 2015; Trautwein et al., 2012). Therefore, it is possible that simply discussing the energy required to engage with statistics did not fully capture our students' perception of their effort towards their statistics courses, resulting in no observable relationship between that effort and their performance on the assessment.

Teaching Implications

Our results have important implications for instructors seeking to better understand their students' motivation, engagement, and performance in biostatistics courses. In particular, because students distinguish between various facets of what makes a task 'useful', instructors should be careful to frame the utility of their content or material in ways which align with their students' specific values, such as how statistics can be useful within a variety of scientific and non-scientific careers or, more simply, in future courses. Instructors can shape their students' utility values through active interventions such as having students write about their personal connections to and perceived utility of statistics (Canning et al., 2018), which typically produce more meaningful and longer-lasting impacts than directly discussing with or explaining the utility of the subject with students (Canning & Harackiewicz, 2015). In our sample of students where statistics utility for school was predictive of their performance on BioVEDA, encouraging students to describe how the concepts in the statistics courses relate to the subsequent life-sciences courses in their major and drawing connections between their personal interests to statistics may increase their utility for statistics.

Additionally, instructors should pay close attention to the costs students perceive towards the content and material, in particular the anxiety students feel towards doing statistics. Previous studies have investigated math and statistics anxiety and describe interventions to address it (Onwuegbuzie & Wilson, 2003; Chang & Beilock, 2016; Ramirez et al., 2018). Simple interventions provide low-stakes activities in-class for students to 'practice' asking for help from their instructors (Pan & Tang, 2004), or increasing instructor and help availability in and out of the classroom by offering more individualized or personal office hours and tutoring sessions for students struggling with anxiety towards the task. More involved interventions have included training instructors in a variety of cognitive and psychosocial techniques to help students manage their own anxiety, such as the reappraisal and regulation of pre-performance anxiety (Chang & Beilock, 2016) or through activities designed to re-frame students' mindsets about failure and their anxiety by giving them opportunities to experience low-stakes setbacks in authentic contexts (Ramirez et al., 2018).

Limitations and Future Directions

This study has several limitations to consider when interpreting our results. A significant limitation stems from our small sample size, both for our factor analyses and our regressions. While we were able to achieve good model fits with multiple factors using only a sample size of 284 students, this is in contrast to the study by Gaspard et al. (2015) which surveyed nearly an order of magnitude more participants. Furthermore, because many students did not complete both the attitude and knowledge surveys or did not report their GPA, our sample size was further reduced to 101 students for our regressions, which limited our ability to model for additional effects. The specific characteristics of our sample also limit the generalizability of our results. The courses from both institutions were fairly ethnically homogenous (our sample was predominantly white students), and the majority of our students identified as female. Future studies should aim to describe task-value facets in underrepresented students and explore how their distinct experiences influence the relationship between their values and their performance. Furthermore, given that there are gender differences in task-value facets (e.g., Gaspard et al., 2015), studies should also seek to better characterize the task-value facets in male life-science students. Lastly, while we aimed to survey introductory statistics courses which closely matched each others' topics and content, future studies would also benefit significantly from surveying additional institutions outside of only public research universities, from a variety of introductory statistics courses with differing structure and instructional styles, to provide a broader picture of students' task-values towards statistics.

Additional limitations arise from the implementation of our survey instruments. The attitude survey drew considerably from the original instrument as described by Gaspard et al. (2015), which was tailored specifically to their study population (secondary school students).

While our study replaced only the subject of ‘math’ with ‘statistics’, it is possible that several of the items as worded by Gaspard et al. (2015) may be insufficiently precise with respect to the task-values of our population of undergraduate life-sciences students. Gaspard et al. (2015) identified the need to test their model in different student populations but additionally warned that populations of a different age group or academic stage may distinguish value facets more finely than their students. Furthermore, some facets such as utility for career or utility for school were measured by only two items, which could reduce the reliability and validity of those scales. Further studies may find it fruitful to revisit, revise, or expand the wording of specific items to tailor them to the study population more closely, although this would likely require further analyses to validate the revisions as a suitable measure.

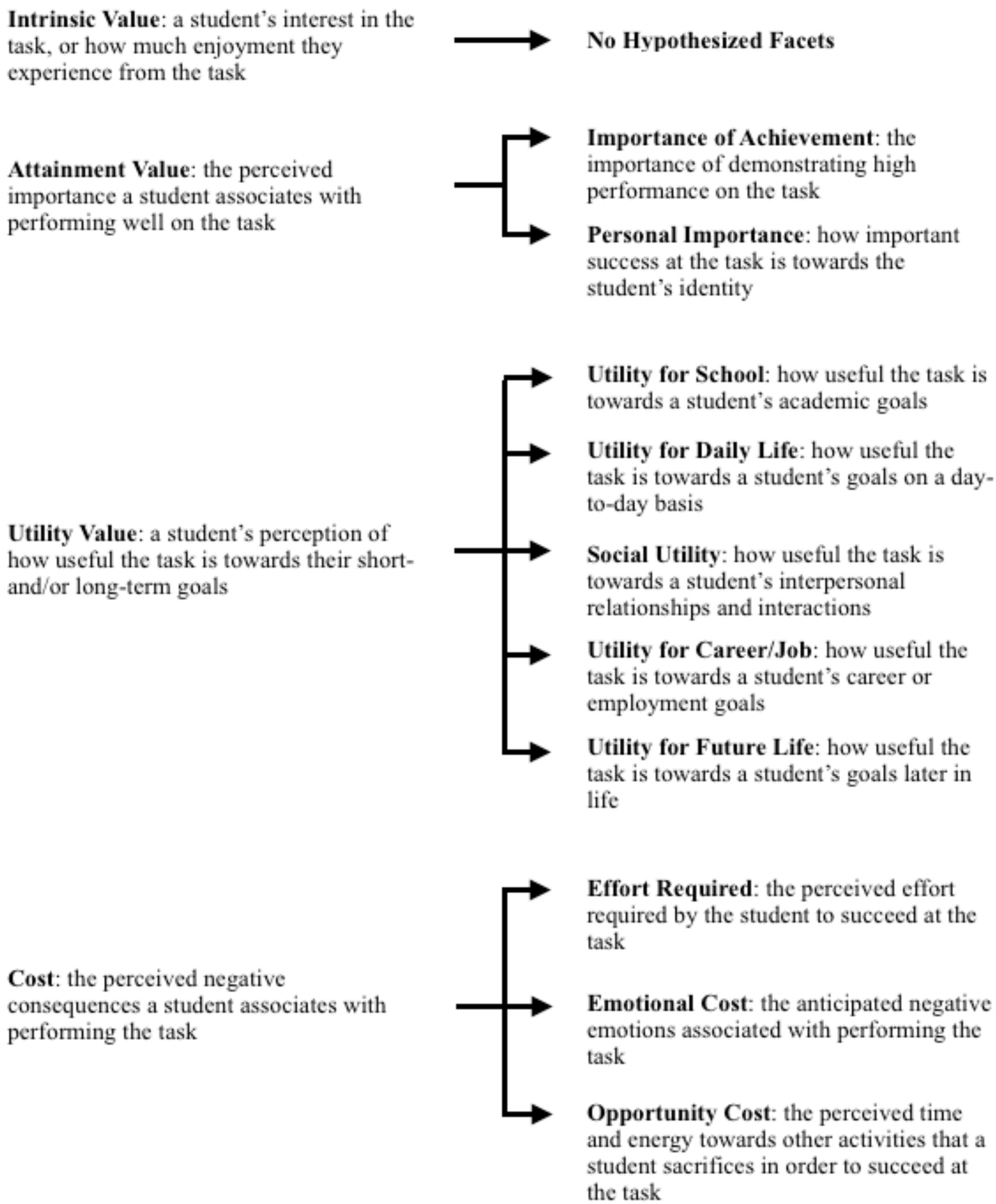
Lastly, while the BioVEDA instrument was validated for use with undergraduate life-sciences students learning statistics, the instrument may not have been ideal for our sample. We selected BioVEDA for its focus on the key concept of variation in the important applied context of experimental design, a concept which was well-covered by the instructors of each course. However, the majority of the instrument’s items used developmental biology experiments as the context for the questions, which may have been unfamiliar to our students. In particular, while each course discusses concepts in relation to experimental design, it is not an explicit focus of either course. Furthermore, given students’ difficulty in generalizing statistical terminology and symbology to underlying concepts (Kim et al., 2016), slight differences in the presentation of the principle idea of variation and its application in experimental design between each course and the survey instrument may have resulted in an overall lack of understanding of the items by our students, impacting their performance. Future studies should carefully consider whether the context of the BioVEDA is familiar enough to their students to ensure that the instrument adequately measures their performance with statistics.

Conclusion

This study found that undergraduate life-science students' task-values towards statistics are better described using multiple specific facets of each canonical task-value, rather than treating the task-values as monolithic constructs. In particular, students differentiate between five utility value facets and three cost facets. The two hypothesized facets of attainment value were not well supported by our models, in contrast to the findings of Gaspard et al. (2015) who initially described these facets. We additionally found that students' statistics utility for school and emotional cost towards statistics were predictive of their performance on an assessment designed to measure their understanding of variation in experimental design. Students who found statistics more useful for their academic goals performed better on the assessment than their peers with lower utility for school, while students who expressed lower emotional costs towards statistics performed better on the assessment than their peers who found statistics more emotionally costly. Further exploration of students' values as measured through these specific task-value facets may provide a clearer or more precise understanding of how they impact performance, and provide a basis for more targeted or tailored interventions designed to increase students' performance. Ultimately, statistics instructors and educators seeking to increase students' engagement with the material and their performance may benefit from content or interventions which more specifically target these facets as opposed to a more general approach to increasing students' utility value or decreasing their cost.

Figures

Figure 2.1: Canonical Task-Values and Task-Value Facets. The hypothesized breakdown of the four canonical task-values into their respective task-value facets. Definitions were drawn from Gaspard et al. (2015), Wigfield and Eccles (2000), and Eccles et al. (1983)



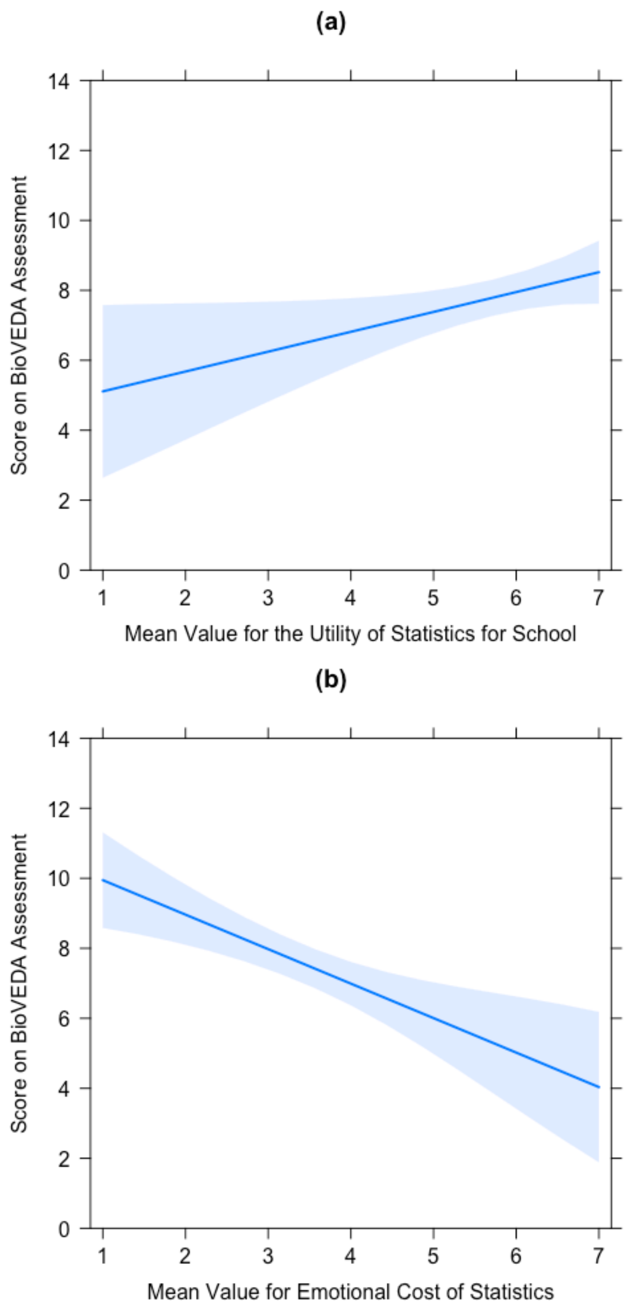


Figure 2.2: Relationship Between Task-Value Facets and BioVEDA Score. n = 101 for both models. Both models controlled for students' self-reported GPA, prior statistics experience, and institution. The shaded regions represent the 95% point-wise confidence interval of the estimated effect. **(a)** The mean value for the utility of statistics for school versus BioVEDA score. Students with higher mean value for utility of statistics for school performed better on the BioVEDA assessment. The shaded region represents the 95% point-wise confidence interval of the estimated effect. This regression model was statistically significant ($F(6, 94) = 2.657$, $p = 0.020$), and accounted for roughly 9% of the variance in BioVEDA scores (Adj. $R^2 = 0.090$). We found two significant predictors: utility for school (β: 0.568; standard error: 0.267; $p = 0.036$) and self-reported GPA (β: 1.359; standard error 0.572: $p = 0.020$). **(b)** The mean value for the emotional cost of statistics versus BioVEDA score. Students with lower mean emotional cost of statistics performed better on the BioVEDA assessment. This regression model was statistically significant ($F(5, 95) = 4.307$, $p = 0.001$), and accounted for slightly over 14% of the variance in BioVEDA scores (Adj. $R^2 = 0.142$). We found a single significant predictor: emotional cost (β: -0.985; standard error: 0.283; $p < 0.001$).

Tables

Table 2.1: Demographics of Study Participants. n = 284. Sample excludes students who were non-life-sciences. Some characteristics may not have percentages which total to 100% due to excluding students who did not or preferred not to respond for a given characteristic.

Characteristic	Northeastern University (n = 232)	Western University (n = 52)
Gender		
Male	33%	33%
Female	66%	67%
Other	1%	0%
Year in School		
First Year	14%	23%
Second Year	57%	25%
Third Year	19%	37%
Fourth Year	7%	14%
Other	3%	1%
Prior Statistics Course?		
Yes	27%	52%
No	73%	48%
Pre-Professional Status		
In a Program	44%	37%
Not in a Program	52%	56%
First Generation Student?		
Yes	10%	10%
No	89%	88%
Unsure	1%	0%
Race / Ethnicity		
American Indian / Alaskan Native	0%	0%
Asian	3%	10%
Black	1%	0%
Pacific Islander	0%	0%
White	90%	71%
Hispanic / Latinx	2%	6%
Other	1%	2%
Multiracial	3%	8%
Mean self-reported GPA (± Standard Deviation)	3.40 ± 0.43	3.33 ± 0.44

Table 2.2: Standardized Factor Loadings of Survey Items. n = 284. Factor loadings for each survey item are listed for each set of models tested: 1) the single-factor models describing each canonical task-value as a single construct; 2) the multi-factor models describing each task-value facet within a canonical task-value as separate constructs; 3) the 3-factor model of Intrinsic Value, Utility Value, and Cost as single constructs; 4) the 9-factor model of all task-value facets for Intrinsic Value, Utility Value, and Cost as separate constructs. For the 3-factor model and the 9-factor model, Attainment Value was excluded, indicated by '~'. As there are no hypothesized facets within Intrinsic Value, we did not test a model with facets, also indicated by '~'.

Survey Item	Item Means (Standard Deviations)	Canonical Task-Values as Single Constructs	Task-Value Facets as Separate Constructs	3-Factor Model of Canonical Task-Values as Single Constructs	9-Factor Model of Task-Value Facets as Separate Constructs
Intrinsic					
Statistics is fun to me.	4.046 (1.525)	0.932	~	0.933	0.933
I like doing statistics.	4.239 (1.531)	0.961	~	0.958	0.959
I simply like statistics.	4.060 (1.489)	0.876	~	0.876	0.876
I enjoy dealing with statistical topics.	4.229 (1.513)	0.790	~	0.795	0.796
Attainment - Importance of Achievement					
It is important to me to be good at statistics.	5.504 (1.215)	0.778	0.856	~	~
Being good at statistics means a lot to me.	4.923 (1.340)	0.819	0.878	~	~
Performing well in statistics is important to me.	5.799 (1.147)	0.585	0.676	~	~
Good grades in statistics are very important to me.	6.127 (0.867)	0.244	0.320	~	~
Attainment - Personal Importance					
I care a lot about remembering the things we learn in statistics.	5.201 (1.224)	0.785	0.761	~	~
Statistics is not meaningful to me.	5.025 (1.442)	0.675	0.673	~	~
I'm really keen on learning a lot in statistics.	4.482 (1.317)	0.761	0.806	~	~

Survey Item	Item Means (Standard Deviations)	Canonical Task-Values as Single Constructs	Task-Value Facets as Separate Constructs	3-Factor Model of Canonical Task-Values as Single Constructs	9-Factor Model of Task-Value Facets as Separate Constructs
Statistics is very important to me personally.	3.975 (1.514)	0.712	0.766	~	~
To be honest, I don't care about statistics.	4.606 (1.605)	0.754	0.777	~	~
It is important to me to know a lot of statistics.	4.768 (1.265)	0.730	0.734	~	~
Utility - School					
It is worth making an effort in statistics, because it will save me a lot of trouble at school in the next years.	5.599 (1.254)	0.579	0.978	0.584	0.987
Being good at statistics pays off, because it is simply needed at school.	5.475 (1.271)	0.440	0.707	0.441	0.702
Utility - Daily Life					
Understanding statistics has many benefits in my daily life.	4.563 (1.559)	0.820	0.889	0.822	0.893
Statistics comes in handy in everyday life and leisure time.	4.025 (1.546)	0.813	0.936	0.811	0.932
Statistics is directly applicable in everyday life.	4.437 (1.547)	0.737	0.823	0.736	0.823
Utility - Social Utility					
Being well versed in statistics will go down well with my classmates.	4.813 (1.188)	0.577	0.665	0.576	0.665
I can impress others with intimate knowledge in statistics.	4.187 (1.488)	0.552	0.859	0.552	0.859
If I know a lot in statistics, I will leave a good impression on my classmates.	4.306 (1.343)	0.568	0.895	0.565	0.894
Utility - Career/Job					

Survey Item	Item Means (Standard Deviations)	Canonical Task-Values as Single Constructs	Task-Value Facets as Separate Constructs	3-Factor Model of Canonical Task-Values as Single Constructs	9-Factor Model of Task-Value Facets as Separate Constructs
Good grades in statistics can be of great value to me later on.	5.708 (1.114)	0.330	0.525	0.329	0.517
Learning statistics is worthwhile, because it improves my job and career chances.	5.849 (1.092)	0.596	0.895	0.599	0.909
Utility - Future Life					
Statistics contents will help me in my life.	5.359 (1.183)	0.730	0.884	0.730	0.889
I will often need statistics in my life.	4.961 (1.410)	0.761	0.877	0.759	0.871
Cost - Effort Required					
Doing statistics is exhausting to me.	4.056 (1.514)	0.913	0.910	0.914	0.910
I often feel completely drained after doing statistics.	3.739 (1.555)	0.953	0.959	0.952	0.959
Dealing with statistics drains a lot of my energy.	3.746 (1.572)	0.945	0.959	0.943	0.960
Learning statistics exhausts me.	3.729 (1.538)	0.934	0.939	0.934	0.939
Cost - Emotional Cost					
I'd rather not do statistics, because it only worries me.	3.077 (1.387)	0.656	0.841	0.661	0.845
When I deal with statistics, I get annoyed.	3.856 (1.585)	0.669	0.764	0.674	0.775
Statistics is a real burden to me.	3.162 (1.461)	0.690	0.895	0.694	0.897
Doing statistics makes me really nervous.	3.116 (1.523)	0.660	0.808	0.662	0.791
Cost - Opportunity Cost					
I have to give up other activities that I like to be successful at statistics.	3.102 (1.603)	0.568	0.873	0.570	0.873

Survey Item	Item Means (Standard Deviations)	Canonical Task-Values as Single Constructs	Task-Value Facets as Separate Constructs	3-Factor Model of Canonical Task-Values as Single Constructs	9-Factor Model of Task-Value Facets as Separate Constructs
I have to give up a lot to do well in statistics.	2.824 (1.462)	0.614	0.954	0.615	0.955
I'd have to sacrifice a lot of free time to be good at statistics.	3.099 (1.607)	0.638	0.915	0.638	0.914

Table 2.3: Thresholds of Model Fit Indices for All Confirmatory Factor Analyses. These thresholds were established based on the recommendations of Hu and Bentler (1999). Thresholds without recommendations are indicated by '~'

Threshold	*MLR χ2 p-value*	CFI	TLI	RMSEA	SRMR
Acceptable Fit	> 0.05	0.90	0.90	0.08	0.10
Good / Excellent Fit	~	0.95	0.95	0.06	0.08

Table 2.4: Fit Summary for Confirmatory Factor Analyses. n = 284. Model fit indices for analyses where each canonical task-value was examined as a single construct (1-factor models) compared to analyses which distinguished each of the facets as separate constructs (multi-factor models). "Combined models" describe where all canonical task-values as single constructs (3-factor model) were conducted in one analysis compared to all task-value facets for Intrinsic Value (1 'facet'), Utility Value (5 facets) and Cost (3 facets) were conducted in one analysis. Thresholds for 'acceptable' fit for CFI and TLI, and 'good fit' for RMSEA and SRMR are indicated in parentheses.

Analysis	χ2	*df*	*p* (>0.05)	CFI (>0.90)	TLI (>0.90)	RMSEA (<0.06)	SRMR (<0.08)
Intrinsic Value							
1 Factor	0.476	2	0.788	1	1	0.005	0.005
Attainment Value							
1 Factor	244	35	0.000	0.810	0.756	0.175	0.089
2 Factors	177	34	0.000	0.874	0.833	0.145	0.076
Utility Value							
1 Factor	663	54	0.000	0.616	0.531	0.226	0.125
5 Factors	91	44	0.000	0.972	0.958	0.068	0.050
Cost							
1 Factor	556	44	0.000	0.728	0.667	0.266	0.132
3 Factors	75	41	0.001	0.985	0.980	0.065	0.029
Combined Models							
All Canonical Task-Values (3-factor)	1739	321	0.000	0.724	0.698	0.140	0.103
Task-Value Facets (9-Factor)	488	288	0.000	0.973	0.967	0.047	0.048

Table 2.5: Regression Outputs for Task-Value Facets and BioVEDA Scores. n = 101.

Asterisks indicate significant predictors.

Task Value / Facet	*β*	*S.E.*	*t*	*p*	*F*	*df*	*p*	Adj. R^2
Intrinsic Value					2.189	96	0.076	0.045
Intercept	2.045	2.043	1.001	0.319				
Intrinsic	0.288	0.176	1.638	0.105				
Prior Stats Exp.	0.103	0.555	0.186	0.853				
GPA	1.233	0.577	2.135	0.035*				
School	0.262	0.578	0.453	0.651				
Utility Value					2.657	94	0.02*	0.090
Intercept	0.108	2.207	0.049	0.961				
Utility - School	0.568	0.267	2.128	0.036*				
Utility - Daily Life	0.238	0.221	1.077	0.2843				
Utility - Career/Job	-0.244	0.305	-0.801	0.425				
Prior Stats Exp.	0.030	0.556	0.054	0.957				
GPA	1.359	0.572	2.374	0.012*				
School	0.240	0.610	0.393	0.6949				
Cost					4.307	95	0.001*	0.142
Intercept	6.660	2.159	3.085	0.003*				
Cost - Effort Required	0.373	0.249	1.500	0.137				
Cost - Emotional Cost	-0.985	0.284	-3.475	0.001*				
Prior Stats Exp.	0.151	0.527	0.287	0.775				
GPA	0.787	0.562	1.399	0.165				
School	0.015	0.552	0.027	0.978				

CONCLUSION

In this thesis, I explored two key aspects of students' motivation towards quantitative biology in different scenarios: characterizing the specific experiences of students when working together in groups and how those experiences shape their math self-efficacy beliefs; and investigating students' task-values towards statistics and how they relate to their understanding of a statistical concept in an applied context. My findings show that these constructs, widely studied in other educational contexts and fields, are useful in understanding biology undergraduates' engagement towards math in biology and provide an avenue for future investigation into how better to integrate quantitative skills into modern biology curricula.

In Chapter 1, I explored how, when working together in small groups to complete quantitative biology tasks like evaluating Hardy-Weinberg Equilibria and modeling population growth, students draw from their experiences in group work, which in turn reflect different sources of self-efficacy, to build or diminish their self-efficacy beliefs. We also asked how students' math self-efficacy related to the sources of self-efficacy they reported which increased or decreased their self-efficacy. When building their self-efficacy, many students reported experiences which reflected a mastery experience, such as succeeding at a problem on the group work assignment, being able to verify their success with their peers, or even teach or guide their struggling peers which reinforced their own confidence in their abilities. These findings support the theory and literature which argue that mastery experiences are a critical source of self-efficacy (Bandura, 1997; Usher & Pajares, 2008; Butz & Usher, 2015). We also found that most students drew confidence from the discussion of different ideas and approaches to problems, and the ability for group members to help one another and ask each other questions in a large-lecture environment. These specific experiences highlight how the unique social dynamics of working in groups can positively impact students' confidence (Nokes-Malach et al., 2015; Felder & Brent, 2016; Lent et al., 2006). We also found relationships between students' self-efficacy levels when entering the group work assignment and the sources of self-efficacy they reported increased their confidence: students with higher self-efficacy tend to experience more mastery and rely less than their lower self-efficacy peers. This reinforces evidence that students of varying self-efficacy levels may develop their self-efficacy beliefs through different sources (Usher & Pajares, 2008; Butz & Usher, 2015) and highlights the importance of providing opportunities for multiple sources of self-efficacy when designing interventions which target it. Additionally, our findings may help explain how students working in diverse or heterogenous groups tend to perform better (Donovan et al., 2018), as the ability to work with others benefits lower self-efficacy students through the availability of help, while also providing opportunities for higher self-efficacy students to demonstrate their mastery.

A more complex story emerges from the experiences which decreased students' confidence, however. While most students encouragingly expressed that group work did not decrease their confidence, those who did experience a decrease reported a huge breadth of negative experiences reflecting a wide range of self-efficacy sources. Once again, mastery experiences—or, rather, a lack thereof—were most prevalent in hurting students' confidence (Bandura, 1997), but the specific experiences of students ranged from simply making mistakes on a problem, to being unsure of their success because their group failed to verify their answers or collaborate with each other, to feeling pressured for time and being unable to keep up with their group mates. Additionally, students frequently expressed that their groups failed to communicate openly or consistently, feeling also that their group mates were unable to provide the support they needed to build their confidence in solving the problems. Some students felt so pressured by their relative progress or success compared to their peers that their confidence in their own abilities diminished. While studies have found negative impacts to engagement and performance when groups become dysfunctional or ineffective (Chang & Brickman, 2018; Donovan et al., 2018; Nokes-Malach et al., 2015), our findings suggest an explanation for how these negative impacts manifest, by shaping students' self-efficacy beliefs. Additionally, we found that students with lower self-efficacy tended to report a lack of mastery more frequently than their higher self-efficacy peers, further highlighting the importance of mastery experiences as a source of self-efficacy but also underscoring the importance for instructors to target lower self-efficacy students given the breadth and variety of negative experiences we observed in our students during group work. Further qualitative work is necessary to dive more deeply into these experiences, both positive and negative, to better understand how they emerge through group work and how they relate to other experiences as students form their self-efficacy beliefs. Interviewing students can provide a focused and individual lens through which to examine these relationships to better understand what is going on when students report their experiences, as our findings show that what students tell us about their confidence provides a unique window through which educators can understand their students' motivation, engagement, and performance, and help them create more effective interventions to support the development of students' self-efficacy towards quantitative biology.

In Chapter 2, I investigated how to better represent and characterize students' task-values towards statistics, and the relationship of their task-values to their understanding of biological variation in experimental design as measured by their performance. We found that students' task-values are better described using a model which differentiates the four canonical task-values—intrinsic value, attainment value, utility value, and cost—into multiple dimensions or ‘task-value facets.’ While intrinsic value is not typically described with multiple facets, our model distinguished between multiple utility value facets and multiple cost facets. These findings extend those of other studies which focused on a specific task-value, cost especially (Perez et al., 2014; Flake et al., 2015) by revealing similar dimensionality within our students’ task-values, and reflect those of Gaspard et al. (2015) who investigated math task-value facets in secondary-school students. In their conclusion, they called for others to examine their model of multiple task-value facets in other populations and other contexts. Our findings represent a step in that direction by exploring the model in biology students and suggest a model of several task-value facets can be a more focused tool for understanding the task-values of biology undergraduates towards math in biology.

We also found that these task-value facets may individually predict students’ performance and understanding. In our students, the utility of statistics for school / academics related positively to students’ performance on the statistical assessment, while the emotional / psychological cost of statistics related negatively to performance on the statistical assessment. Our results emphasize the importance of utility value (Conley, 2012; Luttrell et al., 2010) by revealing how it encompasses a variety of different aspects or domains in students’ lives to different degrees (Peetsma & van der Veen, 2011). They also add to the mounting evidence of the significant influence of cost in students’ motivation and achievement (Flake et al., 2015). Further exploration of task-value facets and how they may affect students of varying backgrounds in different contexts or domains can provide instructors with insight into their students' motivation and benefit them towards developing more targeted or personal interventions to improve engagement and performance in their students.

Overall, these findings highlight how exploring student motivation is a useful and meaningful lens through which to examine both students' performance in using and understanding math in biology as well as how to reinforce the development of their quantitative skills. The constructs of self-efficacy, both in the context of social cognitive theory and expectancy-value theory, and task-values frame a window into how and why students engage with the quantitative lessons we present them as well as how to shape their beliefs about their ability to tackle these problems. These insights into students' motivation reveal promising leads in the investigation of how best to incorporate quantitative biology into new curricula, and how biology educators and instructors can better help their students meet the challenges of modern biology.

REFERENCES

Ainscough, L., Foulis, E., Colthorpe, K., Zimbardi, K., Robertson-Dean, M., Chunduri, P., & Lluka, L. (2016). Changes in biology self-efficacy during a first-year university course. *CBE–Life Sciences Education*, 15(2). https://doi.org/10.1187/cbe.15-04-0092

American Association for the Advancement of Science. (2011). Vision and change in undergraduate biology education: A call to action. https://www.visionandchange.org

Andrews, S. E., Runyon, C., & Aikens, M. L. (2017). The Math–Biology Values Instrument: Development of a tool to measure life science majors' task values of using math in the context of biology. *CBE–Life Sciences Education*, 16(3), ar45. https://doi.org/10.1187/cbe.17-03-0043

Bailey, C. P., Minderhout, V., & Loertscher, J. (2012). Learning transferable skills in large lecture halls: Implementing a POGIL approach in biochemistry. *Biochemistry and Molecular Biology Education*, 40(1), 1–7. https://doi.org/10.1002/bmb.20556

Bandura, A. (1986). Social foundations of thought and action. Englewood Cliffs, New Jersey, 1986, 23–28.

Bandura, A. (1997). Self-efficacy: The exercise of control. W.H. Freeman, New York, 1997, 79–115.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72(1), 187–206. https://doi.org/10.1111/1467-8624.00273

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01

Ben-Zvi, D., & Garfield, J. (2004a). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42–63. http://dx.doi.org/10.1007/978-1-4020-8383-9

Ben-Zvi, D., & Garfield, J. (2004b). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. The challenge of developing statistical literacy, reasoning, and thinking. Springer, Dordrecht, 3–15. https://doi.org/10.1007/1-4020-2278-6_1

Bialek, W., & Botstein, D. (2004). Introductory science and mathematics education for 21st-century biologists. *Science*, 303(5659), 788–790. https://doi.org/10.1126/science.1095480

Bong, M. (2001). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. *Contemporary Educational Psychology*, 26(4), 553–570. https://doi.org/10.1006/ceps.2000.1048

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1–40. https://doi.org/10.1023/A:1021302408382

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, 43(5), 485–499. https://doi.org/10.1002/tea.20131

Butz, A. R., & Usher, E. L. (2015). Salient sources of early adolescents' self-efficacy in two domains. C*ontemporary Educational Psychology*, 42, 49–61. https://doi.org/10.1016/j.cedpsych.2015.04.001

Canning, E. A., & Harackiewicz, J. M. (2015). Teach it, don't preach it: The differential effects of directly-communicated and self-generated utility-value information. *Motivation Science*, 1(1), 47–71. https://doi.org/10.1037/mot0000015

Canning, E. A., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Tibbetts, Y., & Hyde, J. S. (2018). Improving performance and retention in introductory biology with a utility-value intervention. *Journal of Educational Psychology*, 110(6), 834–849. https://doi.org/10.1037/edu0000244

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*, 98–113. https://doi.org/10.1016/j.edurev.2007.04.001

Chang, H., & Beilock, S. L. (2016). The math anxiety-math performance link and its relation to individual and environmental factors: A review of current behavioral and psychophysiological research. *Current Opinion in Behavioral Sciences*, 10, 33–38. https://doi.org/10.1016/j.cobeha.2016.04.011

Chang, Y., & Brickman, P. (2018). When group work doesn't work: Insights from students. *CBE–Life Sciences Education*, 17(3), ar52. https://doi.org/10.1187/cbe.17-09-0199

Chen, J. A., & Usher, E. L. (2013). Profiles of the sources of science self-efficacy. *Learning and Individual Differences*, 24, 11–21. https://doi.org/10.1016/j.lindif.2012.11.002

Chiang, E. S., Byrd, S. P., & Molin, A. J. (2011). Children's perceived cost for exercise: Application of an expectancy-value paradigm. *Health Education & Behavior*, 38(2), 143–149. https://doi.org/10.1177/1090198110376350

Chiel, H. J., McManus, J. M., Shaw, K. M., & Jungck, J. (2010). From biology to mathematical models and back: Teaching modeling to biology students, and biology to math and engineering students. *CBE–Life Sciences Education*, 9(3), 248–265. https://doi.org/10.1187/cbe.10-03-0022

Colon-Berlingeri, M., & Burrowes, P. A. (2011). Teaching biology through statistics: Application of statistical methods in genetics and zoology courses. *CBE—Life Sciences Education*, 10(3), 259–267. https://doi.org/10.1187/cbe.10-11-0137\

Conley, A. M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology*, 104(1), 32–47. https://doi.org/10.1037/a0026042

Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *International Statistical Review,* 75, 372–396. http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(2)_delMas.pdf

Depelteau, A. M., Joplin, K. H., Govett, A., Miller, H. A., III, & Seier, E. (2010). SYMBIOSIS: Development, implementation, and assessment of a model curriculum across biology and mathematics at the introductory level. *CBE–Life Sciences Education*, 9(3), 342-347. https://doi.org/10.1187/cbe.10-05-0071

Donovan, D. A., Connell, G. L., & Grunspan, D. Z. (2018). Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. *CBE–Life Sciences Education*, 17(4), ar60. https://doi.org/10.1187/cbe.17-12-0283

Drott, M. and M.A. Sarvary (2016). Why did the snake cross the road? A population genetics and habitat conversation case study. National Center for Case Study Teaching in Science. http://sciencecases.lib.buffalo.edu

Duffus, D., Olifer, A., & Jungck, J. (2010). Introductory life science mathematics and quantitative neuroscience courses. *CBE–Life Sciences Education*, 9(3), 370–377. https://doi.org/10.1187/cbe.10-03-0026

Durik, A., Hulleman, C., & Harackiewicz, J. (2015). One size fits some: Instructional enhancements to promote interest. Interest in mathematics and science learning. *American Educational Research Association*, 49–62. https://doi.org/10.3102/978-0-935302-42-4_3

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology*, 99(3), 597–610. https://doi.org/10.1037/0022-0663.99.3.597

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. F., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. Achievement and achievement motivation. 75–146.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eddy, S. L., Brownell, S. E., Thummaphan, P., Lan, M.-C., & Wenderoth, M. P. (2015). Caution, student experience may vary: Social identities impact a student's experience in peer discussions. *CBE–Life Sciences Education*, 14(4), ar45. https://doi.org/10.1187/cbe.15-05-0108

Elliott, B., Oty, K., McArthur, J., & Clark, B. (2001). The effect of an interdisciplinary algebra/science course on students' problem solving skills, critical thinking skills and attitudes towards mathematics. *International Journal of Mathematical Education in Science and Technology*, 32(6), 811–816. https://doi.org/10.1080/00207390110053784

Evans, B. (2007). Student attitudes, conceptions, and achievement in introductory undergraduate college statistics. *The Mathematics Educator*, 17(2), 24–30. https://eric.ed.gov/?id=EJ841563

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, 28(2), 161–186. https://doi.org/10.1016/S0361-476X(02)00015-2

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, 41, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Fox, J., Weisberg, S. (2019). An R companion to applied regression. 3rd edition. Sage, Thousand Oaks, CA. http://tinyurl.com/carbook

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, 2, 1–16. https://doi.org/10.1080/10691898.1994.11910471

García-Santillán, A., Venegas-Martinez, F., Escalera Chávez, M. E., & Córdova-Rangel, A. (2013). Attitude toward statistic in college students (An empirical study in public university). *Journal of Statistical and Econometric Methods*, 2(1), 43–60.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review,* 75(3), 372–396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, 107(3), 663–677. https://doi.org/10.1037/edu0000003

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology*, 48, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003

Gill, R. (2011). Effective strategies for engaging students in large-lecture, nonmajors science courses. *Journal of College Science Teaching*, 41(2), 1–8. https://eric.ed.gov/?id=EJ964626

Gillies, R. M. (2013). Structuring cooperative group work in classrooms. *International Journal of Educational Research*, 39, 35–49. https://doi.org/10.1016/S0883-0355(03)00072-7

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2007). Nonscience majors learning science: A theoretical model of motivation. *Journal of Research in Science Teaching*, 44(8), 1088–1107. https://doi.org/10.1002/tea.20181

Gogol, K., Brunner, M., Martin, R., Preckel, F., & Goetz, T. (2017). Affect and motivation within and between school subjects: Development and validation of an integrative structural model of academic self-concept, interest, and anxiety. *Contemporary Educational Psychology*, 49, 46–65. https://doi.org/10.1016/j.cedpsych.2016.11.003

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE–Life Sciences Education*, 11(4), 364–377. https://doi.org/10.1187/cbe.12-03-0026

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction*, 49, 81–91. https://doi.org/10.1016/j.learninstruc.2016.12.007

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., Cambria, J., Flunger, B., Dicke, A.-L., Häfner, I., Brisson, B., & Trautwein, U. (2016). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open*, 2(1), 1–20. https://doi.org/10.1177/2332858415626884

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, 332(6034), 1213–1216. https://doi.org/10.1126/science.1204820

Hagen, B., Awosoga, O., Kellett, P., & Dei, S. O. (2013). Evaluation of undergraduate nursing students' attitudes towards statistics courses, before and after a course in applied statistics. *Nurse Education Today*, 33(9), 949–955. https://doi.org/10.1016/j.nedt.2012.11.005

Harackiewicz, J. M., & Hulleman, C. S. (2010). The importance of interest: The role of achievement goals and task values in promoting the development of interest. *Social and Personality Psychology Compass*, 4(1), 42–52. https://doi.org/10.1111/j.1751-9004.2009.00207.x

Hicks, J., Dewey, J., Brandvain, Y., & Schuchardt, A. (2020). Development of the Biological Variation In Experimental Design And Analysis (BioVEDA) assessment. *PLoS ONE*, 15(7), e0236098. https://doi.org/10.1371/journal.pone.0236098

Hodges, L. C. (2018). Contemporary issues in group learning in undergraduate science classrooms: A perspective from student engagement. *CBE–Life Sciences Education*, 17(2), es3. https://doi.org/10.1187/cbe.17-11-0239

Hoffman, K., Leupen, S., Dowell, K., Kephart, K., & Leips, J. (2016). Development and assessment of modules to integrate quantitative skills in introductory biology courses. *CBE–Life Sciences Education*, 15(2), ar14. https://doi.org/10.1187/cbe.15-09-0186

Holley, J. W., & Guilford, J. P. (1964). A note on the G index of agreement. *Educational and Psychological Measurement*, 24(4), 749–753. https://doi.org/10.1177/001316446402400402

Hood, M., Creed, P. A., & Neumann, D. L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics 6. *Statistics Education Research Journal,* 11(2), 72–85. http://www.stat.auckland.ac.nz/serj

Horton, N. J., & Hardin, J. S. (2015). Teaching the next generation of statistics students to "think with data": Special issue on statistics and the undergraduate curriculum. *The American Statistician*, 69(4), 259–265. https://doi.org/10.1080/00031305.2015.1094283

Hoskinson, A. M. (2010). How to build a course in mathematical-biological modeling: Content and processes for knowledge and skill. *CBE–Life Sciences Education*, 9(3), 333–341. https://doi.org/10.1187/cbe.10-03-0041

Hruschka, D. J., Schwartz, D., St. John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, 16(3), 307–331. https://doi.org/10.1177/1525822x04266540

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, 102(4), 880–895. https://doi.org/10.1037/a0019506

Hutchison, M. A., Follman, D. K., Sumpter, M., & Bodner, G. M. (2006). Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, 95(1), 39–47. https://doi.org/10.1002/j.2168-9830.2006.tb00876.x

Hutchison-Green, M. A., Follman, D. K., & Bodner, G. M. (2008). Providing a voice: Qualitative investigation of the impact of a first-year engineering experience on students' efficacy beliefs. *Journal of Engineering Education*, 97(2), 177–190. https://doi.org/10.1002/j.2168-9830.2008.tb00966.x

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Jones, B. D., Paretti, M. C., Hein, S. F., & Knott, T. W. (2010). An analysis of motivation constructs with first-year engineering students: Relationships among expectancies, values, achievement, and career plans. *Journal of Engineering Education*, 99(4), 319–336. https://doi.org/10.1002/j.2168-9830.2010.tb01066.x

Joshi, M., Aikens, M. L., & Dolan, E. L. (2019). Direct ties to a faculty mentor related to positive outcomes for undergraduate researchers. *Bioscience*, 69(5), 389–397. https://doi.org/10.1093/biosci/biz039

Karpiak, C. P. (2011). Assessment of problem-based learning in the undergraduate statistics course. *Teaching of Psychology*, 38(4), 251–254. https://doi.org/10.1177/0098628311421322

Kim, H. W., Fukawa-Connelly, T., & Cook, S. A. (2016). Student understanding of symbols in introductory statistics courses. The teaching and learning of statistics: International perspectives. Springer International Publishing, 163–174. https://doi.org/10.1007/978-3-319-23470-0_21

Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology*, 25(4), 615–624. https://doi.org/10.1002/acp.1730

Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. The decade ahead: Theoretical perspectives on motivation and achievement. Emerald Group Publishing Limited, Bingley, 1–33 https://doi.org/10.1108/s0749-7423(2010)000016a004

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE–Life Sciences Education*, 18(1), rm1. https://doi.org/10.1187/cbe.18-04-0064

Lee, W., Lee, M.-J., & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology*, 39(2), 86–99. https://doi.org/10.1016/j.cedpsych.2014.02.002

Lent, R. W., Schmidt, J., & Schmidt, L. (2006). Collective efficacy beliefs in student work teams: Relation to self-efficacy, cohesion, and performance. *Journal of Vocational Behavior*, 68(1), 73–84. https://doi.org/10.1016/j.jvb.2005.04.001

Liu, C., & Zhu, R. (2016). Set theory, logic, and probability: The integration of qualitative reasoning into teaching statistics for quantitative biology. *CBE–Life Sciences Education*, 15(4). https://doi.org/10.1187/cbe.16-06-0184

Luttrell, V. R., Callen, B. W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. C. S. (2010). The Mathematics Value Inventory for general education students: Development and initial validation. *Educational and Psychological Measurement,* 70(1), 142–160. https://doi.org/10.1177/0013164409344526

Marsteller, P., de Pillis, L., Findley, A., Joplin, K., Pelesko, J., Nelson, K., Thompson, K., Usher, D., & Watkins, J. (2010). Toward integration: From quantitative biology to mathbio-biomath. *CBE–Life Sciences Education*, 9(3), 165–171. https://doi.org/10.1187/cbe.10-03-0053

Masel, J., Humphrey, P. T., Blackburn, B., & Levine, J. A. (2015). Evidence-based medicine as a tool for undergraduate probability and statistics education. *CBE–Life Sciences Education*, 14(4), ar42. https://doi.org/10.1187/cbe.15-04-0079

Matthews, K. E., Hodgson, Y., & Varsavsky, C. (2013). Factors influencing students' perceptions of their quantitative skills. *International Journal of Mathematical Education in Science and Technology*, 44(6), 782–795. https://doi.org/10.1080/0020739x.2013.814814

McKim, C. (2014). Understanding undergraduate statistical anxiety. *Journal of Research in Education*, 24(2), 204–210. https://eric.ed.gov/?id=EJ1098188

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94(2), 535–546. https://doi.org/10.1037/a0013773

Metz, A. M. (2008). Teaching statistics in biology: using inquiry-based learning to strengthen understanding of statistical analysis in biology laboratory courses. *CBE–Life Sciences Education*, 7(3), 317–326. https://doi.org/10.1187/cbe.07-07-0046

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K. T., & Trautwein, U. (2011). Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, 22(8), 1058–1066. https://doi.org/10.1177/0956797611415540

Nagengast, B., Trautwein, U., Kelava, A., & Lüdtke, O. (2013). Synergistic effects of expectancy and value on homework engagement: The case for a within-person perspective. *Multivariate Behavioral Research*, 48, 428–460. https://doi.org/10.1080/00273171.2013.775060

National Research Council (2003). Bio2010: Transforming Undergraduate Education for Future Research Biologists. National Academies Press, Washington, D.C.

National Research Council. (2009). A New Biology for the 21st Century. National Academies Press, Washington, D.C.

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, 27(4), 645–656. https://doi.org/10.1007/s10648-015-9312-8

O'Brien, R. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41, 673–690. https://doi.org/10.1007/s11135-006-9018-6

Olimpo, J. T., Pevey, R. S., & McCabe, T. M. (2018). incorporating an interactive statistics workshop into an introductory biology course-based undergraduate research experience (CURE) enhances students' statistical reasoning and quantitative literacy skills. *Journal of Microbiology & Biology Education*, 19(1), 1–7. https://doi.org/10.1128/jmbe.v19i1.1450

Öntaş, Ö. C., & Tekindal, M. T. (2015). The effect of group work on the self-efficacy of social work students. *Procedia - Social and Behavioral Sciences*, 174, 1124–1128. https://doi.org/10.1016/j.sbspro.2015.01.804

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments–A comprehensive review of the literature. *Teaching in Higher Education*, 8(2), 195–209. https://doi.org/10.1080/1356251032000052447

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research,* 66(4), 543–578. https://doi.org/10.3102/00346543066004543

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, 19, 139–158. https://doi.org/10.1080/10573560308222

Pajares, F., Johnson, M. J., & Usher, E. L. (2007). Sources of writing self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English*, 42(1), 104–120. http://www.jstor.org/stable/40171749

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. Contemporary Educational Psychology, 20(4), 426–443. doi:10.1006/ceps.1995.1029

Pan, W., & Tang, M. (2004). Examining the effectiveness of innovative instructional methods on reducing statistics anxiety for graduate students in the social sciences. *Journal of Instructional Psychology*, 31(2), 149–159. https://doi.org/10.2466/11.04.PR0.115c12z3

Peetsma, T., & van der Veen, I. (2011). Relations between the development of future time perspective in three life domains, investment in learning, and academic achievement. *Learning and Instruction*, 21(3), 481–494. https://doi.org/10.1016/j.learninstruc.2010.08.001

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, 106(1), 315–329. https://doi.org/10.1037/a0034027

Perez, T., Dai, T., Kaplan, A., Cromley, J. G., Brooks, W. D., White, A. C., Mara, K. R., & Balsai, M. J. (2019). Interrelations among expectancies, task values, and perceived costs in undergraduate biology achievement. *Learning and Individual Differences*, 72, 26–38. https://doi.org/10.1016/j.lindif.2019.04.001

Poladian, L. (2013). Engaging life-sciences students with mathematical models: Does authenticity help. *International Journal of Mathematical Education in Science and Technology*, 44(6), 865–876. https://doi.org/10.1080/0020739x.2013.811301

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: past research, promising interventions, and a new interpretation framework. *Educational Psychologist*, 53(3), 145–164. https://doi.org/10.1080/00461520.2018.1447384

Revelle W (2019). Psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois. R package version 1.9.12. https://CRAN.R-project.org/package=psych

Rheinlander, K., & Wallace, D. (2011). Calculus, biology and medicine: A case study in quantitative literacy for science students. *Numeracy*, 4(1), 1–19. https://doi.org/10.5038/1936-4660.4.1.3

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2). http://dx.doi.org/10.18637/jss.v048.i02

Ruggeri, K., Dempster, M., & Hanna, D. (2011). The impact of misunderstanding the nature of statistics. *Psychology Teaching Review*, 17(1), 35–40. https://files.eric.ed.gov/fulltext/EJ932187.pdf

Ryan, R. M., & Deci, E. L. (2016). Facilitating and hindering motivation, learning, and well-being in schools: Research and observations from self-determination theory. Handbook of motivation at school. 2nd edition. Erlbaum, New Jersey, 96–119

Saldaña, J. (2015). The Coding Manual for Qualitative Researchers. 3rd edition. Sage, 96.

Savadori, L., Van Swol, L. M., & Sniezek, J. A. (2001). Information sampling and confidence within groups and judge advisor systems. *Communication Research*, 28(6), 737–771. https://doi.org/10.1177/009365001028006002

Sawtelle, V., Brewe, E., & Kramer, L. H. (2012). Exploring the relationship between self-efficacy and retention in introductory physics. *Journal of Research in Science Teaching*, 49(9), 1096–1121. https://doi.org/10.1002/tea.21050

Scager, K., Boonstra, J., Peeters, T., Vulperhorst, J., & Wiegant, F. (2016). Collaborative learning in higher education: evoking positive interdependence. *CBE–Life Sciences Education*, 15(4), ar69. https://doi.org/10.1187/cbe.16-07-0219

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, 57, 149–174.

Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. Development of achievement motivation. San Diego, CA: Academic Press, 15–31.

Schunk, D. H., & Pajares, F. (2005). Competence beliefs in academic functioning. Handbook of competence and motivation. Guilford Press, New York, 85–104.

Shechter, O., Durik, A., Miyamoto, Y., & Harackiewicz, J. (2011). The role of utility value in achievement behavior: the importance of culture. *Personality & Social Psychology Bulletin*, 37, 303–317. https://doi.org/10.1177/0146167210396380

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, 42(1), 70–83. https://doi.org/10.1037/0012-1649.42.1.70

Speth, E. B., Momsen, J. L., Moyerbrailean, G. A., Ebert-May, D., Long, T. M., Wyse, S., & Linton, D. (2010). 1, 2, 3, 4: Infusing quantitative literacy into introductory biology. *CBE–Life Sciences Education*, 9(3), 323–332. https://doi.org/10.1187/cbe.10-03-0033

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21–51. https://doi.org/10.2307/1170643

Steinmayr, R., Weidinger, A. F., Schwinger, M., & Spinath, B. (2019). The importance of students' motivation for their academic achievement - Replicating and extending previous findings. *Frontiers in Psychology*, 10(1730), 1–11. https://doi.org/10.3389/fpsyg.2019.01730

Theobald, E. (2018). Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE–Life Sciences Education*, 17(3), rm2. https://doi.org/10.1187/cbe.17-12-0280

Theobald, E. J., Eddy, S. L., Grunspan, D. Z., Wiggins, B. L., & Crowe, A. J. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PLoS One*, 12(7), e0181336. https://doi.org/10.1371/journal.pone.0181336

Theobald, E. J., Aikens, M., Eddy, S., & Jordt, H. (2019). Beyond linear regression: A reference for analyzing common data types in discipline based education research. *Physical Review Physics Education Research*, 15(2). https://doi.org/10.1103/physrevphyseducres.15.020110

Theobald, R., & Freeman, S. (2014). Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE–Life Sciences Education*, 13(1), 41–48. https://doi.org/10.1187/cbe-13-07-0136

Thompson, K. V., Nelson, K. C., Marbach-Ad, G., Keller, M., & Fagan, W. F. (2010). Online interactive teaching modules enhance quantitative proficiency of introductory biology students. *CBE–Life Sciences Education*, 9(3), 277–283. https://doi.org/10.1187/cbe.10-03-0028

Thompson, K. V., Cooke, T. J., Fagan, W. F., Gulick, D., Levy, D., Nelson, K. C., Redish, E. F., Smith, R. F., & Presson, J. (2013a). Infusing quantitative approaches throughout the biological sciences curriculum. *International Journal of Mathematical Education in Science and Technology*, 44(6), 817–833. https://doi.org/10.1080/0020739x.2013.812754

Thompson, K. V., Chmielewski, J., Gaines, M. S., Hrycyna, C. A., & LaCourse, W. R. (2013b). Competency-based reforms of the undergraduate biology curriculum: Integrating the physical and biological sciences. *CBE–Life Sciences Education*, 12(2), 162–169. https://doi.org/10.1187/cbe.12-09-0143

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, 104(3), 763–777. https://doi.org/10.1037/a0027470

Trujillo, G., & Tanner, K. D. (2014). Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE–Life Sciences Education*, 13(1), 6–15. https://doi.org/10.1187/cbe.13-12-0241

Usher, D. C., Driscoll, T. A., Dhurjati, P., Pelesko, J. A., Rossi, L. F., Schleiniger, G., Pusecker, K., & White, H. B. (2010). A transformative model for undergraduate quantitative biology education. *CBE–Life Sciences Education*, 9(3), 181–188. https://doi.org/10.1187/cbe.10-03-0029

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 78(4), 751–796. https://doi.org/10.3102/0034654308321456

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34(1), 89–101. https://doi.org/10.1016/j.cedpsych.2008.09.002

Wachsmuth, L. P., Runyon, C. R., Drake, J. M., & Dolan, E. L. (2017). Do biology students really hate math? Empirical insights into undergraduate life science majors' emotions about mathematics. C*BE–Life Sciences Education*, 16(3). https://doi.org/10.1187/cbe.16-08-0248

Wang, S.-L., & Lin, S. S. J. (2007). The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning. *Computers in Human Behavior*, 23(5), 2256–2268. https://doi.org/10.1016/j.chb.2006.03.005

Watkins, J. C. (2010). On a calculus-based statistics course for life science students. *CBE–Life Sciences Education*, 9(3), 298–310. https://doi.org/10.1187/cbe.10-03-0035

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, 30(1), 1–35. https://doi.org/10.1016/j.dr.2009.12.001

Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. Advances in motivation science. Elsevier, 161–198. https://doi.org/10.1016/bs.adms.2019.05.002

Wightman, B., & Hark, A. T. (2012). Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. *Biochemistry & Molecular Biology Education*, 40(5), 310–319. https://doi.org/10.1002/bmb.20637

Woolcock, A. D., Creevy, K. E., Coleman, A. E., Moore, J. N., & Brown, S. A. (2016). Assessing academic self-efficacy, knowledge, and attitudes in undergraduate physiology students. *American Journal of Educational Research*, 4(9), 652–657. doi: 10.12691/education-4-9-1

Wright, R., Cotner, S., & Winkel, A. (2009). Minimal impact of organic chemistry prerequisite on student performance in introductory biochemistry. *CBE–Life Sciences Education*, 8(1), 44–54. https://doi.org/10.1187/cbe.07-10-0093

Xu, S., & Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohen's kappa. *Journal Consulting and Clinical Psychology*, 82(6), 1219–1227. https://doi.org/10.1037/a0037489

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), 1081–1094. https://doi.org/10.1080/0950069032000052207

APPENDICES

Appendix A: Survey Instruments for Chapter 1

Hardy-Weinberg Equilibrium Pre-Survey

Name: ___________________________

Please consider the following problem about Hardy-Weinberg Equilibrium. **You do not have to solve it**.

A gene has two alleles: A and B. The number of individuals in a population with each genotype is shown in the table to the right.

AA	AB	BB
42	96	62

Please rate your confidence (circle the number) in your ability to successfully do the following:

	Not at all confident	A little confident	Fairly confident	Very confident	Completely confident
Calculate the predicted number of individuals of each genotype under the conditions of Hardy-Weinberg Equilibrium.	**1**	**2**	**3**	**4**	**5**
Justify whether the population is evolving or not using the Hardy-Weinberg Equilibrium model.	**1**	**2**	**3**	**4**	**5**

Hardy-Weinberg Equilibrium Post-Survey

Name: ___________________________

Please consider the following problem about Hardy-Weinberg Equilibrium. **You do not have to solve it**.

A gene has two alleles: A and B. The number of individuals in a population with each genotype is shown in the table to the right.

AA	AB	BB
42	96	62

Please rate your confidence (circle the number) in your ability to successfully do the following:

	Not at all confident	A little confident	Fairly confident	Very confident	Completely confident
Calculate the predicted number of individuals of each genotype under the conditions of Hardy-Weinberg Equilibrium.	**1**	**2**	**3**	**4**	**5**
Justify whether the population is evolving or not using the Hardy-Weinberg Equilibrium model.	**1**	**2**	**3**	**4**	**5**

Describe any experiences and/or interactions during group work today that **increased** your confidence in your ability to calculate the predicted number of individuals of each genotype under the conditions of Hardy-Weinberg Equilibrium.

Describe any experiences and/or interactions during group work today that **decreased** your confidence in your ability to calculate the predicted number of individuals of each genotype under the conditions of Hardy-Weinberg Equilibrium.

Describe any experiences and/or interactions during group work today that **increased** your confidence in your ability to justify whether the population is evolving or not using the Hardy-Weinberg Equilibrium model.

Describe any experiences and/or interactions during group work today that **decreased** your confidence in your ability to justify whether the population is evolving or not using the Hardy-Weinberg Equilibrium model.

With which gender do you identify?

○ Male (1)

○ Female (2)

○ Other (3) ____

○ Prefer not to respond (4)

What year are you in college?

○ First year (1)

○ Second year (2)

○ Third year (3)

○ Fourth year (4)

○ Other (5) ____

○ Prefer not to respond (6)

What is your major?

____

Of the following, which is the highest mathematics course you took in high school?

○ Algebra or Geometry (1)

○ Trigonometry (2)

○ Pre-calculus (3)

○ Calculus (4)

○ Prefer not to respond (5)

Population Growth Pre-Survey

Name: ___________________________

Please consider the following problem about population growth. **You <u>do not</u> have to solve it**.

Cod is an economically important fish species in the fishing industry. Unfortunately, overfishing has depleted cod populations in some areas. A group of fisheries biologists is monitoring one particular cod population that is currently closed to fishing. The biologists estimated that the population size at the beginning of 2019 was 150 cod. Over the course of the year, they recorded 240 births and 60 deaths in the population. Assume the per capita population growth rate is the same every year, the carrying capacity of the population is 1000 cod, and the population can be modeled with the logistic growth model:

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right)$$

The fisheries biologists have agreed to re-open the population for fishing once the population surpasses its maximum growth rate. Will the population size in 2022 be large enough to allow fishing?

Please rate your confidence (circle the number) in your ability to successfully do the following:

	Not at all confident	A little confident	Fairly confident	Very confident	Completely confident
Predict the population size in the year 2022	**1**	**2**	**3**	**4**	**5**

Population Growth Post-Survey

Name: ___________________________

Please consider the following problem about population growth. **You <u>do not</u> have to solve it**.

Cod is an economically important fish species in the fishing industry. Unfortunately, overfishing has depleted cod populations in some areas. A group of fisheries biologists is monitoring one particular cod population that is currently closed to fishing. The biologists estimated that the population size at the beginning of 2019 was 150 cod. Over the course of the year, they recorded 240 births and 60 deaths in the population. Assume the per capita population growth rate is the same every year, the carrying capacity of the population is 1000 cod, and the population can be modeled with the logistic growth model:

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right)$$

The fisheries biologists have agreed to re-open the population for fishing once the population surpasses its maximum growth rate. Will the population size in 2022 be large enough to allow fishing?

Please rate your confidence (circle the number) in your ability to successfully do the following:

	Not at all confident	A little confident	Fairly confident	Very confident	Completely confident
Predict the population size in the year 2022	**1**	**2**	**3**	**4**	**5**

Describe any experiences and/or interactions during group work today that **<u>increased</u>** your confidence in your ability to predict the population size in the year 2022.

Describe any experiences and/or interactions during group work today that **<u>decreased</u>** your confidence in your ability to predict the population size in the year 2022.

With which gender do you identify?

- ◯ Male (1)
- ◯ Female (2)
- ◯ Other (3) ____
- ◯ Prefer not to respond (4)

What year are you in college?

- ○ First year (1)
- ○ Second year (2)
- ○ Third year (3)
- ○ Fourth year (4)
- ○ Other (5) ____
- ○ Prefer not to respond (6)

What is your major?

____

Of the following, which is the highest mathematics course you took in high school?

- ○ Algebra or Geometry (1)
- ○ Trigonometry (2)
- ○ Pre-calculus (3)
- ○ Calculus (4)
- ○ Prefer not to respond (5)

Appendix B – Survey Instruments for Chapter 2

Task-Value Facets

Intrinsic Value

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Statistics is fun to me. (1)	○	○	○	○	○	○	○
I like doing statistics. (2)	○	○	○	○	○	○	○
I simply like statistics. (3)	○	○	○	○	○	○	○
I enjoy dealing with statistical topics. (4)	○	○	○	○	○	○	○

Attainment Value – Importance of Achievement

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
It is important to me to be good at statistics. (1)	○	○	○	○	○	○	○
Being good at statistics means a lot to me. (2)	○	○	○	○	○	○	○
Performing well in statistics is important to me. (3)	○	○	○	○	○	○	○
Good grades in statistics are very important to me. (4)	○	○	○	○	○	○	○

Attainment Value – Personal Importance

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I care a lot about remembering the things we learn in statistics. (1)	○	○	○	○	○	○	○
Statistics is not meaningful to me. (2)	○	○	○	○	○	○	○
I‘m really keen on learning a lot in statistics. (3)	○	○	○	○	○	○	○
Statistics is very important to me personally. (4)	○	○	○	○	○	○	○
To be honest, I don't care about statistics. (5)	○	○	○	○	○	○	○
It is important to me to know a lot of statistics. (6)	○	○	○	○	○	○	○

Utility Value - School

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
It is worth making an effort in statistics, because it will save me a lot of trouble at school in the next years. (1)	○	○	○	○	○	○	○
Being good at statistics pays off, because it is simply needed at school. (2)	○	○	○	○	○	○	○

Utility Value – Daily Life

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Understanding statistics has many benefits in my daily life. (1)	○	○	○	○	○	○	○
Statistics comes in handy in everyday life and leisure time. (2)	○	○	○	○	○	○	○
Statistics is directly applicable in everyday life. (3)	○	○	○	○	○	○	○

Utility Value – Social Utility

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Being well versed in statistics will go down well with my classmates. (1)	○	○	○	○	○	○	○
I can impress others with intimate knowledge in statistics. (2)	○	○	○	○	○	○	○
If I know a lot in statistics, I will leave a good impression on my classmates. (3)	○	○	○	○	○	○	○

Utility Value – Career/Job

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Good grades in statistics can be of great value to me later on. (1)	○	○	○	○	○	○	○
Learning statistics is worthwhile, because it improves my job and career chances. (2)	○	○	○	○	○	○	○

Utility Value – Future Life

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Statistics contents will help me in my life. (1)	○	○	○	○	○	○	○
I will often need statistics in my life. (2)	○	○	○	○	○	○	○

Cost – Effort Required

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
Doing statistics is exhausting to me. (1)	○	○	○	○	○	○	○
I often feel completely drained after doing statistics. (2)	○	○	○	○	○	○	○
Dealing with statistics drains a lot of my energy. (3)	○	○	○	○	○	○	○
Learning statistics exhausts me. (4)	○	○	○	○	○	○	○

Cost – Emotional Cost

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I'd rather not do statistics, because it only worries me. (1)	○	○	○	○	○	○	○
When I deal with statistics, I get annoyed. (2)	○	○	○	○	○	○	○
Statistics is a real burden to me. (3)	○	○	○	○	○	○	○
Doing statistics makes me really nervous. (4)	○	○	○	○	○	○	○

Cost – Opportunity Cost

Please rate how much you agree/disagree with each statement. There are no right or wrong answers.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I have to give up other activities that I like doing in order to be successful at statistics. (1)	○	○	○	○	○	○	○
I have to give up a lot to do well in statistics. (2)	○	○	○	○	○	○	○
I'd have to sacrifice a lot of free time to be good at statistics. (3)	○	○	○	○	○	○	○

Demographics

What year are you in college?

○ First year (1)

○ Second year (2)

○ Third year (3)

○ Fourth year (4)

○ Other (5) ____

○ Prefer not to respond (6)

Have you taken a statistics course prior to this course (at college, in high school, or elsewhere)?

○ Yes (1)

○ No (2)

○ Prefer not to respond (3)

Which of the following describes you with relation to pre-professional programs?

○ I am pre-medicine (1)

○ I am pre-dental (2)

○ I am pre-pharmacy (3)

○ I am pre-veterinary medicine (4)

○ I am in a pre-professional science program not listed here (5)
____

○ I am not in a pre-professional program (6)

○ Prefer not to respond (7)

What is your major? (If you prefer not to respond, please indicate that in the text box)

____

What is your current cumulative GPA? (If you prefer not to respond, please indicate that in the text box)

____

What is your gender?

- ○ Male (1)
- ○ Female (2)
- ○ Other (3) ____
- ○ Prefer not to respond (4)

With which race(s)/ethnicity do you most closely identify? Please choose all that apply.

- ☐ American Indian or Alaska Native (1)
- ☐ Asian (2)
- ☐ Black or African American (3)
- ☐ Native Hawaiian or other Pacific Islander (4)
- ☐ White (not Hispanic or Latinx) (5)
- ☐ Hispanic or Latinx (6)
- ☐ Other (7) ____
- ☐ Prefer not to respond (8)

What is the highest level of education obtained by any parents or guardians in your household?

- ○ Some high school (1)
- ○ High school/GED (2)
- ○ Some college (3)
- ○ Trade/technical school degree (4)
- ○ Associate's degree (5)
- ○ Bachelor's degree (6)
- ○ Master's degree (7)
- ○ Doctorate degree (8)
- ○ Professional degree (9)
- ○ Other (10) ____
- ○ I don't know (11)
- ○ Prefer not to respond (12)

Appendix C – Correlation Table for Survey Items

Correlations in parentheses were not statistically significant, all other correlations reported significant at $p < 0.05$.

Task Value	Task-Value Facet	#	1	2	3	4	5	6	7	8	9	10	11
Intrinsic value	Intrinsic value	1	~										
Attainment value	Importance of acheivement	2	0.46	~									
	Personal importance	3	0.70	0.65	~								
Utility value	Utility for school	4	0.33	0.43	0.46	~							
	Utility for daily life	5	0.49	0.38	0.61	0.33	~						
	Social utility	6	0.37	0.31	0.44	0.39	0.49	~					
	Utility for career/job	7	0.32	0.55	0.46	0.53	0.34	0.29	~				
	General utility for future life	8	0.44	0.51	0.67	0.49	0.60	0.44	0.54	~			
Cost	Effort required	9	-0.39	-0.12	-0.30	-0.18	-0.17	-0.16	-0.10	-0.15	~		
	Emotional cost	10	-0.53	-0.28	-0.51	-0.27	-0.29	-0.22	-0.22	-0.32	0.71	~	
	Opportunity cost	11	-0.25	(-0.08)	-0.17	(-0.10)	-0.12	(-0.04)	(-0.09)	(-0.11)	0.59	0.59	~

Appendix D – IRB Approval Letters

IRB Approval for # 7005

University *of* New Hampshire

Research Integrity Services, Service Building
51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

23-Aug-2018

Aikens, Melissa L
Dept of Biological Sciences
Rudman Hall
Durham, NH 03824-2618

IRB #: 7005
Study: Exploring Student Motivation and Learning of Quantitative Skills in Introductory Biology
Approval Date: 21-Aug-2018

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved the protocol for your study as Expedited as described in Title 45, Code of Federal Regulations (CFR), Part 46, Subsection 110.

Approval is granted to conduct your study as described in your protocol for one year from the approval date above. At the end of the approval period, you will be asked to submit a report with regard to the involvement of human subjects in this study. If your study is still active, you may request an extension of IRB approval.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the attached document, *Responsibilities of Directors of Research Studies Involving Human Subjects.* (This document is also available at http://unh.edu/research/irb-application-resources.) Please read this document carefully before commencing your work involving human subjects.

Note: IRB approval is separate from UNH Purchasing approval of any proposed methods of paying study participants. Before making any payments to study participants, researchers should consult with their BSC or UNH Purchasing to ensure they are complying with institutional requirements. If such institutional requirements are not consistent with the confidentiality or anonymity assurances in the IRB-approved protocol and consent documents, the researcher may need to request a modification from the IRB.

If you have questions or concerns about your study or this approval, please feel free to contact Melissa McGee at 603-862-2005 or melissa.mcgee@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F Simpson

Julie F. Simpson
Director

cc: File
,
,

IRB Approval for # 8077

University *of* New Hampshire

Research Integrity Services, Service Building
51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

18-Apr-2019

Kulacki, Alexander
Biological Sciences, Spaulding Hall
38 Academic Way
Durham, NH 03824

IRB #: 8077
Study: Exploring Student Motivation and Conceptual Understanding in Applied Biostatistics
Approval Date: 15-Apr-2019

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved the protocol for your study as Exempt as described in Title 45, Code of Federal Regulations (CFR), Part 46, Subsection 104(d). Approval is granted to conduct your study as described in your protocol.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the attached document, *Responsibilities of Directors of Research Studies Involving Human Subjects.* (This document is also available at http://unh.edu/research/irb-application-resources.) Please read this document carefully before commencing your work involving human subjects.

Note: IRB approval is separate from UNH Purchasing approval of any proposed methods of paying study participants. Before making any payments to study participants, researchers should consult with their BSC or UNH Purchasing to ensure they are complying with institutional requirements. If such institutional requirements are not consistent with the confidentiality or anonymity assurances in the IRB-approved protocol and consent documents, the researcher may need to request a modification from the IRB.

Upon completion of your study, please complete the enclosed Exempt Study Final Report form and return it to this office along with a report of your findings.

If you have questions or concerns about your study or this approval, please feel free to contact Melissa McGee at 603-862-2005 or melissa.mcgee@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F. Simpson

Julie F. Simpson
Director

cc: File
Aikens, Melissa